

Bildsequenzanalyse
für die passive Navigation

Von der Fakultät für
Bauingenieur- und Vermessungswesen
der Universität Stuttgart
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

vorgelegt von

Dipl.-Ing. Michael Hahn

München 1995

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

Adresse der Deutschen Geodätischen Kommission:

Deutsche Geodätische Kommission
Marstallplatz 8 • D – 80 539 München

Telefon (089) 23 031 113 • Telefax (089) 23 031 – 100/– 240
Telex 5 213 550 dgfi d • E-mail hornik@dgfi.badw-muenchen.de

Hauptberichter: Prof.i.R. Dr.-Ing. Dr.h.c.mult. Friedrich Ackermann

Mitberichter: Prof. Dr.-Ing.habil. Dieter Fritsch
Prof. Dr.-Ing.habil. Wolfgang Förstner

Tag der Einreichung: 16.5.1994

Tag der mündlichen Prüfung: 14.7.1994

© 1995 Deutsche Geodätische Kommission, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| I | Einführung in die Bildfolgenanalyse | 9 |
| 1 | Einleitung | 9 |
| 1.1 | Historie | 9 |
| 1.2 | Übersicht über die Arbeit | 10 |
| 1.3 | Bildentstehung: Bildfolge | 11 |
| 1.4 | Bildfolgenanalyse und Bewegungsdeutung | 13 |
| 1.4.1 | Änderungsdetektion | 13 |
| 1.4.2 | Bewegungsschätzung und -deutung | 13 |
| 1.4.3 | Kriterien zur Bewegungsdeutung | 14 |
| 1.5 | Bewegungsmessung in Bildsequenzen | 15 |
| 1.5.1 | Kriterien zur Klassifizierung von Zuordnungsverfahren | 16 |
| 1.5.2 | Intensitätsbasierte Verfahren | 16 |
| 1.5.3 | Merkmalsbasierte Verfahren | 18 |
| 1.6 | 3D-Rekonstruktion von Bewegung und Struktur | 18 |
| 1.6.1 | Inversionsansätze | 19 |
| 1.6.2 | Projektionsansätze | 20 |
| 1.7 | Passive Navigation | 21 |
| 1.7.1 | Abgrenzung passiver von aktiver visueller Navigation | 22 |
| 1.7.2 | Rekursive kinematische Modellierung | 22 |
| 1.7.3 | Der Realwelt-Datensatz: Bildsequenz Rastatt | 24 |
| 1.8 | Problemstellung und Zielsetzung dieser Arbeit | 26 |
| 2 | Bewegungsmodellierung | 29 |
| 2.1 | Bewegung im dreidimensionalen Raum | 30 |
| 2.1.1 | Starrkörperbewegung | 30 |
| 2.1.2 | Differentielle Bewegung | 32 |
| 2.2 | Rotation | 33 |
| 2.2.1 | Kriterien für die Parametrisierung | 34 |
| 2.2.2 | Parametrisierung mit den Elementen des Einheitsquaternions | 35 |
| 2.2.3 | Interpretation der Quaternionen-Parametrisierung | 37 |
| 2.2.4 | Zusammengesetzte Rotationen | 38 |
| 2.2.5 | Ermittlung von Quaternionen-Parametern aus gegebener Rotationsmatrix | 39 |
| 2.2.6 | Differentielle Rotationsparameter | 39 |
| 2.3 | Projizierte Bewegung | 40 |

| | | |
|-------|---|----|
| 2.3.1 | Projektion | 40 |
| 2.3.2 | Differentielle projizierte Bewegung | 41 |
| 2.3.3 | Starrheitsannahme | 42 |
| 2.3.4 | Objektmodell Ebene | 43 |
| 2.3.5 | Lokal quadratische Oberflächen | 45 |
| 2.4 | Rekonstruktionsansätze | 46 |
| 2.4.1 | Die beobachteten Daten | 47 |
| 2.4.2 | Lösungswege und Anwendungen bei geringer Anzahl von Bildern | 48 |
| 2.4.3 | Rekonstruktionsansatz für die lange Bildsequenz | 51 |

II Bildzuordnung 56

3 Schätzung des optischen Flusses 59

| | | |
|-------|--|----|
| 3.1 | Regularisierung und finite Elemente Analyse | 60 |
| 3.2 | Variationsformulierungen zur Schätzung des optischen Flusses | 63 |
| 3.2.1 | Variationsprobleme und Euler'sche Gleichung | 63 |
| 3.2.2 | Beobachtungsmodelle des optischen Flusses | 64 |
| 3.2.3 | Variationslösung von Horn und Schunk | 67 |
| 3.2.4 | Weitere Ansätze von Hildreth und von Nagel | 68 |
| 3.2.5 | Diskussion | 70 |
| 3.3 | Bewegungsschätzung in örtlich begrenzten Bildausschnitten | 71 |
| 3.3.1 | Starre Szene | 71 |
| 3.3.2 | Lokal schwach gekrümmte Flächen | 72 |
| 3.3.3 | Lokal ebene Objekte | 73 |
| 3.3.4 | Differentialgeometrische Interpretation | 73 |
| 3.3.5 | Konturorientierte Beobachtungen | 74 |
| 3.3.6 | Affinapproximation | 75 |
| 3.3.7 | Lokale Fenster | 75 |
| 3.3.8 | Differentialgeometrische Modellierung der Bildfunktion | 76 |
| 3.3.9 | Diskussion | 78 |
| 3.4 | Ein Vergleich von Variations- und Kleinste-Quadrate-Formulierungen | 80 |
| 3.4.1 | Bildzuordnung in der Photogrammetrie | 80 |
| 3.4.2 | Variations- versus Kleinste-Quadrate-Formulierung: Ein Beispiel | 82 |
| 3.5 | Bildflußparadigma | 85 |

| | | |
|------------|---|------------|
| 4 | Markante Punkte und Korrespondenz | 90 |
| 4.1 | Kriterien an die Extraktion markanter Punkte | 91 |
| 4.2 | Operatoren zur Extraktion markanter Punkte | 93 |
| 4.2.1 | Klassifizierung in indirekte und direkte Verfahren | 93 |
| 4.2.2 | Punktfinder: Statistische Verfahren | 95 |
| 4.2.3 | Punktfinder: Differentialgeometrische Verfahren | 97 |
| 4.2.4 | Punktfinder: Klassifizierende Verfahren | 102 |
| 4.2.5 | Weitere Verfahren und abschließende Bewertung | 109 |
| 4.3 | Merkmalsbasierte Punktverfolgung - Experimentelle Untersuchung | 112 |
| 4.3.1 | Verfahren zur Punktverfolgung | 112 |
| 4.3.2 | Experimente und Resultate zur Punktextraktion und -verfolgung | 114 |
| 4.3.3 | Beispiele verfolgter Punkte und Diskussion | 117 |
| 5 | Intensitätsbasierte Bildzuordnung im Scale-Space | 122 |
| 5.1 | Bildpyramiden und Scale-Space | 122 |
| 5.2 | Ein Verfahren zur Bildzuordnung im Scale-Space | 124 |
| 5.2.1 | Hintergründe und theoretische Grundlagen | 124 |
| 5.2.2 | Konzeption des Verfahrens | 125 |
| 5.2.3 | Parametrischer Ansatz und Kleinste-Quadrate-Lösung | 127 |
| 5.3 | Intensitätsbasierte Punktverfolgung - Experimentelle Untersuchung | 128 |
| 5.3.1 | Vorgehensweise | 129 |
| 5.3.2 | Voruntersuchungen | 130 |
| 5.3.3 | Ergebnisse und Diskussion | 132 |
| III | 3D-Rekonstruktion von Bewegung und Struktur | 137 |
| 6 | Entwurf und Realisierung eines 3D-Kalman-Filters | 137 |
| 6.1 | Kalman-Filter | 137 |
| 6.2 | Entwurf eines 3D-Kalman-Filters | 140 |
| 6.2.1 | Konzeption des Verfahrens | 140 |
| 6.2.2 | Realisierung | 141 |
| 6.3 | Experimentelle Untersuchungen | 142 |
| 6.3.1 | Kontrollinformation | 142 |
| 6.3.2 | Verschiedene Versionen der Filterung | 143 |
| 6.3.3 | Bereinigung von Zuordnungsfehlern | 144 |
| 6.3.4 | Schätzung der Eigenbewegung | 145 |
| 6.3.5 | Schätzung von Strukturparametern | 150 |
| 6.4 | Zusammenfassende Beurteilung | 154 |

| | |
|---------------------------------------|------------|
| 7 Zusammenfassung und Ausblick | 155 |
| Literatur | 157 |
| Danksagung | 168 |
| Lebenslauf | 168 |

Abbildungsverzeichnis

| | | |
|-----|---|-----|
| 1.1 | Raum-Zeit-Bildblock (entnommen aus Baker, 1987) | 12 |
| 1.2 | Unbestimmtheit der Bewegung an einer Kante | 17 |
| 1.3 | Random-dot stereogram. (Ein räumlicher Eindruck entsteht z. B. bei beidäugiger Betrachtung mit einem kleinen Stereoskop.) | 20 |
| 1.4 | Prinzipskizze eines passiv navigierenden Sensors | 23 |
| 1.5 | Aufnahme 40 und 120 der Rastatt-Bildsequenz | 24 |
| 1.6 | Kalibrierszene | 26 |
| 2.1 | Das Fußgängermodell umfaßt ein Körpermodell und ein Bewegungsmodell (aus Rohr und Nagel, 1990, S. 53) | 30 |
| 2.2 | Interpretation der Bewegungsparameter: Translatorische und rotatorische Geschwindigkeit der Starrkörperbewegung sind mit Bezugspunkt O dargestellt (detaillierte Erläuterung im Text). | 33 |
| 2.3 | Kameramodell: Die Bildkoordinaten eines projizierten Punktes können in der Bildebene durch (x, y) oder im Kamerakoordinatensystem durch $(X = x, Y = y, Z = c)$ angegeben werden. | 41 |
| 3.1 | Aspekte der Zuordnungsproblematik | 59 |
| 3.2 | Geometrische Interpretation der Beobachtungsgleichung | 65 |
| 3.3 | Bildpyramide und rekonstruiertes DHM (aus Ackermann und Hahn, 1991). Links: 3 Ebenen der Bildpyramide eines Stereobildpartners. Rechts: Rekonstruktionsresultate der entsprechenden Pyramidenebenen | 83 |
| 3.4 | Elemente des Bildflußparadigmas (vereinfachte Fassung aus Waxman, 1987) | 88 |
| 4.1 | Modell einer Grauwertecke (aus Nagel, 1983, S. 89) | 101 |
| 4.2 | Keilförmige Struktur (ideale Ecke) | 107 |
| 4.3 | 1. Reihe: L -Knoten, Y -Knoten, kreisförmige Bildstruktur, 2. Reihe: Grauwertkeil vor konstantem Hintergrund, Y -Knoten als 2.5 D-Ecke, Scheibe vor Hintergrund mit Lichtabfall | 110 |
| 4.4 | Suchbereiche bei Elimination der rotatorischen Bewegungsanteile | 113 |
| 4.5 | Übertragung des FOE zur Approximation der Sensorrotation zwischen Bild i und $i + 1$. | 114 |
| 4.6 | Verschiebungsvektoren an ausgewählten Beispielen | 116 |
| 4.7 | Punktverfolgung über Abschnitte von jeweils 20 Bildern. Dargestellt ist der translatorische Bewegungsanteil der Verschiebungsvektoren | 117 |

| | | |
|------|---|-----|
| 4.8 | Beispiele verfolgter Punkte | 119 |
| 4.9 | Beispiele verfolgter Punkte | 120 |
| 4.10 | Beispiele verfolgter Punkte | 121 |
| 5.1 | Nulldurchgänge der zweiten Ableitungen auf verschiedenen Glättungsniveaus (aus Witkin, 1983) | 123 |
| 5.2 | Bildzuordnungsaufgabe: Zu bestimmen sind der Verschiebungsvektor \mathbf{u}_0 und die Schärfedifferenz β_0 zwischen beiden Bildausschnitten | 125 |
| 5.3 | Bildzuordnung im Scale-Space | 126 |
| 5.4 | Konvergenzverhalten bei Scale-Space-basierter Bildzuordnung (aus Hahn, 1990) | 126 |
| 5.5 | Konvergenzverhalten für β_0 | 130 |
| 5.6 | Konvergenzverhalten für u, v (Bildpaar 132/133) | 131 |
| 5.7 | Histogramm: Aufgetragen ist die Anzahl der Zuordnungen über der vom Zuordnungsverfahren benötigten Anzahl Iterationen | 133 |
| 5.8 | Iterationsverhalten im Scale-Space. Kontrollpunkt 5032, Bildpaar 35/36. Der ungeglättete Bildausschnitt (jeweils die linke Spalte) zeigt die jeweiligen Verschiebungen. | 134 |
| 6.1 | Systemeinheit "Einzelbild" | 140 |
| 6.2 | Systemeinheit "Bildpaar" | 140 |
| 6.3 | Bündellösung: Geschätzte Winkel | 146 |
| 6.4 | Kalman Version A: Geschätzte Winkel | 146 |
| 6.5 | Winkeldifferenz zwischen Bündellösung und Kalman Version A | 147 |
| 6.6 | Winkeldifferenz zwischen Bündellösung und Kalman Version D | 148 |
| 6.7 | Positionsdifferenz zwischen Bündellösung und Kalman Version A | 148 |
| 6.8 | Theoretische Genauigkeit der äußeren Orientierung (Bündellösung) | 149 |
| 6.9 | Theoretische Genauigkeit der äußeren Orientierung (Kalman Version A) | 149 |
| 6.10 | Koordinatendifferenzen in ausgewählten Objektpunkten zwischen der Bündellösung und der Kalman-Filterung (Versionen A und C). | 152 |
| 6.11 | Theoretische Genauigkeit der geschätzten Objektpunktkoordinaten von Version A und C der Kalman-Filterung. | 153 |



Teil I

Einführung in die Bildfolgenanalyse

1 Einleitung

Die Analyse von Bildfolgen ist ein dominanter Themenschwerpunkt in den Bereichen Rechner Sehen, Bildanalyse und Bildverstehen. Diese Bereiche der Bildverarbeitung befassen sich damit, aus den erfaßten Bilddaten bedeutungsvolle Interpretationen und Beschreibungen über die abgebildete Szene zu gewinnen. Was als "bedeutungsvoll" gilt, ist direkt vom Ziel der Verarbeitung und damit von der Anwendung abhängig. Eine häufig genannte Forschungsthema der Bildfolgenanalyse ist die visuelle Navigation. Hier zielt die Bildanalyse auf die Extraktion einer räumlich-zeitlichen Beschreibung der beobachteten Szene. Auf dieser Grundlage lassen sich beispielsweise Verkehrssituationen analysieren oder entsprechende Aktionen eines Industrieroboters in der Szene planen und ausführen.

Die Analyse von Bildfolgen hat schon relativ früh im Bereich der digitalen Bildverarbeitung einen Forschungsschwerpunkt gebildet. Die Einführung soll deshalb mit einem Blick in die Historie dieses Themas beginnen.

1.1 Historie

Als Geburtsdatum der digitalen Bildverarbeitung werden die frühen 20er Jahre dieses Jahrhunderts betrachtet, in denen es erstmals gelang, ein digitalisiertes Bild über Kabelverbindung zwischen London und New York zu übertragen. Danach gab es einen verhältnismäßig langen Zeitraum, in dem wenig Spektakuläres über die Bildverarbeitung zu berichten war. Erst 1964 hat mit der Übertragung der durch das unbemannte Raumfahrzeug *Ranger VII* erfaßten Mondaufnahmen die Bildverarbeitung neuen Auftrieb erhalten (Castleman, 1979, Gonzalez und Wintz, 1977), zumal auch die Computerentwicklung in dieser Zeit gewaltige Fortschritte machte. Von diesem Raumfahrzeug wurden insgesamt 4316 Bilder zur Erde übertragen. Der Erfolg der *Ranger VII* und nachfolgender Raumfahrtmissionen, die alle mit Videokameras ausgestattet wurden, hat euphorische Ideen bezüglich autonom agierender Roboter genährt. Die ersten Schritte in diese Richtung erfolgten postwendend in 1966, als die "lunar soft-landers" des *Surveyor I* - Raumfahrzeugs mit Videokameras versehen waren. Mit Hilfe der Videobilder wurde eine gezielte Auswahl der zu entnehmenden Bodenproben ermöglicht. Die Entwicklung algorithmischer Lösungen zur Analyse von Bildern bzw. Bildfolgen - in den fünf *Surveyor* Missionen wurden 87674 Bilder auf die Erde übertragen - wurden in der Folgezeit zunehmend zu Zielen des Rechner Sehens (Computer Vision).

Die grundlegende Idee, dem Rechner Aufgaben zu übertragen, die vom menschlichen visuellen System geleistet werden, wird Selfridge (1955) zugeschrieben. Allerdings sieht sich die Erforschung des menschlichen visuellen Sehvorgangs heute selbst als eine Disziplin, die noch in den Anfängen steckt (Gibson, 1950, Cornsweet, 1970, Marr, 1982). Dennoch sind durch algorithmische Nachbildungen der beim Sehvorgang erfolgten Leistungen (Marr, 1982, Grimson, 1981, Pentland, 1986) beachtliche Teilerfolge hinsichtlich einer automatischen Analyse und Interpretation von Bildern durch den Rechner erzielt worden. So gibt es heute Beispiele aus der Qualitätskontrolle, der Inspektion, der Überwachung und anderen industriellen Aufgaben, in denen durch die maschinelle Verarbeitung von Bilddaten Aufgaben wahrgenommen werden, ohne daß menschliche Eingriffe stattzufinden brauchen.

Die Entwicklungen seit den 60er Jahren waren gekennzeichnet von Paradigmenwechseln, mit denen sich die Forschungsausrichtung auf neue Schwerpunkte konzentrierte (Neumann, 1988). Einer der ersten Schwerpunkte lag in der Zeichenerkennung, aus der das Mustererkennungsparadigma abstrahiert wurde. Damit konzentrierte man sich auf das Ziel, Objekte anhand ihrer Merkmalsvektoren zu klassifizieren. Eine Umorientierung brachte die Arbeit von Roberts (1965) und anderer Forscher am MIT. Hier wurde die dreidimensionale Welt, die sogenannte Blockswelt, in den Mittelpunkt der Forschung gerückt. Schwerpunkte lagen vor allem in der Analyse von 2D-Linienzeichnungen und deren Zuordnung zu Blocksweltmodellen. Für die Bildfolgenanalyse begann in den 70er Jahren schon relativ früh ein Rückbesinnen auf die Realwelt und ihre Anwendungen. In der Folge

wurden im Bereich der Bewegungsanalyse eine Vielzahl algorithmischer Lösungswege für Einzelproblemstellungen ausgearbeitet. Diese betrafen die Schätzung des optischen Flusses, d. h. der visuell sichtbaren Bewegung im Bild und die Rekonstruktion von Bewegung aus 2D- oder 3D-Punkt- und Linienkorrespondenzen, um die sich mehrere Forschergruppen bemüht haben (Huang, 1990). Rückblickend konstatierte Horn (1986, S. 417): “... in the process, much of what is known in photogrammetry has been rediscovered”, wobei er insbesondere die Orientierungsverfahren der analytischen Photogrammetrie anspricht.

Der Blick hin zu einem theoretischen Fundament mit der Intention, Erkenntnisse über die menschliche visuelle Wahrnehmung zur Formulierung von Berechnungstheorien für das Sehen aufzugreifen, wurde von Marr (1982) forciert und führte zu einem neuen Paradigmenwechsel. Eine solche Erkenntnis aus psychologischen Studien der menschlichen Wahrnehmung ist z. B. die, daß der Mensch in der Lage ist, Bewegung und Form eines (unbekannten) Objekts schon allein aufgrund der visuell beobachteten Bewegung zu erkennen (Ullman, 1987, 1979). Bewegungsanalyse ist aus heutiger Sicht im wesentlichen eine Informationsquelle zur Szenerekonstruktion und -deutung. Sie stellt sich als ein zweistufiger Prozeß dar, in dem die implizit in der Bildfolge enthaltene Bewegungsinformation in einem ersten Schritt ermittelt wird. Die im zweiten Schritt durchgeführte dreidimensionale Rekonstruktion von Form und Bewegung beinhaltet eine erste Deutung der in der Bildebene erfaßten Bewegung. Diese Zweiteilung mit bildraumspezifischer Messung und deren Interpretation durch eine objektraumspezifische Rekonstruktion ist bis heute aktueller Stand der Forschung. Eine Vielzahl von Anwendungen zur Bildsequenz- und Szeneanalyse sind in Huang (1981, 1983) dokumentiert und zeigen den Entwicklungsstand in der Signalübertragung (bewegungskompensierte Kodierung der Bildsequenzen), der Robotik (Verfolgen von Objekten, Erkennen von Werkstücken, Montage), der Meteorologie (Wolkenbewegung aus Satellitenbildern), dem Verkehrswesen (Überwachung von Straßen oder Bahnsteigen) und in der Medizin (Bewegung von Zellen, Herz, Blutzirkulation) auf. Auffallend ist die Beschränkung auf die Zweibildkonfiguration.

Marr's Theorie über das Bildverstehen hat insbesondere das Interesse an der Strukturierung von Systemen zur Analyse und zum Verstehen einer dynamischen Szene geweckt. Die meisten Vorschläge lehnen sich an die Vorstellungen Marr's an, in der in einer datengetriebenen (bottom-up) Verarbeitungsrichtung die algorithmische Transformation zur Repräsentationsebene der rekonstruierten Szenelemente (3D-Oberflächen, Objektgrenzen, Volumina, aber auch Position und Geschwindigkeit von Objekt oder Sensor) führt. Aufgaben wie die Objekterkennung und das Erkennen von Ereignissen und Situationen sind erwartungsgesteuerte (top-down) Prozesse, die gezielt das Vorhandensein bestimmter Objekte oder Ereignisse prüfen. Die einzelnen in der Literatur zu findenden Strukturierungsvorschläge, z. B. von Kanade (1978), Niemann (1985), Neumann (1988) oder Huang (1990) unterscheiden sich vor allem durch die Betonung einer besonderen Sichtweise, z. B. der Rolle des wissensbasierten Aspektes. Obwohl die algorithmischen Verfahren aller Abstraktionsebenen auf Wissen basieren, greifen Verfahren der Objekterkennung auf ein Objektmodell aus einer Wissensbasis explizit zurück, wohingegen die Verfahren der Bildanalyse bis hin zur Rekonstruktion von Szenelementen Wissen, z. B. über die Gesetzmäßigkeiten aus der projektiven Geometrie, meist implizit enthalten. Eine explizite Repräsentation wäre hier zu ineffektiv (Neumann, 1988).

Die Arbeit will einen Beitrag zur Analyse und zum Verstehen dynamischer Szenen leisten. Ein Schwerpunkt der Arbeit liegt dabei in dem Entwurf eines Systems zur Bestimmung von Bewegungs- und Strukturparametern, wobei die Analyse langer Bildsequenzen für die passive Navigation bearbeitet werden soll. Für ein solches System ist die Bewegungsmessung in der Bildsequenz von großer Bedeutung, weshalb hierin ein zweiter Schwerpunkt der Arbeit liegt. Es werden Meß- und Rekonstruktionsverfahren als Komponenten eines Systems entwickelt und bezüglich ihrer Leistungsfähigkeit an einer Realweltsequenz experimentell erprobt.

1.2 Übersicht über die Arbeit

Die vorliegende Arbeit zur Bildsequenzanalyse für die passive Navigation ist in drei Teile gegliedert. Sie beginnt mit einer ausführlichen Einführung in die verschiedenen Aspekte der Analyse von Bildsequenzen. Die Bewegungsmessung in der Bildsequenz bildet den Schwerpunkt im zweiten Teil der Arbeit. Der dritte Teil widmet sich schließlich der Bestimmung von 3D-Bewegung und -Struktur unter dem Aspekt eines passiv navigierenden Systems.

Ziel des ersten Kapitels ist es, in die wichtigsten Aspekte der Bildsequenzanalyse und deren Ausrichtung auf die passive Navigation einzuführen. Diese allgemeine Einführung kann von Lesern übersprungen werden, denen die Literatur im Bereich der Bildsequenzanalyse vertraut ist. Am Ende der Einführung (Abschnitt 1.7.3) wird die in den Experimenten verwendete Bildsequenz charakterisiert. Die Einführung dient auch der Vorbereitung von Abschnitt 1.8, in dem die spezifische Problemstellung und Zielsetzung dieser Arbeit ausführlich dargelegt wird. Kapitel 2 diskutiert verschiedene Aspekte der Modellierung von Starrkörperbewegung, wobei Schwerpunkte in der Parametrisierung rotatorischer Bewegung, in der projizierten Bewegung und der Einbeziehung verschiedener Oberflächenmodelle liegen. Aus dieser sehr grundlegenden Aufarbeitung des Geometrieaspektes wird ein Rekonstruktionsansatz für die Bestimmung von Bewegung und Struktur aus einer langen Bildsequenz entwickelt.

Der zweite Teil dieser Arbeit zielt auf die Messung von Bewegung in der Bildsequenz. Eine Vielfalt verschiedener Ansätze und Lösungswege zur Bestimmung des optischen Flusses wird in Kapitel 3 aufgezeigt und vergleichend diskutiert. Die in der Photogrammetrie weitgehend unabhängig entstandenen Arbeiten zur Bildzuordnung sind in diese auf die Theorie und Systematisierung bedachte Aufarbeitung eingegliedert. Die Messung der Bewegung in der Bildebene muß sich auch daran orientieren, in welcher Art und Weise das Resultat des Meßprozesses weiter verwendet wird. Die 3D-Bewegungs- und -Strukturrekonstruktion soll in dieser Arbeit auf Punktmessungen aufgebaut werden. Essentiell für die Punktmessung im Bild sind Operatoren zur Extraktion markanter Punkte, die im Kapitel 4 vertieft werden. Experimente zur merkmalsbasierten Punktverfolgung zeigen Diskrepanzen von Theorie und Wirklichkeit auf, z. B. bezüglich der Operatoreigenschaft, die fordert, daß die extrahierten Punkte eine hohe Stabilität aufweisen sollen. Im Kapitel 5 wird ein Verfahren zur intensitätsbasierten Bildzuordnung präsentiert, das den sogenannten Scale Space - eine Art kontinuierliche Bildpyramide - in den Zuordnungsprozeß einbezieht. Die theoretischen Ausführungen in diesem zweiten Teil der Arbeit werden von Experimenten zur Punktverfolgung auf Merkmals- und Intensitätsbasis begleitet, in der die Analyse der oben angesprochenen Realwelt-Bildsequenz im Mittelpunkt steht.

Der dritte Teil der Arbeit behandelt die Rekonstruktion von 3D-Bewegung und -Struktur aus langen Bildsequenzen. Für die Anwendung in der passiven Navigation ist die fortlaufende Bestimmung von Bewegung und Struktur essentiell. Hierfür bieten sich rekursive Schätztechniken an. In Kapitel 6 wird eine Kalman-Filter Lösung vorgeschlagen, die die zeitlich zurückliegenden und bereits erfaßten Informationen über das Objekt und die Bewegung des Sensors mit den aktuellen Meßdaten sowie mit einem Bahnmodell integriert. Die experimentelle Verifikation und Analyse des entworfenen Systems erfolgt wiederum mit der Realweltsequenz. Diese Experimente bauen auf den im zweiten Teil der Arbeit gewonnenen Zuordnungsergebnissen auf und gehen mit Qualitätsanalysen der Meßdaten sowie der geschätzten Größen einher.

Jedem der folgenden Kapitel 2 bis 6 ist eine Kurzfassung vorangestellt, die den wissenschaftlichen Beitrag der Arbeit in dem entsprechenden Bereich hervorhebt. In Kapitel 7 werden schließlich die Ergebnisse dieser Arbeit zusammengefaßt.

In der Fortsetzung dieser Einführung erfolgt zunächst eine Präzisierung des Begriffs Bildfolge, danach eine Eingrenzung der Themenbereiche Bewegungsmessung, 3D-Rekonstruktion und Navigation.

1.3 Bildentstehung: Bildfolge

Die meisten Verfahren der Bildsegmentierung und Bildzuordnung legen implizit oder explizit eine Bildfunktion zugrunde, die im Fall der Bildfolgen als kontinuierliche Funktion eines dreidimensionalen Raum-Zeitblocks ("spatio-temporal volume") dargestellt werden kann. Raum steht hier für die in ihrer Ausdehnung begrenzte Bildebene (x, y) , Zeit (t) für die über der Bildebene abgetragene dritte Dimension eines Bildblocks, der beispielsweise wie in Abbildung 1.1 dargestellt werden kann.

Die häufig mit dem Begriff *Bild* assoziierte Vorstellung einer *Momentaufnahme* ist durch die Betrachtung eines wählbaren Zeitpunktes gegeben. Die Bildentstehung einer Momentaufnahme, auch wenn sie nur sehr vereinfacht betrachtet werden soll, umfaßt zumindest vier Gesichtspunkte:

- Die Lage eines Objekts in der Bildebene soll durch eine perspektive Projektion der dreidimensionalen Szene definiert sein.

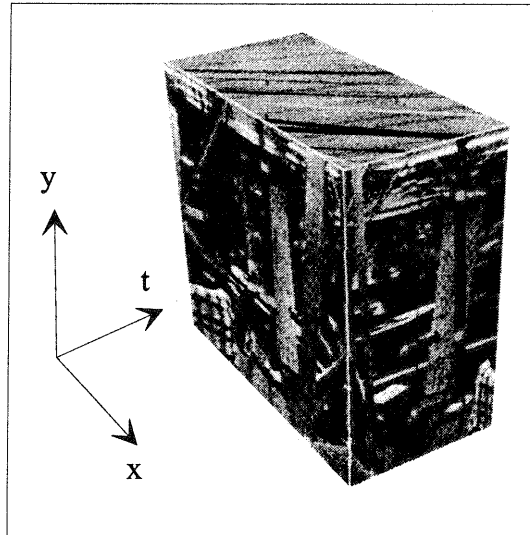


Abbildung 1.1: Raum-Zeit-Bildblock (entnommen aus Baker, 1987)

- Die Charakteristik des optischen Systems Kamera soll durch die Punktantwort, modelliert als 2D-Gaußfunktion, beschrieben werden.
- Da der Einfluß der Beleuchtung, aber auch der Reflexionseigenschaften des Objekts, auf Segmentierungs- und Zuordnungsverfahren meist sehr komplex ist, wird die Realität durch diffuse Beleuchtungsverhältnisse und Lambert'sche Reflexion approximiert.
- Die räumliche Diskretisierung der Bildfunktion und die Überlagerung durch Rauschen führen zum digitalen Intensitäts- oder Grauwertbild der Momentaufnahme.

Der Schritt vom Digitalbild einer Momentaufnahme hin zur Bildfolge korrespondiert mit der Diskretisierung der Zeitachse des Bildblocks. Die Anforderungen an das temporale Abtastintervall sollten einerseits direkt an die Bewegung der Kamera relativ zur Szene in der dreidimensionalen Welt geknüpft sein, andererseits werden sie meist über die visuell sichtbaren Änderungen der Grauwertmuster in der Bildebene formuliert und reichen von temporal nahezu kontinuierlichen Aufnahmen bis zu sehr großen Zeitabständen bei Szenen, die sich zeitlich nur sehr langsam ändern. Umgekehrt ist technisch bedingt häufig eine Abtastfrequenz vorgegeben, z. B. mit 25 Hz bei der europäischen Videonorm. Die Folge ist, daß sich die Methoden zur Verarbeitung entsprechender Bildsequenzen an dieser Vorgabe orientieren müssen.

Erfolgt die Charakterisierung von Bildfolgen in Abhängigkeit von der Bewegung in der Bildebene zwischen zwei aufeinanderfolgenden Aufnahmezeitpunkten, so gibt eine grobe Unterscheidung in

- dichte Bildfolge - kleine Bewegung von Bild zu Bild von maximal 1-2 Pixel
- lose Bildfolge - große Bewegung von Bild zu Bild von 20 bis zu 100 oder mehr Pixel,

bei einem Bildformat von 500×500 Bildelementen deutliche Hinweise für die Auswertung, z. B. durch Bildzuordnungsverfahren. Die temporale Diskretisierung der dichten Bildfolge ist in etwa der räumlichen Diskretisierung in der Bildebene angepaßt. Offensichtlich trifft dies für die lose Bildfolge nicht zu. Die bildraumbezogene Charakterisierung nach dichter und loser Sequenz läßt insbesondere 3D-Aspekte offen, zum Beispiel Informationen bezüglich der Dynamik der beobachteten Welt bzw. des Sensors. Eine sehr gleichförmige Bewegung in 3D führt in der Regel auch zu einer entsprechend glatten projizierten Bewegung. Im Kontext der Bewegungsdeutung sollen weitere Gesichtspunkte der Bildfolgenanalyse im nächsten Abschnitt vertieft werden.

1.4 Bildfolgenanalyse und Bewegungsdeutung

1.4.1 Änderungsdetektion

Die einfachste Form der Detektion bewegter Objekte besteht in der Analyse von Grauwertänderungen in aufeinanderfolgenden Bildern. Unter der Voraussetzung, daß die Bildsequenz von einer stationären Kamera erfaßt wurde, lassen sich die Grauwertdifferenzbilder zeitlich aufeinanderfolgender Aufnahmen direkt pixelbezogen bewerten. Die klassischen, statistischen Techniken der Formulierung von Hypothesentests mit Null- und Alternativhypothesen, und damit auch der Zuverlässigkeits- und Trennbarkeitsanalyse, können zur Beurteilung herangezogen werden (Kories, 1988). Ein Bildpunkt wird dann als Änderungshinweis angesehen und markiert, wenn die Grauwertdifferenz einen statistisch begründeten Schwellwert überschreitet. Durch Clusterverfahren werden die binären Änderungshinweise zu Bereichen gruppiert. Die Interpretation ist meist simpel: Ist mindestens ein Bereich hinreichend groß, definiert dieser Bereich eine im Bild detektierte Änderung in der überwachten Szene und kann einer vermuteten Objektbewegung oder Beleuchtungsänderung zugeschrieben werden. Kleine, isoliert liegende Bereiche werden als rauschbedingt zurückgewiesen. Die Festlegung der Schwellwerte für die pixelbezogene Grauwertänderung und noch mehr für die Bereichsgröße der lokalen Ballungsgebiete hat sich in umfangreichen Untersuchungen (Nagel 1985) als äußerst problematisch erwiesen.

Verfeinerungen, in denen die Grauwerte lokaler Bereiche zum Beispiel durch bivariate Polynome modelliert werden und die Änderungsdetektion auf dem Vergleich dieser Beschreibung des lokalen Grauwertmodells beruht, verbessern das Detektionsergebnis in gewissem Maße. Anstelle der Intensitätswerte können Schätzwerte der Signalstärke, d. h. lokal ermittelte Grauwertvarianzen, zur Änderungsdetektion herangezogen werden. Der Schwellwert für den Kompatibilitätstest kann hier ebenfalls statistisch, als Fraktile der Fisher-Verteilung, begründet werden. Bei der polynomischen Modellierung der Bildfunktion bleibt offen, wie die Form und Größe der lokalen Umgebung eines Bildpunktes festgelegt werden sollen. Für eine zuverlässige Parameterbestimmung soll der lokale Bereich hinreichend groß sein, ohne daß allerdings die Änderungsdetektion dadurch merklich negativ beeinflusst wird.

1.4.2 Bewegungsschätzung und -deutung

Mit der Frage nach der Deutung einer möglicherweise zuverlässig detektierten Intensitätsänderung wird offenkundig, daß die Änderungsdetektionsverfahren eine nur sehr undifferenzierte Aussagekraft besitzen. Nagel (1985) listet eine Reihe möglicher Ursachen auf, die bei der Interpretation von Grauwertänderungen in Erwägung zu ziehen sind: Änderungen der Form oder der Materialeigenschaften von Oberflächen, Änderungen, die in der Beleuchtung begründet sind (Position und Stärke einer Lichtquelle, Reflexe, Schatten), Änderungen der Kamerablende oder Fokussierung, aber auch die Bewegung zwischen Sensor und abgebildetem Objekt. Die Unzulänglichkeit der Detektion von Änderungen wird beim Übergang von einer stationären Kamera auf eine bewegte Kamera besonders deutlich, bzw. allgemeiner, wenn die Änderungen durch eine Relativbewegung zwischen abgebildeter Szene und Kamera verursacht werden. Bildbereiche ohne Änderungen sind dann nur zufällig oder in Sonderfällen vorhanden.

Bewegung als Ursache einer Änderung ist für die Bildfolgenanalyse dennoch sicherlich der am häufigsten auftretende Fall. In der Regel wird entsprechendes Wissen, z. B. daß eine stationäre Szene von einer bewegten Plattform aufgezeichnet wird, meist schon durch die Aufgabenstellung impliziert. Das bedeutet aber nicht, daß andere Ursachen als die der Bewegung für eine Änderungsdetektion grundsätzlich ausgeschlossen werden können.

Als Alternative zur intensitätsbezogenen Detektion von Bewegungsänderungen haben Verri und Poggio (1987) vorgeschlagen, zuerst den sog. optischen Fluß - die visuell sichtbare Bewegung im Bild - zu bestimmen. Die Detektion von Änderungsbereichen im Bild kann über die ermittelte ortsabhängige Bewegung bzw. Verschiebung erfolgen, wobei wiederum Clusterverfahren zur entsprechenden Regionenbildung herangezogen werden können. Die Ermittlung von Verschiebung im Bild wird allgemein auch als Korrespondenzproblem bezeichnet und besteht in der Aufgabe, die Abbilder desselben Szenepunktes in zwei oder mehr Aufnahmen der Sequenz einander zuzuordnen. Mit der Bewegungsänderung ist eine Interpretation detektierter Grauwertänderungen

gefunden. Die ermittelte Bewegung solcher Bereiche besitzt potentiell wiederum eine Vielfalt weiterer Interpretationsmöglichkeiten, z. B. als 3D-Starrkörperbewegung eines Objekts. Eine Klassifizierung entsprechender Aufgaben zur Bewegungsdeutung erfolgt im nächsten Abschnitt.

1.4.3 Kriterien zur Bewegungsdeutung

Für die Klassifizierung von Aufgaben, die eine Deutung der visuell sichtbaren Bewegung in der Bildsequenz beinhalten, hat Neumann (1989) verschiedene Kriterien formuliert. Bewegungsdeutung grenzt sich von der Szeneinterpretation insbesondere dadurch ab, daß keine spezifische Objektmodellierung einbezogen wird. Das Interesse gilt dem Sensorsystem, der Aufnahmeart, dem Objekttyp und vor allem der Bewegung von Objekten und Sensor in der dreidimensionalen Welt. Die Bewegungsdeutung ist nach folgenden Kriterien zu klassifizieren, die jeweils kurz kommentiert werden sollen:

(1) *stationärer oder bewegter Beobachter*

Ein ortsfester Beobachter nimmt häufig Überwachungsaufgaben wahr oder dient der Bewegungsbestimmung von Objekten innerhalb seines Blickfeldes, wohingegen der bewegte Beobachter typisch ist für Anwendungen in der Navigation und der Robotik.

(2) *stationäre oder bewegte Szene*

Die Stationarität bzw. Bewegung von Szene und Beobachter sind oft komplementär zueinander. Die Bewegung im Bild hängt bei bewegtem Sensor stark von der räumlichen Gliederung der stationären Szene ab und ist im allgemeinen im ganzen Bild von Null verschieden. Für eine bewegte Szene typisch ist die Bewegung eines einzelnen Objektes oder mehrerer Objekte vor einem stationärem Hintergrund. Hier lassen sich Regionen im Bild segmentieren, die die bewegten Objekte vom Hintergrund trennen.

(3) *kleine oder beliebige Bewegungen*

Bei der Beurteilung der Bewegung in der Bildebene entsprechen kleine Bewegungen den dichten Bildfolgen. Die Bestimmung der Bewegung erfolgt meist durch Lösung differentieller Ansätze, die in Folge der Datendiskretisierung Verschiebungsvektorfelder ergeben. Beliebige Bewegungen korrespondieren mit losen Bildfolgen. Die Ermittlung der Bewegung erfolgt meist an ausgewählten Punkten oder Konturen, und geht oft einher mit Mehrebenenprozeduren. Bei kleiner Bewegung im Objektraum bietet sich für die 3D-Rekonstruktion von Bewegungsparametern ein differentieller Ansatz und kleinste Quadrate Minimierung an. Unter den Lösungswegen bei beliebigen Bewegungen dominieren Verfahren, die Lösungen nicht iterativ, sondern direkt ermitteln. Solche direkten Lösungen können aber auch als Verfahren zur Näherungswertbestimmung für differentielle Ansätze betrachtet werden.

(4) *Rotation, Translation, uneingeschränkte Bewegung*

Die Einschränkung der Freiheitsgrade einer Sensor- und/oder Objektbewegung kann z. B. aufgrund industrieller Rahmenbedingungen vorgegeben sein. Setzt man voraus, daß die Bewegung ausschließlich aus Rotation oder aus Translation besteht, vereinfacht dies sowohl die Modellierung des Meßprozesses der Bewegung in der Bildebene als auch die Ermittlung der Bewegungsparameter im Objektraum beträchtlich. Auch wenn ein Sensor uneingeschränkte Bewegungsfreiheitsgrade besitzt, kann wie zum Beispiel in der Fahrzeugnavigation durch Straßenmodelle (Objektmodelle) und darauf aufbauende Bewegungsmodelle Vorwissen in die Rekonstruktionsproblematik eingebracht werden.

(5) *starre Objekte, gelenkige Objekte oder verformbare Objekte*

Die Starrheitsannahme ist die am häufigsten postulierte Restriktion an die Objekte, obwohl viele Objekte wie auch der "stationäre" Hintergrund diese Forderung oft nur in grober Näherung erfüllen. Bei gelenkigen Objekten ist die Bestimmung der größeren Zahl von Bewegungsparametern aus Bildfolgen oft schon problematisch. Um die Bewegung komplexer Objekte bestimmen zu können, wie sie verformbare Gegenstände darstellen, treten anstelle der Starrheit Glattheitsannahmen bezüglich des Deformationsverhaltens, die eine vielparametrische Deformationsbewegung regularisieren und eine Schätzung erst ermöglichen. Solche allgemeinen Annahmen werden insbesondere bei der Bestimmung des optischen Flusses zugrundegelegt.

(6) *perspektive Projektion oder orthographische Projektion*

Das Standardmodell zur geometrischen Beschreibung der Abbildung der 3D-Welt auf eine 2D-Bildebene ist die perspektive Projektion. Die orthographische Projektion ist eine gute Approximation der Perspektive, wenn der betrachtete Bildausschnitt relativ zur Brennweite der Kamera sehr klein ist oder das Objekt sich in großer Entfernung von der Kamera befindet, so daß die Tiefe des Objekts gering gegenüber der Entfernung ist.

(7) *Graubilder oder Farbbilder*

Graubilder und Farbbilder konkurrieren im praktischen Einsatz, wobei dem Gewinn an Information durch Farbe eine höhere Datenmenge entgegen steht, die verarbeitet werden muß. Während Farbe für die Segmentierung und Objekterkennung ein wichtiges Merkmal sein kann, wird es für die Bewegungsschätzung als von nicht so großer Bedeutung eingeschätzt. Dennoch sind viele Kanten- und Punktefinder ebenso wie Zuordnungsalgorithmen sowohl ein- und mehrkanalig definiert. Gerade unter dem Navigationsaspekt werden aber andere Sensortypen, wie zum Beispiel entfernungsmessende Sensoren, als wichtiger betrachtet.

(8) *kurze oder lange Sequenz*

Kurze Sequenzen beschränken sind in der Regel auf zwei oder drei Bilder. Die meisten algorithmischen Lösungen zur Bewegungsschätzung in Bildfolgen sind für die Zweibildkonfiguration ausgearbeitet. Sieht man von der wenig optimalen rekursiven Verwendung von Zwei- oder Dreibildlösungsalgorithmen ab, sind bislang noch sehr wenig Vorschläge zur Analyse langer Sequenzen unterbreitet worden. Für die lange Sequenz kommt dem Gesamtsystem eine große Bedeutung zu, das die gewonnenen Informationen über die Bewegung und Struktur der Szene aus bereits erfaßten und verarbeiteten Bildern für die Analyse des aktuellen Bildes bzw. der zukünftigen Aufnahmen als Vorwissen bereithält. Dabei spielt die Modellierung des kinematischen oder dynamischen Bewegungsablaufes (z. B. Glattheit der sich zeitlich ändernden Bewegungsparameter) unmittelbar in der 3D-Welt eine wichtige Rolle.

(9) *Monobilder oder Stereobilder*

Eine mit einer Kamera erfaßte Sequenz gestattet die Bestimmung von Bewegung und Struktur in der 3D-Welt bis auf einen frei zu wählenden Maßstab. Die Maßstabsgebung ist ebenso wie die Lagerung eines festen 3D-Bezugssystems aus Bildern nicht schätzbar, wenn nicht Informationen aus der 3D-Welt, wie zum Beispiel Paßpunkte, eine bestimmte Strecke im Objektraum oder die gemessene Orientierung des Sensors in den Schätzprozeß eingebracht werden.

Mit dem Begriff der Stereobildsequenz eng verknüpft ist die Vorstellung, daß zu jedem Zeitpunkt ein Stereobildpaar mit bekannter Basislänge und bekannter gegenseitiger Orientierung aufgenommen wird. Da eine bekannte Länge genügt, um den Maßstabsdefekt zu beseitigen, gestattet schon ein Stereopaar die 3D-Rekonstruktion z. B. der Entfernung von Punkten. Die Stereosequenz erlaubt eine geometrisch sehr viel besser kontrollierte Analyse als die Monosequenz, impliziert aber wie schon die Farbsequenz einen höheren Erfassungs- und Verarbeitungsaufwand.

Anhand dieser Kriterien fällt es nun leicht, die vorliegende Arbeit einzuordnen. Das primäre Interesse gilt einem navigierenden Sensor, der sich in einer im wesentlichen stationären aber weitgehend unbekanntem Umgebung bewegt. Zwischen zwei Aufnahmezeitpunkten sollen größere Bewegungen zugelassen werden. Die Sensorplattform kann sich rotatorisch und translatorisch frei bewegen. Der Bewegungsmessung und 3D-Rekonstruktion sollen eine perspektive Projektion und Graubilder zugrundegelegt werden. Der Schwerpunkt der Arbeit konzentriert sich dann auf die Analyse von langen Monobildsequenzen.

Schwerpunkte der Analyse sind die Messung von Bewegung in der Bildsequenz, die Rekonstruktion von Bewegung und Struktur in 3D und die Fokussierung auf Anwendungen in der Navigation. Sie werden in den nächsten Abschnitten weiter vertieft.

1.5 Bewegungsmessung in Bildsequenzen

Die Messung von Bewegung in der Bildebene steht synonym für die Bildzuordnung in der Bildfolgenanalyse. Allgemein umfaßt Bildzuordnung die Aufgabe, die korrespondierenden Größen zweier Bilddatensätze zu bestimm-

men. Beinhalten die Datensätze Größen, die das Bild und ein Objektmodell beschreiben, zielt die Bild-Modell-Zuordnung auf die Identifikation oder Lokalisierung des Objekts. Die fortgesetzte Lokalisierung eines bekannten Objekts im jeweils aktuell erfaßten Bild einer Sequenz wird auch als Objektverfolgung bezeichnet. Sie führt zur Bewegungsbestimmung, die sowohl im Bildraum als auch direkt im Objektraum erfolgen kann. Der zweite, weit intensiver erforschte Bereich der Bildzuordnung ist die Bild-zu-Bild-Zuordnung. Die Messung der in der Bildfolge visuell sichtbaren Bewegung ist die typische Meßaufgabe bei Bildsequenzen. Die Aufgabe besteht darin, explizit oder implizit korrespondierende Größen zweier Bilddatensätze zu bestimmen. Hinter dieser Formulierung des Korrespondenzproblems verbirgt sich eine nichttriviale Interpretationsaufgabe: Es gilt Größen in beiden Bildern zu identifizieren, die von einem Objektelement herrühren. Implizite Korrespondenz ist typisch für Ansätze bei dichten Bildsequenzen. Der zu bestimmenden Geschwindigkeit eines Helligkeitsmusters in der Bildebene entspricht diskret eine Verschiebungsvektor eines Bildpunktes, der auf die Lage des korrespondierenden Punktes im nächsten Bild zeigt. Bei expliziter Korrespondenz wird hingegen zuerst die Lage des oft durch Merkmale beschriebenen Objektelements in beiden Bildern bestimmt. Bei erfolgreicher Zuordnung der Merkmale definiert die Lageänderung die Bewegung des Objektelements in der Bildebene.

1.5.1 Kriterien zur Klassifizierung von Zuordnungsverfahren

Die Klassifizierung von Bildzuordnungsverfahren kann nach verschiedenen Gesichtspunkten erfolgen, die hier kurz erwähnt werden sollen. Sehr verbreitet ist die Klassifizierung nach dem *Suchverfahren*. Beispiele sind hier Baumsuchverfahren, bei denen der Suchraum auf einen Graphen abgebildet wird, das simulierte Abkühlen (simulated annealing), ein auf einer Monte-Carlo-Methode beruhendes Optimierungsverfahren, und die generalisierte Hough-Transformation, die sich insbesondere durch die Ermittlung eines Optimums im aufgerasterten Lösungsraum auszeichnet. Weit verbreitet sich auch kleinste Quadrate Minimierungsverfahren zur Bestimmung der unbekanntenen Transformationsparameter zwischen Datensätzen und das relaxation labeling, bei dem über die Nachbarschaftsbeziehungen der Daten einer Beschreibung untereinander in einem iterativen Prozeß eine eindeutige Abbildung angestrebt wird. Ein anderes Klassifizierungskriterium ist das *Bewertungsmaß*, insbesondere die Bewertung der Zuordnung durch Kostenfunktionen oder durch Distanzmaße, oder auch durch das klassische Ähnlichkeitsmaß, den Korrelationskoeffizienten. Eine eher an der Praxis orientierte Klassifizierung ist die nach der *Anwendung*, deren Vielfalt im Bereich der Bildsequenzen bereits im Abschnitt 1.4.2 angedeutet wurde. Die Klassifizierung nach der *Kontrollstrategie* wird zunehmend bedeutsamer, da ein Gesamtsystem zur Bewegungsmessung insbesondere nach seiner Zuverlässigkeit beurteilt wird. Hier sind vor allem auf Bildpyramiden aufbauende Mehrebenenprozeduren, in der Bildsequenz aber auch Filterverfahren zur Verfolgung von Punkten oder anderen Bildstrukturen von Bedeutung. Kontrolle auf wissensbasiertem Niveau steuert und überwacht die Prozesse (Gruppierung, Klassifizierung) bis hin zur Interpretation. Pyramidales Vorgehen kann entfallen, wenn insbesondere bei dichten Sequenzen nur sehr kleine Verschiebungen auftreten. Ein letztes Kriterium für eine Klassifizierung der Zuordnungsverfahren ist die Einteilung nach dem *Abstraktionsgrad der Bildbeschreibung* in intensitätsbasierte, merkmalsbasierte und relationale Verfahren. Intensitätsbasierte Verfahren nützen die Information der sich änderenden Grauwerte direkt, während merkmalsbasierte Verfahren die Änderung der Position der aus den Bildern extrahierten Merkmalen analysieren. Die zusätzliche Extraktion und Bewertung von Relationen hat insbesondere für die Objekterkennung und Lokalisierung eine große Bedeutung; relationale Verfahren nehmen aber in der Bildsequenzanalyse bislang eine untergeordnete Rolle ein. Intensitätsbasierte und merkmalsbasierte Ansätze sollen in den nächsten beiden Abschnitten noch näher betrachtet werden. Sie entsprechen grob der in der Bildsequenzanalyse auch üblichen Unterteilung der Verfahren in kontinuierliche und diskrete Ansätze.

1.5.2 Intensitätsbasierte Verfahren

Intensitätsbasierte Verfahren zeichnen sich dadurch aus, daß die Formulierung des Modells zur Bewegungsbestimmung in der Bildebene direkt auf den Bildintensitäten bzw. -gradienten beruht. Die Gruppe der kontinuierlichen Verfahren geht von einer kontinuierlichen Beschreibung der Bildintensitäten in der Bildebene und auch in der temporalen Dimension aus. Das Ziel ist die Bestimmung des optischen Flusses, bzw. bei der diskreten Realisierung, ein am Bildpunktraster orientiertes dichtes Verschiebungsvektorfeld. Aus der Annahme, daß sich

die Intensität $I(x, y, t)$ eines abgebildeten Punktes innerhalb eines kurzen Zeitintervalls nicht ändert, folgt durch Linearisierung die sogenannte Restriktionsgleichung des optischen Flusses $I_x u + I_y v + I_t = 0$ (3-19). Verfahren, die auf dieser Restriktion beruhen, werden auch als gradientenbasierte Verfahren bezeichnet, da die linearisierte Bildfunktion über den räumlichen und temporalen Bildgradienten in dieser Bedingung auftritt. Diskret betrachtet kommen auf eine Restriktion zwei unbekannte Komponenten des Geschwindigkeitsvektors pro Bildpunkt. Dies impliziert, daß ohne weitere Annahme der optische Fluß nicht bestimmt werden kann. Deshalb sind zusätzliche Glattheitsbedingungen vorgeschlagen worden, die im Kapitel 3.2 ausführlich dargestellt werden.

Die Unbestimmtheit bei der Schätzung der Bildbewegung existiert nicht nur bei punktbezogener Betrachtung der Restriktionsgleichung des optischen Flusses, sondern auch in allgemeinerer Form und wird als Apertur- oder Blendenproblem bezeichnet. Die Problematik ist in Abbildung 1.2 verdeutlicht. Auf der linken Seite sind die

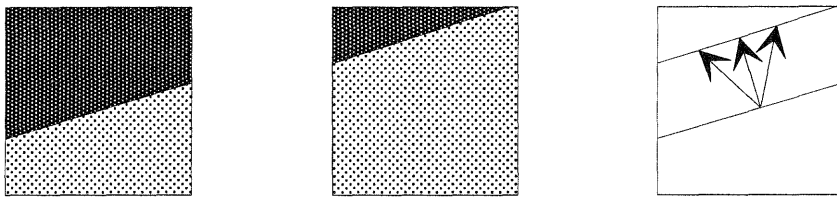


Abbildung 1.2: Unbestimmtheit der Bewegung an einer Kante

lokalen Bildauschnitte einer Kante in zwei aufeinanderfolgenden Bildern dargestellt, so wie sie bei der Betrachtung des Bildes durch eine Blende gesehen werden. Entlang der Kante gibt es keine Unterscheidungsmerkmale der Punkte, weshalb die Verschiebung der Kantenpunkte unterbestimmt ist. Eindeutig festgelegt ist nur der Verschiebungsbetrag jeweils senkrecht zur Kantenrichtung.

Eine zweite Gruppe bilden intensitätsbasierte Verfahren, die von diskreten Bildintensitäten ausgehen. Obgleich meist bezüglich der Radiometrie allgemeinere Transformationen als bei der optischen Flußbedingung einbezogen werden, sind die Ansätze dadurch wesentlich restriktiver, daß die Transformationen (radiometrisch und geometrisch) mit wenigen, festen Parametern für einen ganzen Bildausschnitt beschrieben werden. Ein Beispiel hierfür ist die klassische Korrelation, die fordert, daß innerhalb eines Korrelationsfensters das Bild sich um einen festen Betrag translatorisch bewegt.

Die Restriktionen für eine möglichst eindeutige Bestimmung von Bewegung in der Bildsequenz sind in der Regel heuristischen Ursprungs. In der Literatur sind mehrere verschiedenartige Heuristiken für die intensitätsbasierten Verfahren diskutiert (Hannah, 1974, Koenderink und van Doorn, 1976, Horn und Schunk, 1981, Hoff und Ahuja, 1986):

- H1: Kleine oder gar differentiell kleine Verschiebung der Bildpunkte
- H2: Konstante Intensität $I(x, y, t)$ innerhalb eines kurzen Zeitintervalls
- H3: Isotrope oder gerichtete Glattheit der Bildbewegung (Kontinuität des optischen Flußfeldes) in großen Teilen des Bildes
- H4: Verschiebungsfelder können analytisch, meist durch eine Polynom, beschrieben werden
- H5: Das Gradientenfeld des optischen Flusses überschreitet eine feste Obergrenze nicht

Die meisten der vorgeschlagenen Verfahren stützen sich auf diese Heuristiken, wobei in der Regel mehrere simultan einbezogen werden. Die Heuristiken sind in der Absicht formuliert, eine möglichst widerspruchsfreie, der Realität weitgehend angepaßte Lösung zu erzielen. Sie schränken den Suchraum für die Zuordnung beträchtlich ein und implizieren oft auch ein Modell für die Bewegung im Bild.

1.5.3 Merkmalsbasierte Verfahren

Merkmalsbasierte Verfahren werden in der Bildsequenzanalyse auch als diskrete Verfahren bezeichnet. Diskret sind sowohl die Zeitpunkte, zu denen die Bilder erfaßt werden, als auch die Bildprimitive oder Merkmale, die aus den Bilddaten extrahiert werden. Die Schätzung von Bewegung in der Bildfolge erfolgt in zwei Schritten: In einem ersten Schritt wird markante Bildinformation extrahiert. Insbesondere sind dies markante Punkte, beispielsweise Grauwertecken oder helle oder dunkle Flecken, aber auch Kantensegmente, gerade Linien und geschlossene Konturen, z. B. die zero crossings eines LoG-gefilterten Bildes (Abschnitt 3.2.4). Auch komplexere Strukturen sind in einzelnen Arbeiten verwendet worden (Gu et al., 1987). Der zweite Schritt des merkmalsbasierten Meßprozesses besteht in der Korrespondenz. Durch Such- oder Relaxationsverfahren werden korrespondierende Primitive ermittelt, wobei die Differenz in der Lage und eventuell anderer Attribute der Punkt-, Linien- oder Regionenprimitive die Bewegung zwischen Bildern angibt. Der Transfer zur symbolischen Bildrepräsentation durch Primitive ist ein Datenreduktionsprozeß. Deshalb ist auch die Dichte der aus der Korrespondenz resultierenden Verschiebungsvektoren entsprechend gering. Soll dennoch ein dichtes Verschiebungsvektorfeld ermittelt werden, wie es typischerweise von den Verfahren zur Bestimmung des optischen Flusses erhalten wird, können hierfür entsprechende Oberflächeninterpolationsverfahren (siehe auch Abschnitt 3.1) verwendet werden.

Die Heuristiken für die merkmalsbasierten Verfahren sind mit ähnlicher Intention wie die der intensitätsbasierten Verfahren formuliert worden. Die Vorschläge beinhalten (Ballard und Brown, 1982, Marr und Poggio, 1979, Mayhew und Frisby, 1981, Koller und Nagel, 1990):

- H1: Verschiebungsschätzung nur an ausgewählten Primitiven (Merkmalen), Stabilität der Primitive in der Sequenz
- H2: Ähnlichkeit der Merkmale auch bei größerem Zeitintervall, eventuell auch strukturelle Konsistenz (z. B. Endpunkte von Kanten sollen entsprechend korrespondieren)
- H3: Glattheit der Verschiebungsvektoren entlang von Konturen
- H4: Kontinuität von Verschiebungsvektoren innerhalb von Regionen und über mehrere Bilder der Sequenz
- H5: Feste Obergrenze für die Verschiebungsvektoren, aber auch Mindestverschiebung der Merkmale

Die Heuristiken H1 - H5 der intensitätsbasierten und merkmalsbasierten Verfahren sind hier so geordnet, daß sie im direkten Vergleich betrachtet werden können. Die Verschiedenartigkeit spiegelt insbesondere wieder, daß die merkmalsbasierten Heuristiken für eine eher geringe Menge allerdings bedeutungsvoller Beobachtungen konzipiert sind, die intensitätsbasierten Heuristiken sich dagegen an der Bestimmung eines dichten Verschiebungsvektorfeldes bei kleiner Bewegung in der Bildebene orientieren.

1.6 3D-Rekonstruktion von Bewegung und Struktur

In den vorangehenden Abschnitten wurde schon ausgeführt, daß sich die Bestimmung von Bewegung und Struktur aus heutiger Sicht als zweistufiger Prozeß darstellt. Die implizit in der Bildfolge enthaltenen Informationen werden durch die Bewegungsmessung explizit erfaßt und in einem zweiten Schritt durch Inversion der 3D - 2D Projektion mittels objektraumspezifischer Größen - Bewegung von Objekt bzw. Sensor und Struktur der Szene - erklärt. Die 3D-Rekonstruktion von Bewegung und Struktur ist in erster Linie ein analytisches Problem, das bei gegebener Modellierung durch Parameterschätzung gelöst werden kann. Allerdings sind die Beziehungen zwischen 2D-Beobachtungen und 3D-Unbekannten oft nichtlinear, so daß für die Parameterschätzung in linearisierten Modellen Näherungswerte benötigt werden oder zur Lösung ein lineares Ersatzproblem gefunden werden muß. Lineare Ersatzprobleme, aber auch andere Lösungswege, z. B. auf der Grundlage polynomischer Ansätze, sind meist schwierig abzuleiten und fast immer nur für eine sehr spezielle Aufgabenstellung vorgeschlagen worden.

Für die Rekonstruktion von *Struktur* oder *Form* gibt es eine Vielzahl von Ansätzen, die unter den Sammelbegriffen *structure from X* oder *shape from X* bekannt sind. *Struktur* oder *Form* stehen für die im Objektraum

spezifizierte Parametrisierung eines Modells der abgebildeten Szene bzw. deren Objekte. In der einfachsten Form kann ein Objekt durch eine Ebene, eine Linie und auch nur einen Punkt modelliert werden. Komplexere Beispiele sind quadratische Flächen, Zylinder oder Polyeder.

Eine Alternative zu diesen Inversions- oder Rückprojektionsansätzen sind die Projektionsansätze, die von einer grob vorliegenden Beschreibung der Szene, z. B. mittels parametrisierter Objektmodelle, ausgehen. Die Projektion des parametrisierten Objektmodells führt zu einer unendlich großen Zahl von Modellbildern, praktisch meist zu projizierten objektspezifischen Größen verbunden mit Intervallen, die die Unsicherheit der projizierten Merkmale, z. B. der Position, Orientierung oder auch Ausdehnung, angeben. Die detailliertere, präzisere Beschreibung der Szene ergibt sich aus der Suche nach dem zum Bild am besten passenden Modellbild.

1.6.1 Inversionsansätze

Inversionsansätze sind dadurch gekennzeichnet, daß sie von den beobachteten Daten in der Bildebene ausgehen und durch die Beziehungen zu Größen des 3D-Objektraumes Information in die 3D-Szene zurückprojizieren. Beispielsweise wird ein Punkt in der Bildebene in einen Punkt zurückprojiziert, der sich auf der Geraden durch das Projektionszentrum und den Bildpunkt befindet. Die unbekannte Entfernung des Punktes auf dieser Geraden ist eine Möglichkeit zur Parametrisierung der Struktur. Das Beispiel macht deutlich, daß die Rückprojektion eine unendliche Vielfalt an Kandidaten bezüglich der 3D-Struktur erzeugt. Restriktionen, wie die Starrheit des Objekts oder bekannte Längen oder Winkel, beschränken die Vielfalt der 3D-Kandidaten direkt im Objekt-raum. Die Inversionsansätze werden von Kanatani (1990) als 3D euklidische Verfahren bezeichnet, da alle zu rekonstruierende Größen im euklidischen 3D-Raum parametrisiert sind. Der Schwerpunkt der Forschung liegt bis heute in der Entwicklung bzw. Verbesserung entsprechender Verfahren. Sie können danach unterscheiden werden, ob Monobilder, Stereobilder oder Bildfolgen verarbeitet werden.

Viele Verfahren haben das Leistungsvermögen des menschlichen visuellen Systems zum Vorbild. Eine qualitative Beschreibung der Szenengeometrie gelingt dem Menschen schon bei monoskopischer Betrachtung eines Bildes äußerst erfolgreich. Das *shape from shading* rekonstruiert die 3D-Oberflächenform (Oberflächenorientierung) aus dem Intensitätsbild, wobei die Reflexionseigenschaften der Oberfläche und deren Ausrichtung zur Lichtquelle und Beobachter bekannt sein müssen. Zur Regularisierung der unterbestimmten Aufgabe werden Glattheitsrestriktionen bezüglich der Krümmung der Oberfläche eingeführt (Horn, 1977). *Shape from texture* setzt eine gleichförmige Texturierung der betrachteten Oberfläche voraus. In der Projektion ist die Regelmäßigkeit gestört bzw. verzerrt. Art und Größe der Verzerrung geben direkt Hinweise auf die Form und Orientierung der Oberfläche. Mit der statistischen Charakterisierung der Textur durch Isotropie oder Homogenität hat Witkin (1981) eine sehr allgemeine Lösung für die Rekonstruktion aufgrund gegebener Textureigenschaften vorgeschlagen. Mehr spezifische Informationen über die Strukturierung einer Oberflächentextur wie z. B. Parallelität, Orthogonalität, Symmetrie, und anderes führen zu noch sehr viel mehr Rekonstruktionsansätzen aus Monobildern, wie z. B. *shape from contour*, *shape from angle*, *shape from line drawings* (siehe Aloimonos, 1988). Gemeinsam haben diese Verfahren, daß die gewonnene formbezogene Information über das Objekt in der 3D-Orientierung von Kanten oder Oberflächen besteht.

Die Grundidee hinter den *shape from stereo* Ansätzen korrespondiert mit der Fähigkeit des Menschen zum Stereosehen aufgrund der räumlichen Trennung der Augen (Stereoparallaxe). Julesz (1971) hat mit dem random-dot stereogram (Abbildung 1.3) in beeindruckender Weise nachgewiesen, daß Stereotiefenwahrnehmung ohne Interpretation des Einzelbildes (bekannte Formen, Objekte, u. a.) ausschließlich aufgrund der Stereoparallaxe erfolgen kann. Die Stereorekonstruktion aus gegebenen Disparitäten bei bekannter Orientierung der Kameras erfolgt durch räumlichen Vorwärtsschnitt, in dem die 3D-Position von Punkten der Szene bestimmt wird. Zur Ableitung einer regelmäßigen Formbeschreibung durch dichte Tiefenkarten oder Oberflächenmodelle schließt sich oft noch eine Oberflächeninterpolation an.

Die Ansätze für die Rekonstruktion von Bewegung und Struktur bei gegebener Bewegung in der Bildfolge werden als *structure from motion* oder teilweise auch als *shape from motion* bezeichnet. Psychologische Studien von Wallach und O'Connell (Ullman, 1987) anhand der Schattenbilder projizierter 3D-Objekte zeigen, daß es den Versuchspersonen in den meisten Fällen gelingt, sowohl die 3D-Form eines Objekts, als auch dessen



Abbildung 1.3: Random-dot stereogram. (Ein räumlicher Eindruck entsteht z. B. bei beidäugiger Betrachtung mit einem kleinen Stereoskop.)

Bewegung im Raum richtig zu beschreiben. Das bedeutet, das Erkennen von 3D-Form und 3D-Bewegung eines bewegten Objekts ist dem Menschen selbst bei nicht näher identifizierbaren Objekten möglich.

Da *structure from motion* das für Bildfolgen typische Rekonstruktionsverfahren ist (*shape from stereo* ist darin oft eingeschlossen), werden die Ansätze hierzu ausführlich im Abschnitt 2.4 besprochen. Voraussetzung für die Rekonstruktion ist, daß mindestens zwei Bilder einer Bildsequenz vorliegen. Viele Verfahren sind für kurze Sequenzen von zwei oder drei Bildern vorgeschlagen worden. Zum Beispiel bestimmen Bruss und Horn (1983) Bewegung und Tiefe aus dem optischen Fluß, wobei sie diskret das gesamte Verschiebungsvektorfeld in eine kleinste Quadrate Minimierung einbeziehen. Ein Ansatz zur Bestimmung von Bewegung und Struktur (Tiefe) direkt auf der ikonischen Bilddatenebene unter Verwendung räumlicher und zeitlicher Bildgradienten wurde von Negahdaripour und Horn (1987) vorgestellt. Merkmalsbasierte Ansätze beruhen auf der durch die Korrespondenz gegebenen Position (und eventuell auch anderer Größen) von Primitiven. Die Ermittlung der minimalen Anzahl korrespondierender Primitive und die Entwicklung direkter Lösungen sind wichtige Fragestellungen dieser Verfahren. So hat beispielsweise Ullman (1979) nachgewiesen, daß die Bewegungs- und Strukturparameter bestimmbar sind, wenn die Punktkorrespondenz von 4 Punkten in 3 orthographisch projizierten Bildern vorliegt.

Die Arbeiten in langen Sequenzen können nach Mono- und Stereosequenzen unterschieden werden (Matthies et al. 1988, Weng et al., 1987). Die Schwerpunkte liegen in der algorithmischen Entwicklung rekursiver Rekonstruktionsverfahren, der Integration der bereits aus vorhergehenden Aufnahmezeitpunkten extrahierten Information über die Szene, und im Systementwurf, der die Verquickung von Zuordnung und Rekonstruktion aufzeigt. In nahezu allen Arbeiten beruht die Rekonstruktion aus Bildfolgen auf dem geschätzten optischen Fluß bzw. bei den merkmalsbasierten Verfahren auf den punktbezogenen Bewegungsvektoren. Diese Daten, aber auch die Bildintensitäten direkt oder andere extrahierte Größen, wie die Parameter von Objektkonturen sind die Observablen (Kanatani, 1990), die die bildbezogene Information für die Rekonstruktion darstellen. Auf die Diskussion der verschiedenen Rekonstruktionsansätze in Abschnitt 2.4 sei verwiesen.

1.6.2 Projektionsansätze

Projektionsansätze sind dadurch gekennzeichnet, daß sie von einer vorliegenden groben Beschreibung der Szene ausgehen und diese dann aufgrund der Bilddaten präzisieren. Dieses Aufstellen einer geeigneten Modellierung der Szene steht an erster Stelle. Durch die Projektion der modellierten Größen auf die Bildebene entsteht eine Vielzahl von Modellbildern mit definierten Merkmalen, die dann entsprechend im Bild extrahiert werden müssen, um sie vergleichen zu können. Die Projektionsansätze zeichnen eine szenenbezogene Modellierung aus, die projiziert im Bild durch eine Bild-Modellbild-Zuordnung die eigentliche Rekonstruktion erwirkt. Die Projektionsansätze werden von Kanatani (1990) als 3D nicht-euklidische Verfahren bezeichnet. Der Hintergrund

dieser Bezeichnungsweise soll an einem Beispiel verdeutlicht werden. Parallelität von Linien in 3D geht bei der Projektion in die Bildebene verloren, wenn eine euklidische Metrik in der Bildebene zugrundegelegt wird. Definiert man aber eine 2D nicht-euklidische Metrik in der Bildebene, kann eine geeignete Modellierung des Objekts zugrunde gelegt werden, mit der Eigenschaft, daß Parallelität in der Bildebene eine entsprechende Parallelität in der Szene reflektiert. Auch andere Invarianten, beispielsweise die rotationsinvariante Parametrisierung des optischen Flusses eines ebenen Objekts werden in Kanatani diskutiert.

Obwohl viele Ansätze die Idee invarianter Merkmale für die Rekonstruktion nicht streng befolgen, sind sie den Projektionsansätzen zuzurechnen. Ein bekannter Vertreter ist das System ACRONYM (Brooks, 1981). Es beschränkt sich auf die Analyse eines Einzelbildes und zielt auf die Bestimmung der Position und Orientierung von Objekten in der Szene. Vorgeben werden vom Anwender des Systems Klassen von dreidimensionalen Körpern. Merkmale, wie die Anzahl der sichtbaren Kanten im Bild werden verwendet, um grobe Schätzungen der Position und Orientierung der Objekte vorzunehmen. Eine wesentliche Charakteristik des Systems ist das Modul zur Verwaltung der Restriktionen, die die 3D-2D Projektion der Merkmale beschreiben. Aufgrund der Zuordnung der projizierten Merkmale zu den im Bild extrahierten Merkmalen erfolgt dann die Präzisierung der Beschreibung der Szene. Eine Erweiterung dieses Konzepts eines Projektionsansatzes für Anwendungen in bewegten Szenen wird in Pauli (1992) vorgestellt, der seinen Ansatz als erklärungs-basierte Szenerekonstruktion bezeichnet. In Pauli (1992) werden auch weitere Arbeiten aus dem Bereich der Projektionsansätze diskutiert.

Heuristiken und Modelle

Wie schon die Bildzuordnungsverfahren beruhen auch die 3D-Rekonstruktionsverfahren mit den Inversions- und Projektionsansätzen auf Heuristiken bzw. Vorinformationen über die dreidimensionale Welt. Die am meisten postulierte Annahme ist das Starrkörpermodell, das für die aufnehmende Kamera und - bei der Schätzung der Eigenbewegung einer Kamera - auch für die Szene zugrundegelegt wird. Die Kamerabewegung kann dann durch drei Translationsfreiheitsgrade und drei Rotationsfreiheitsgrade beschrieben werden. Dasselbe gilt für die Modellierung der 3D-Bewegung eines starren Objektes bezüglich der fest installierten Kamera und darüber hinaus für Relativbewegungen von Objekt und Kamera. Heuristisch motiviert sind spezifischere Annahmen bezüglich der Bewegungsform, insbesondere die Glattheit der Bewegungsparameter. Die Annahme bekannter Objektmodelle führt in der Regel zu einer beträchtlichen Reduktion der Anzahl der zu rekonstruierenden Strukturparameter. Beispiele für Objektmodelle und deren Parametrisierung durch Strukturparameter sind: eine Linie, repräsentiert durch einen Punkt und einen Richtungsvektor, eine Ebene, repräsentiert durch einen Punkt und den Normalenvektor der Ebene, ein Kugel, parametrisiert durch den Mittelpunkt und den Radius, aber auch Zylinder, Kegel, Vielflächner, und vieles andere. In der Praxis ist der Punkt und seine 3D-Koordinate als elementarstes Strukturprimitiv weit verbreitet.

1.7 *Passive Navigation*

Die Diskussion von Bewegungsdeutungs-aufgaben im Abschnitt 1.4 hat viele Möglichkeiten und Randbedingungen der Bildsequenzanalyse aufgezeigt. Mit der Ausrichtung auf die passive Navigation wird diese Vielfalt beschränkt und die Zielrichtung dieser Arbeit präzisiert. Hieraus ergeben sich sowohl für die Bildzuordnung als auch die 3D-Rekonstruktion entsprechende Konsequenzen.

Der Begriff der passiven Navigation wird in der Literatur nicht einheitlich definiert. Im Bereich des maschinellen Sehens haben Bruss und Horn (1983) damit die Aufgabe identifiziert, die Bewegung einer Kamera aus dem vorab bestimmten optischen Fluß zu rekonstruieren. Vorausgesetzt wurde dabei, daß die Kamera sich in einer stationären Umgebung bewegt. Überraschenderweise fällt quasi als Nebenprodukt der 3D-Bewegungsrekonstruktion die Form der abgebildeten Objektoberfläche mit an (Horn, 1986, S. 400), vorausgesetzt die Bewegung beinhaltet eine Translation.

Navigation beschäftigt sich mit der Führung von Objekten vom aktuellen Standpunkt zu einem Zielort, wobei Aspekte wie die Sicherheit, eine kostengünstige Lösung, und andere, eine große Rolle spielen. Die Sensorik und Meßtechnik sind dabei meist von zentraler Bedeutung: Es existiert eine Vielfalt von Meßgeräten und Auswertetechniken, um die eigene Position und die des Zielortes als Funktion der Zeit zu bestimmen. Typische Aufgaben

der Navigation sind die Positions- und Richtungsbestimmung, die Geschwindigkeitsmessung, die Zeitbestimmung, aber auch speziellere Fragestellungen, wie die Kollisionsvermeidung oder die Routenplanung. Auf dieser Grundlage definiert Hartl (1989) Navigation gemäß: "Navigation is the continuous updating of such information".

1.7.1 Abgrenzung passiver von aktiver visueller Navigation

Die Navigation im Bereich des maschinellen Sehens wird auch als visuelle Navigation bezeichnet. Typische Sensoren sind hier flächenhafte Sensoren, die Graubilder und Distanzbilder in Form von Sequenzen erfassen können.

Die Abgrenzung der passiven von der aktiven Navigation kann nach mindestens drei Gesichtspunkten erfolgen. Kennzeichnend für die passive Navigation ist der Verzicht auf

- eine aktive Steuerung der bewegten Plattform

Die autonome Fahrzeugführung setzt eine Steuerungseinrichtung voraus, um aktiv Bahn- und Kurskorrekturen vornehmen zu können. Hier steht eine regelungstechnische Aufgabe im Mittelpunkt, insbesondere die Ermittlung von Steuerungsgrößen aufgrund der Bewegungsschätzung. Die passive Navigation beschränkt sich auf die Bewegungsbestimmung und gibt diese z. B. an ein Modul weiter, das eine Überwachungsaufgabe wahrnimmt.

- aktive, die Szene abtastende Sensoren

Unter den aktiven flächenhaften Sensoren sind für Navigationszwecke vor allem scannende Laser- und Radarentfernungsmesser im Einsatz. Bezüglich der Echtzeitfähigkeit einer flächenhaften Bilderfassung sind sie den passiv Graubilder oder Farbbilder erfassenden CCD-Sensoren unterlegen.

- aktive, ein Detail der Szene verfolgende Sensoren

Die auch als aktives Sehen bezeichnete Technik beruht auf dem für den Menschen typischen Verhalten, daß er fortwährend die Augen auf ein Objekt ausrichtet, das seine momentane Aufmerksamkeit hervorruft. Entsprechend sollen aktive Kameras ständig auf ein Detail der Szene ausgerichtet sein, was durch ein Zuordnungsmodul mit entsprechender Nachführung der Kameras realisiert werden kann.

Aufgaben eines visuellen Navigationssystems sind z. B. das Verfolgen einer Straße, oder allgemeiner das Navigieren in einer bekannten oder unbekanntem Umgebung, die Planung der Route zu einem Zielort, oder die Kollisionsvermeidung bei Hindernissen, die entlang der Bahn lokalisiert wurden. Zur Lösung solcher Aufgaben ist meist eine sehr hohe Aufdatierungsrate für die zu bestimmenden Größen wie die Position, Richtung, Geschwindigkeit, Distanz usw. notwendig.

1.7.2 Rekursive kinematische Modellierung

Als Konsequenz aus der Forderung nach einer möglichst kontinuierlichen Bestimmung navigatorischer Information wird von der algorithmischen Auswertekomponente eines Navigationssystems erwartet, daß entsprechende Schätzprozesse jeweils aktuelle Daten liefern, was meist mit rekursiven (oder sequentiellen, inkrementellen) Formulierungen einhergeht.

Zur Bestimmung navigatorischer Daten bei kinematischen oder dynamischen Positionieraufgaben werden üblicherweise Kalman-Filter oder ähnliche sequentielle Schätztechniken benutzt. Die zu schätzenden Parameter sind die Zustandsgrößen des zeitlich variierenden Systems. Abhängig von der Anwendung umfaßt der Zustandsvektor Bewegungsparameter ebenso wie typisch statische Größen, z. B. die Geometrie einer stationären Szene in einem szenefesten Objektkoordinatensystem. Die Formulierung eines solchen Filters beinhaltet zwei Modellierungskomponenten:

1. ein kinematisches (dynamisches) Modell

Dieses Modell wird vorwiegend zeitdiskret benötigt und beschreibt die Beziehung der Zustandsgrößen zwischen zwei aufeinanderfolgenden Zeitpunkten (Epochen). Darin enthalten sind die Bahnmodellierung eines bewegten Sensors und bei gesteuerten Systemen auch Steuerungskorrekturen. Mit Hilfe dieses Systemmodells erfolgt die Prädiktion der Zustandsgrößen, wobei der Zustand zum Startzeitpunkt sowie verschiedene stochastische Informationen bekannt sein müssen.

2. ein Beobachtungsmodell

Mit der Beziehung zwischen Beobachtungsdaten und Zustandsgrößen erfolgt die Aufdatierung oder Korrektur der prädizierten Zustandsgrößen zum aktuellen Zeitpunkt. Auch hier wird die Kenntnis der stochastischen Eigenschaften der Beobachtungen vorausgesetzt.

Die Prinzipskizze eines passiv navigierenden Sensors soll den rekursiven Prozeß erläutern (Abbildung 1.4). Die Rekursion beschreibt alle wesentlichen Komponenten eines Systems: Die Prädiktion der Zustandsgrößen stellt die bereits bis zum Zeitpunkt t_{i-1} rekonstruierten Informationen über die Bewegung und Struktur zum neuen

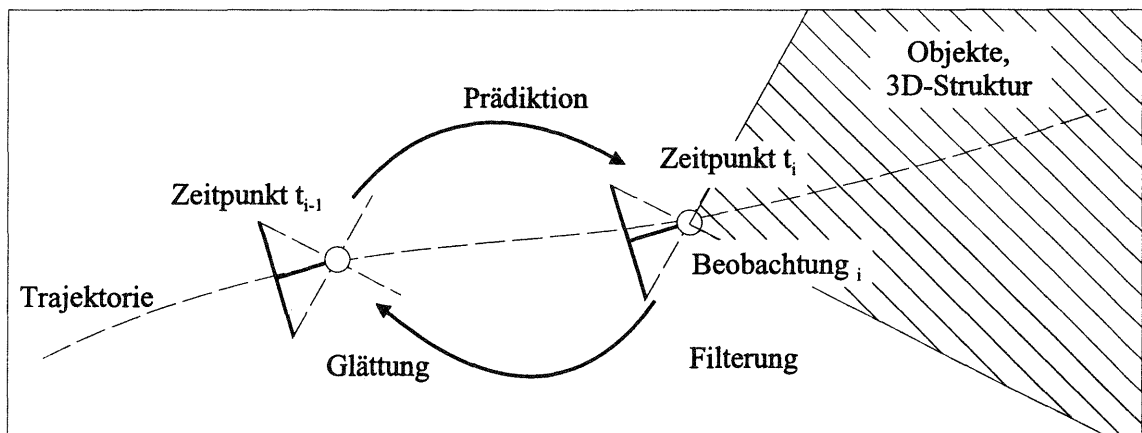


Abbildung 1.4: Prinzipskizze eines passiv navigierenden Sensors

Zeitpunkt t_i bereit. Ebenfalls prädiziert wird mithilfe des Systemmodells und seiner stochastischen Beschreibung die Fehlerkovarianz des Zustandsvektors. Abhängig von der Komplexität der Strukturparameter liegt im Sinne der Projektionsansätze (Abschnitt 1.6) eine mehr oder minder komplexe Modellierung der Szenegeometrie vor, auch wenn dabei die Modellierung meist nicht der Idee der invarianten Merkmale folgt. Die nächste Systemkomponente besteht in der Beobachtung zum aktuellen Bezugszeitpunkt t_i . Die in den prädizierten Zustandsgrößen repräsentierten Objektstrukturen können ins Bild projiziert werden und bilden damit Vorinformation für den Meßprozeß der Bewegung in der Bildebene. Die Extraktion entsprechender Primitive und die Korrespondenz liefern die Meßdaten. Als weiterer Schritt schließt sich die Aufdatierung des prädizierten Zustandes an, wobei auch hier die stochastische Beschreibung durch Kovarianzen aufdatiert wird. Die Aufdatierung kann im Sinne eines Inversionsansatzes interpretiert werden. Die Beziehung der gemessenen Größen zu den im euklidischen 3D-Raum parametrisierten Bewegungs- und Strukturparametern wird in der Aufdatierungsphase direkt genutzt, um die Bestimmung des Bewegungszustandes zu aktualisieren und die rekonstruierten Objektstrukturen zu präzisieren. Diese Aufdatierung ist ein Filterprozeß, der in einem weiteren Schritt auch zu einer verstärkten Glättung der Zustandsgrößen der Bezugszeitpunkte ($t < t_i$) ausgebaut werden kann. Die Systemkomponenten Prädiktion, Beobachtung, Aufdatierung und evtl. Glättung sind die vier charakteristischen Verarbeitungsschritte, die innerhalb einer Rekursion bei der Kalman-Filterung anfallen. Offene Fragestellungen, beispielsweise wie die strukturspezifischen Zustandsgrößen bei verschwindenden oder neu auftretenden Objekten zu modifizieren sind, der konkrete Entwurf eines 3D-Kalman-Filters, die Behandlung der Nichtlinearität des Beobachtungsmodells und anderes mehr, werden im Teil III der Arbeit vertieft.

Aus Sicht der Navigation sind neben Anforderungen an die algorithmische Lösung insbesondere Anforderungen an die Navigationsgenauigkeit von Bedeutung:

- bei schnell beweglichen Fahrzeugen oder Flugzeugen wird im allgemeinen ein eher geringes Genauigkeitsniveau (Positionsunsicherheiten im 10 m- oder 100 m-Bereich) erwartet. Die Bedeutung der Genauigkeit tritt hinter den Anspruch kurzer Meß-, Analyse- oder auch Reaktionszeiten zurück.
- bei langsameren Landfahrzeugen, Schiffen, mobilen Robotern bleibt mehr Zeit für den Meß- und Auswerteprozess, wobei dann oft ein höheres Genauigkeitsniveau (z.B. bei der Schiffsführung im dm bis m-Bereich) erzielt werden soll.
- das gegenteilige Extrem zum schnell beweglichen Fahrzeug mit kurzer Reaktionszeit ist bei geodätischen und photogrammetrischen Vermessungsaufgaben die Regel, beispielsweise bei Fahrzeugen, die eine Bestandsaufnahme einer Autobahn erlauben (Schwarz, 1992). Hierbei werden teilweise sehr hohe Meßgenauigkeiten (3D-Koordinierung im cm-Bereich) erwartet, hingegen spielt die Meß- und Auswertedauer eine untergeordnete Rolle, da die Bilddaten meist interaktiv nach Abschluß der Mission ausgemessen werden.

1.7.3 Der Realwelt-Datensatz: Bildsequenz Rastatt

In den Experimenten, die diese Arbeit begleiten, wird eine Bildfolge von 160 Bildern verwendet. Diese Sequenz ist in einer Experimentieranlage der Daimler Benz AG in Rastatt aufgezeichnet worden. Es handelt sich dabei um ein Gelände von mehreren hundert Metern Ausdehnung in Nord-Süd und in Ost-West Richtung, in dem verschiedenste Versuche zur autonomen Fahrzeugführung vorgenommen werden können. Als Plattform diente

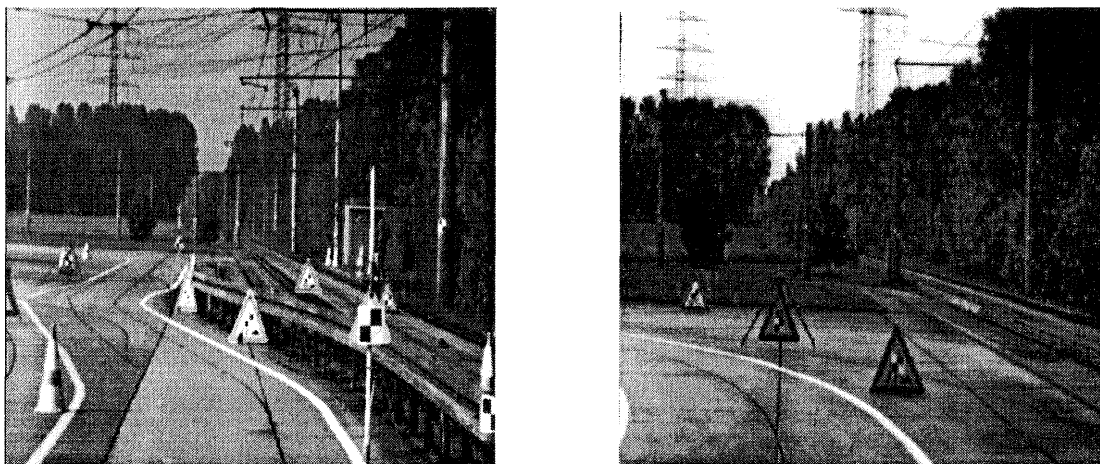


Abbildung 1.5: Aufnahme 40 und 120 der Rastatt-Bildsequenz

ein Fahrzeug, das mit einem Bildaufzeichnungssystem (SR151 der Fa. ITI) ausgestattet war und an dem eine handelsübliche CCD Video-Kamera angeschlossen wurde. Die Befestigung der Kamera erfolgte auf dem Dach des Fahrzeugs, wobei die Aufnahmeachse grob an der Fahrzeugachse ausgerichtet wurde, die Kamera folglich in Fahrtrichtung blickt. Einen Eindruck von der auszuwertenden Szene bzw. der Experimentieranlage, in der das Fahrzeug einen Weg von ca. 160 Metern Länge zurückgelegt hat, geben das in Abbildung 1.5 dargestellte 40ste und 120ste Bild der Sequenz. Weitere Daten sind in Tabelle 1.1 zusammengefaßt.

Die Erfassung geometrischer Kontrollinformationen durch geodätische Messungen und die Kalibrierung der Kamera sind für die experimentellen Untersuchungen eine wichtige Voraussetzung. Sie sollen dennoch nur in knapper Form angesprochen werden, da zu diesen Themen in der Arbeit kein wissenschaftlicher Beitrag geleistet wird.

Ziel der geodätischen Messungen ist es, eine unabhängige Kontrolle von Ergebnissen der Bildsequenzanalyse zu ermöglichen. Hierzu wurde ein lokales geodätisches Netz angelegt und die Position der Kamera bei jeder zweiten Aufnahme von zwei Standpunkten aus mit Theodoliten eingemessen. Die dreidimensionale Koordinierung erfolgte durch Standardtechniken der geodätischen Netzausgleichung. Um Kontrollinformation über die Szene

| | |
|---|---|
| Bild, Kamera | |
| Bildformat, -typ | : 480 × 512 Pixel, 8 bit tiefes Graubild |
| Pixelgröße (horizontal, vertikal) | : 8 μ m × 12 μ m |
| Öffnungswinkel der Kamera | : 11° × 17° |
| Brennweite | : 20 mm bzw. 2000 Pixel |
| Kalibrierungsergebnisse, innere Orientierung | |
| Maßstabsfaktor (Affinkorrektur der mittleren Pixelgröße) | |
| in der Spaltenkoordinate | : 1 + 0.193 |
| in der Zeilenkoordinate | : 1 - 0.193 |
| Brennweite | : 2107.7 Pixel |
| Bildhauptpunkt | : Bildmitte |
| Bewegung in der Sequenz | |
| Verschiebung eines Bildpunktes zwischen zwei Aufnahmen aufgrund | |
| translatorischer Bewegung | : bis zu 25 Pixel |
| rotatorischer Bewegung | : bis zu 50 Pixel |
| Sequenz | |
| Länge der Sequenz | : 160 Bilder |
| Aufnahmedichte | : 1 m Basis zwischen zwei aufeinanderfolgenden Aufnahmen |
| Kameraausrichtung | : ist horizontal um ca 10° gegenüber Fahrzeuglängsachse verschwenkt |

Tabelle 1.1: Charakterisierung der Bildsequenz Rastatt

zu erhalten, wurde die Position einzelner markierter Kontrollpunkte bestimmt. Hierzu sind vor der Aufzeichnung der Sequenz Signaltafeln entlang der Fahrbahn aufgestellt worden (siehe Abbildung 1.5). Zur eindeutigen Definition der Bezugspunkte auf diesen Signaltafeln sind zusätzlich Zielmarken aufgebracht, die aufgrund des einfachen Musters mit großer Wahrscheinlichkeit bei der Extraktion markanter Punkte detektiert werden. Werden diese Punkte in der Analyse verfolgt und in 3D koordiniert, können sie mit den Kontrolldaten verglichen werden. Die Erfassung dieser geometrischen Kontrollinformation in der Szene erfolgte im Zuge der Anlage des lokalen geodätischen Netzes. Ziel der geodätischen Messung war die Bestimmung geometrischer Positionsdaten mit einer Genauigkeit von 1 cm, durch die eine Kontrolle der Bewegung des Sensors und stichprobenhaft auch der Struktur der Szene mit einer übergeordneten Genauigkeit möglich ist. Diese Genauigkeitsvorgabe stellt aus geodätischer Sicht keine besonderen Ansprüche an die Meßtechnik und konnte auch leicht erreicht werden. Eine ausführliche Beschreibung der geodätischen Arbeiten und Resultate findet sich in Hirt (1990).

Die Anzielung der Kamera mit Theodoliten mußte manuell vorgenommen werden. Dies bedingt, daß das Fahrzeug bei der Erfassung einer Aufnahme kurzzeitig (mehrere Sekunden) still stehen mußte, die Aufzeichnung der Sequenz also nicht aus dem bewegten Fahrzeug heraus erfolgen konnte. Um eine Aufzeichnung mit Videotakt bei einer Fahrt mit einer Geschwindigkeit von 25 Metern pro Sekunde nachzubilden, wurde das Fahrzeug translatorisch um 1 Meter zwischen zwei Aufnahmen fortbewegt. Die rotatorische Bewegung ist durch die Lenkbewegung beim Führen des Fahrzeugs auf der Bahn festgelegt. Die im Zuge der Auswertung ermittelten Auswirkungen auf die projizierte Bewegung sind in Tabelle 1.1 dargestellt. Verschiebungen von Bildpunkten um bis zu 50 und mehr Pixel machen deutlich, daß es sich bei der erfaßten Sequenz um eine lose Bildfolge handelt.

Der zweite Aspekt betrifft die Kalibrierung der Kamera. Es wurde eine sehr einfache Kalibrierszene vor der Erfassung der Sequenz vor Ort aufgebaut. Dabei handelt es sich um fünf Signaltafeln, die aus unterschiedlicher Entfernung aufgenommen wurden (siehe Abbildung 1.6). Ziel war es, die Brennweite der Kamera und den Maßstabsfaktor zur Berücksichtigung der rechteckigen Zellengröße des Sensors zu bestimmen. Die Kameraposition und die Objektkoordinaten der Signale wurden geodätisch eingemessen. Zur Bestimmung der Lage der Signale

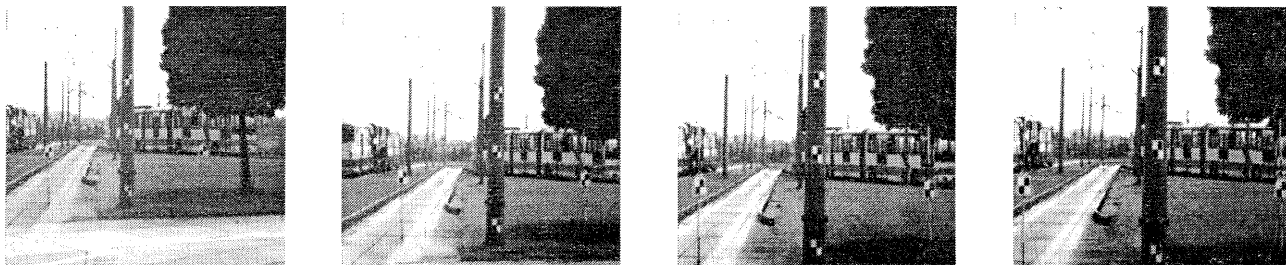


Abbildung 1.6: Kalibrierszene

im Bild ist ein Eckenfinder (Kapitel 4) verwendet worden, der eine Lokalisiergenauigkeit von $1/3$ Pixel besitzt. Mit den vier Aufnahmen erfolgte eine Bildtriangulation, wofür Standardroutinen der analytischen Photogrammetrie verwendet wurden. Die Brennweite und der Bildmaßstab in Zeilen- und Spaltenrichtung können als Selbstkalibrierungsparameter der Triangulation mitbestimmt werden. Diese die innere Orientierung der Kamera betreffenden Größen sind in der Einheit Pixel angegeben, wobei eine mittlere Pixelgröße von $10\mu\text{m}$ zugrunde liegt. Die Resultate sind in Tabelle 1.1 aufgelistet. Weitere Selbstkalibrierungsparameter des von Ebner (1976) vorgeschlagenen Parametersatzes und auch eine Abweichung des Bildhauptpunktes von der Bildmitte haben sich nicht oder nicht signifikant bestimmen lassen. Diese einfache Kalibrierung des Sensors ist in Hirt (1990) ausführlicher dargestellt.

1.8 Problemstellung und Zielsetzung dieser Arbeit

Für die Bewegungsmessung in Bildsequenzen und die Extraktion dreidimensionaler Bewegungs- und Strukturparameter sind eine Vielzahl von Techniken und algorithmischen Lösungen vorgeschlagen worden. Gerade in dieser Vielfalt bildet die Analyse von Bildsequenzen heute ein zentrales Forschungsthema im Bereich Computer Vision. Die bisherigen Erfolge werden von Zhang und Faugeras (1992) dennoch sehr kritisch bewertet: "Die Bedeutung dieses Forschungsthemas leitet sich daraus ab, daß Erfolg oder Mißerfolg festlegen werden, ob Vision-Systeme¹ als Sensoren in diesem Aufgabenfeld bestehen können". Aus dieser Einschätzung läßt sich unmittelbar ableiten, daß weder der Entwurf solcher Systeme allgemein geklärt ist, noch die Einzelalgorithmen (Systemkomponenten) hinreichend analysiert bzw. entwickelt sind. Zu dieser allgemeinen Problemstellung will die Arbeit einen Beitrag leisten, wobei die spezifische Aufgabe des Systems die Analyse langer Bildsequenzen zur passiven Navigation ist.

Einen weiteren gewichtigen Schwachpunkt bei nahezu allen bislang vorgestellten Arbeiten zur Analyse langer Bildsequenzen formulieren Barron et al. (1990) so: "Es ist etwas enttäuschend, daß nahezu keine dieser Rekonstruktionstechniken erfolgreich auf Bewegungsfelder angewandt wurde, die aus der Berechnung realistischer Szenen hervorgingen." Eine besondere Herausforderung für die Untersuchungen in dieser Arbeit besteht deshalb darin, die entwickelten Algorithmen zur Zuordnung und Rekonstruktion erfolgreich an der im Abschnitt 1.7.3 vorgestellten Realweltsequenz zu erproben. Dabei sollen durch die Experimente auch Schwächen bestehender Techniken aufgezeigt werden.

Für die theoretischen Betrachtungen bestehen die Schwerpunkte der Arbeit in der Bewegungsmodellierung, im Meßprozeß zur Bestimmung von Bewegung in Bildsequenzen und in der Integration von Modellierung und Messung innerhalb eines Systementwurfs, der Navigationsansprüchen nach einer rekursiven Parameterschätzung gerecht wird.

Bewegungsmodellierung

Die Bewegungsmodellierung bildet die Grundlage für die Formulierung der Bildzuordnungsansätze und der Rekonstruktionsansätze. Die Beschreibung der Bewegung starrer Körper im dreidimensionalen Raum, wie sie in der Kinematik üblich ist, wird im Computer Vision Bereich meist übernommen. Bewegungsparameter sind

¹Systeme, deren primäre Datenquelle Bilder sind

dabei die rotatorischen und translatorischen Geschwindigkeiten, die bei einer navigierenden Kamera vorwiegend im mitbewegten Koordinatensystem parametrisiert werden. Das Komplement in der Photogrammetrie führt zum sogenannten Bündelansatz, in dem die Position und Richtung einer Kamera zum jeweiligen diskreten Aufnahmezeitpunkt parametrisiert werden, wobei hier ortsfeste Bezugskordinatensysteme verwendet werden. Beide Vorgehensweisen haben Vorteile, z. B. die Parametrisierung der Rotation im mitbewegten Sensorsystem und die Verwendung von Quaternionen oder die Koordinierung der Objektgeometrie im szenefesten Koordinatensystem. Die Beziehungen zwischen den Vorgehensweisen werden grundlegend aufgearbeitet und diskutiert und führen zu einem Ansatz, der sich für die Rekonstruktion von Bewegung und Struktur bei langen Bildsequenzen besonders eignet. Hier grenzt sich die Arbeit sehr stark von den vor allem für die Zwei- und Dreibildeinheit unterbreiteten Vorschlägen ab. Die Generalisierung auf lange Sequenzen ist ein bislang sehr wenig beachteter Aspekt.

Für die Beschreibung der projizierten Bewegung in der Bildebene einer bewegten Kamera ist das Modell-Wissen über den Objektraum von großer Bedeutung. Bei der Navigation in einer unbekanntem Umgebung beschränkt sich dieses Wissen in der Regel darauf, daß die Umgebung stationär ist. Lokal kann dabei die Objektoberfläche oft durch eine einfache Objektform approximiert werden. Solche Heuristiken werden zur Ableitung der projizierten Bewegung entsprechender Objekte eingeführt. Insbesondere für Objektebenen werden die Differentialgleichungen der projizierten Bewegung und die Beziehungen zu den Bewegungs- und Strukturparametern abgeleitet, da bei der Bildzuordnung darauf zurückgegriffen wird.

Bewegungsmessung in der Bildsequenz

Die Vielzahl der Algorithmen zur Bewegungsmessung in der Bildsequenz läßt die Gemeinsamkeiten und Abgrenzungen nur schwer erkennen. Ein Ziel der Arbeit ist deshalb eine systematische Aufarbeitung dieser Ansätze, in der insbesondere Querbezüge aufgezeigt und diskutiert werden. Ein Beispiel hierfür ist die vergleichende Analyse einer Variationslösung zur Bestimmung des optischen Flusses mit einer Kleinste-Quadrate-Lösung, die im photogrammetrischen Bereich vorgeschlagen wurde. Hierbei ergeben sich sehr interessante Einblicke in die Eigenschaften der Verfahren.

Die Schwerpunkte bisheriger Arbeiten liegen in der Bestimmung des optischen Flusses und in der merkmalsbasierten Verfolgung von Primitiven. Die Bestimmung des optischen Flusses ist ein stark unterbestimmtes und bislang eher unbefriedigend gelöstes Problem. Da zuverlässige Schätzungen des optischen Flusses nur in Bildbereichen mit Intensitätsdiskontinuitäten erzielt werden, liegt es nahe, Verfahren zu favorisieren, die auf der symbolischen Beschreibungsebene angesiedelt sind. Ein großer Schwachpunkt solcher korrespondenzbasierter Ansätze sind die Primitive selbst. Von extrahierten Punkten wie z. B. Ecken wird erwartet, daß sie sehr zuverlässig extrahiert werden können und dabei möglichst lange in einer Sequenz verfolgbar sind. Komplexere Primitive, wie z. B. gerade Linien, haben das Problem, daß sie sich nur einigermaßen befriedigend in polyedrischen Szenen extrahieren und verfolgen lassen.

Aus der theoretisch orientierten Aufarbeitung der Bildzuordnungsverfahren bei Bildsequenzen heraus wird ein neuer Ansatz entwickelt, um die Position von markanten Punkten zu verfolgen. Er unterscheidet sich von bisherigen Ansätzen dadurch, daß eine spezielle Repräsentationsform einer Bildpyramide, der sogenannte Scale Space, in den Zuordnungsprozeß einbezogen wird.

Experimente zur Punktverfolgung auf Merkmals- und Intensitätsbasis dienen auch dazu, bestehende Schwachpunkte empirisch zu erfassen. Hier interessieren die Stabilität der Punkte in der Sequenz und deren Qualität. Ferner ist das Zusammenwirken von Zuordnung und Rekonstruktion für die erfolgreiche Verarbeitung der Realweltsequenz von Bedeutung.

Integration durch ein 3D-Kalman-Filter

Die unter navigatorischen Gesichtspunkten möglichst unmittelbar an die Messung anschließende Bewegungs- und Strukturbestimmung kann algorithmisch durch rekursive Parameterschätzung gelöst werden. Unter den bisherigen Ansätzen besteht bislang noch sehr wenig Konsens darüber, wie die Verallgemeinerung der sehr viel intensiver erforschten Zwei- und Dreibildprozeduren für eine rekursive Rekonstruktion bei langen Sequenzen

erfolgen soll. Hingegen besteht Einigkeit darüber, daß sich aus systemtheoretischer Sicht Kalman-Filter zur Lösung der Aufgabe empfehlen.

In der Integration verschiedener Informationsquellen durch ein zu entwerfendes 3D-Kalman-Filter kommt die konzeptionelle Geschlossenheit in dieser Arbeit zum Ausdruck. Der Schlüssel bei der rekursiven Rekonstruktion aus der Monosequenz liegt darin, daß die Information aus den zurückliegenden Schätzungen zum aktuellen Zeitpunkt der Messung theoretisch vollständig zur Verfügung gestellt werden kann. Dies erfolgt durch Prädiktion der Bewegungs- und vor allem der Strukturparameter einschließlich deren Stochastizität, womit im Sinne der Projektionsansätze eine aus dem Prozeß heraus gewonnene Modellierung der Szene vorliegt. Die Projektion dieser Szeneinformation in das Bild gibt die Vorinformationen für den Meßprozeß. Nach erfolgter Korrespondenz liegen die Meßdaten für die sogenannte Aufdatierung des Zustandsvektors vor, die im Sinne der Inversionsansätze zur präziseren Beschreibung der Szene führt. Hier wird deutlich, daß der zumindest partiellen Szenerekonstruktion in einem solchen Filter eine wichtige Rolle zukommt.

Vorgeschlagen und algorithmisch realisiert wird ein Filter, dessen Zustandsvektor die Bewegungsparameter einer Zweibildeinheit (Position, Richtung und die diskreten translatorischen und rotatorischen Geschwindigkeiten) und die 3D-Struktur der Szene in Form von 3D-Punktkoordinaten beinhaltet. Hinter den Systemgleichungen dieser Zweibildeinheit steht ein heuristisch formuliertes Bewegungsmodell, in dem das eher qualitative Wissen über das bewegte System (die Beschleunigung wird konstant angenommen bzw. dem Systemrauschen zugerechnet) quantifiziert wird. Da die Beobachtungsgleichungen des Filters aus dem oben angeführten nichtlinearen Rekonstruktionsansatz bestehen, handelt es sich um ein sogenanntes erweitertes Kalman-Filter. Die Gewinnung von Näherungswerten, die Sensitivität des Systems gegen Meß- und Modellierungsfehler, die Behandlung des Zustandsvektors bei verschwindenden oder neu auftretenden Objektpunkten einschließlich der zugehörigen Kovarianzmatrix und andere Fragestellungen sind in diesem Teil auszuarbeiten.

Der Rekonstruktionsansatz gilt nicht nur für eine Mono-, sondern auch für eine Multibildsequenz. Die einzelnen Bilder mehreren Sequenzen müssen dabei zeitlich nicht synchronisiert sein. Ein entsprechender Filterentwurf muß dann um die Bewegungsparameter entsprechender Kameras erweitert werden. Diese Erweiterung auf Multibild- ebenso wie auf Stereobildsequenzen soll in der Arbeit nicht weiter verfolgt werden.

In Experimenten mit den Ergebnissen des Meßprozesses aus der Realweltsequenz soll auch hier die Praxis-tauglichkeit nachgewiesen werden, wobei unabhängige geodätische Messungen eine Absolutkontrolle der experimentell erzielten Resultate erlauben. Gleichrangig diesem Ziel ist, wie schon bei der Messung, auch bei der Rekonstruktion Schwachpunkte aufzuzeigen, zu diskutieren und soweit wie möglich zu beseitigen.

2 Bewegungsmodellierung

Wissenschaftlicher Beitrag der Arbeit in diesem Kapitel:

Zur Bewegungsmodellierung in Bildsequenzen wird in Arbeiten aus dem Bereich des maschinellen Sehens die zeitkontinuierliche Betrachtung bevorzugt. Oft findet der Übergang auf eine zeitdiskrete Betrachtung erst bei der algorithmischen Lösung einer Problemstellung statt. Im Gegensatz hierzu gehen die in der Photogrammetrie üblichen Modellierungsansätze und Lösungswege z. B. zur Rekonstruktion von Objektpunktkoordinaten und der Orientierung der Bilder immer von einer diskreten Betrachtung aus. Aus der Aufarbeitung dieser Bereiche und der Diskussion von Querbezügen wird ein Ansatz für die Rekonstruktion von Bewegungs- und Strukturparametern vorgeschlagen. Er vereint Vorteile, die sich durch die Parametrisierung der Rotationsbewegung im mitbewegten Koordinatensystem unter Einbeziehung der Quaternionen-Repräsentation ergeben mit Vorteilen, die ein szenefestes Koordinatensystem für die Prozessierung bei langen Bildsequenzen bietet. Die Modellierung der projizierten Bewegung verschiedener Objektformen sowie der Rekonstruktionsansatz bilden die Grundlage für die nachfolgenden Kapitel.

Die Bewegung materieller Körper zählt zu den frühesten Forschungsthemen, denen Pioniere der Physik nachgingen (Goldstein, 1989). Insbesondere in elementaren Lehrbüchern zur Mechanik wird die Beschreibung von Bewegungsvorgängen üblicherweise mit der Unterteilung in Kinematik und Dynamik abgehandelt. Das Objekt bzw. der Körper, dessen Bewegung es zu beschreiben gilt, kann durch sehr verschiedenartige Ursachen in Bewegung gesetzt werden. Als Ursachen, bzw. auch als Rahmenbedingungen der Bewegung sind z. B. Stöße, Gravitationskräfte, Strahlungsdruck, u.v.a.m. wie Eigenschaften des Körpers und der ihn umgebenden Materie in Betracht zu ziehen. Die Bewegung findet in Einklang mit den Gesetzen der Dynamik statt. Das ausschließlich auf die Geometrie beschränkte Studium der Bewegungsformen wird als Kinematik bezeichnet. Eine weit verbreitete Betrachtungsweise zur Bestimmung von Bewegung sieht eine Zweiteilung der Aufgabenstellung vor, in der zuerst die Bewegung aufgrund kinematischer Modellierung rekonstruiert wird. Die Kräfte, welche die Bewegung verursachen, werden in einem zweiten Schritt aufgrund der dynamischen Modellierung und der rekonstruierten Bewegung ermittelt (Beggs, 1983).

Für die Beschreibung der Bewegung eines Körpers wollen wir uns auf die Kinematik beschränken. Die dynamischen Rahmenbedingungen für die Bewegungsmodellierung sind in allgemeinen sehr unterschiedlich, wohingegen die Geometrie der Bewegung eines starren Körpers durch Rotation und Translation beschrieben werden kann. Die Starrheitsannahme ist wohl die am meisten postulierte Heuristik (siehe auch Abschnitt 1.3.3). Ihre Rechtfertigung muß in Abhängigkeit vom Objekt, aber auch in Abhängigkeit vom registrierenden Sensor, vom Maßstab, etc. überprüft werden. D. h. generell bezieht sich die Starrheitsannahme sowohl auf das registrierende Sensorsystem als auch auf das beobachtete Objekt. Das Sensorsystem kann und wird auch in der Regel als starres Koordinatensystem angenommen. Für die vom Sensor beobachtete dynamische Welt stellt die Starrheitsheuristik eine signifikantere Beschränkung dar. So haben z. B. Rohr und Nagel (1990) die Modellierung der Bewegung einer Person durch Starrheitsannahme einzelner zusammenhängender Körperteile (Volumenprimitive) vorgenommen (vgl. Abbildung 2.1).

Für andere Objekte, wie z. B. einfache Fahrzeuge, erscheint die Heuristik "ein starrer Körper" weniger kritisch. Bei statischen Szenen ist die Starrheit implizit gegeben, so daß das Bewegungsmodell für den Sensor durch rotatorische und translatorische Parameter beschrieben werden kann.

Exakte dynamische Modelle müssen für die verschiedenen Objekte in ihrer entsprechenden Umgebung individuell formuliert und zur Vereinfachung des Rekonstruktionsprozesses in der Regel auch approximiert werden. Ein solcher Weg wurde z. B. von Mezger (1988) für das Objekt *Schiff* ausgearbeitet. Die Bewegungsfreiheitsgrade auf dem Wasser sind auf zwei Translationen und eine Rotation beschränkt worden.

In den nachfolgenden Abschnitten dieses Kapitels wollen wir zunächst die Starrkörperbewegung im dreidimensionalen Raum darstellen (Abschnitt 2.1) und dabei der Parametrisierung der Rotationsmatrix (Abschnitt 2.2) besondere Aufmerksamkeit schenken. Mit der Projektion einer beobachteten Szene auf die Bildebene und der Beschreibung der projizierten Bewegung (Abschnitt 2.3) werden die geometrischen Grundlagen der Bildzuordnung (Teil II der Arbeit) und der Rekonstruktion von Bewegung und Struktur (Abschnitt 2.4) erarbeitet.

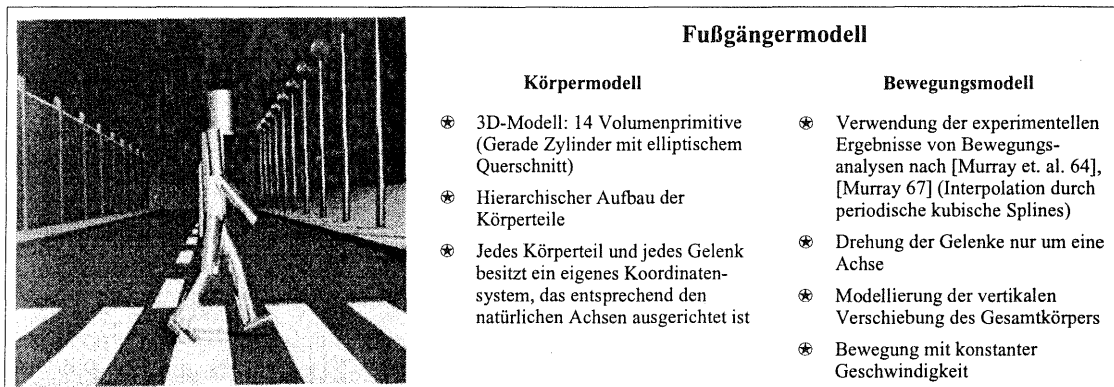


Abbildung 2.1: Das Fußgängermodell umfaßt ein Körpermodell und ein Bewegungsmodell (aus Rohr und Nagel, 1990, S. 53)

2.1 Bewegung im dreidimensionalen Raum

2.1.1 Starrkörperbewegung

Um Bewegung zu beschreiben, muß zunächst ein Bezugssystem festgelegt werden. Üblicherweise werden hierfür links- oder rechtshändige, kartesische Koordinatensysteme gewählt. Im weiteren werden rechtshändige, kartesische Koordinatensysteme verwendet. Zur Beschreibung der Bewegung eines starren Körpers werden zwei Koordinatensysteme, ein ortsfestes und ein bewegtes System, festgelegt. Die Position eines Punktes bezogen auf das ortsfeste Koordinatensystem soll mit \mathbf{U} , bezogen auf das bewegte Koordinatensystem mit \mathbf{X} bezeichnet werden. Die Transformation des Punktes vom bewegten in das ortsfeste System wird durch

$$\mathbf{U} = R \mathbf{X} + \mathbf{U}_0 \quad (2-1)$$

beschrieben, wobei mit R eine Rotationsmatrix und mit \mathbf{U}_0 der Koordinatenursprung des bewegten Systems hinsichtlich seiner Position im ortsfesten System bezeichnet wird. Sowohl $R = R(t)$ als auch $\mathbf{U}_0 = \mathbf{U}_0(t)$ sind Funktionen der Zeit, so daß durch (2-1) eine Form des aus der Kinematik bekannten *Bewegungsmodells* für starre Körper formuliert ist.

Anwendungsbezogen lassen sich vor allem zwei typische Situationen skizzieren:

1. Das sich bewegende starre Objekt \mathbf{X} wird im ortsfesten System (zu bestimmten Zeitpunkten) registriert. Das lokale System bewegt sich mit dem starren Körper mit, d. h. die Position \mathbf{X} eines beliebigen Punktes des starren Körpers ist unabhängig von der Zeit. Mit $\mathbf{X}(t) = \mathbf{X}$ ergibt sich

$$\mathbf{U}(t) = R(t) \mathbf{X} + \mathbf{U}_0(t) \quad (2-2)$$

mit

$$\mathbf{X} = R^T(t_0)(\mathbf{U}(t_0) - \mathbf{U}_0(t_0))$$

Die zweite Zeile in (2-2) bringt die Startsituation (Bezugszeitpunkt t_0) zum Ausdruck; die erste Zeile beschreibt die Bewegung des starren Körpers (\mathbf{X}) im ortsfesten Referenzsystem.

2. Die stationäre Szene, repräsentiert durch \mathbf{U} , wird im bewegten System (zu bestimmten Zeitpunkten) registriert. Das bewegte System steht für das Bezugssystem z. B. eines sich bewegenden Sensors. Wegen der Stationarität der Szene ist $\mathbf{U}(t) = \mathbf{U}$ und es ergibt sich

$$\mathbf{X}(t) = R^T(t) (\mathbf{U} - \mathbf{U}_0(t)) \quad (2-3)$$

mit

$$\mathbf{U} = R(t_0)\mathbf{X}(t_0) + \mathbf{U}_0(t_0)$$

In Analogie zu (1.) gilt, daß wenn die Startsituation $(R(t_0), \mathbf{X}(t_0), \mathbf{U}_0(t_0))$ bekannt ist, die Positionsbeschreibung der stationären Szene im bewegten System durch Rotation $R(t)$ und Translation $\mathbf{U}_0(t)$ festgelegt wird.

Deutlich wird aus dieser Diskussion, daß die Bewegungsbeschreibung eines bezüglich seiner Geometrie (\mathbf{X} bzw. \mathbf{U}) bekannten starren Objekts zwar in verschiedenen Koordinatensystemen, aber dennoch immer vollständig durch Rotation und Translation erfolgen kann.

Diese beiden Varianten (1.) und (2.) des Bewegungsmodells lassen sich darüber hinaus vielfältig kombinieren. So können z. B. anstelle der stationären Szene ein oder mehrere bewegte Objekte im System eines bewegten Sensors registriert werden. Die Rotationsmatrix und der Translationsvektor sind für jedes Objekt individuell vorzusehen und geben dann die relativen rotatorischen und translatorischen Bewegungen bezüglich der Objekte an.

Eine Rotationsmatrix ist orthogonal und durch 3 frei wählbare Parameter festgelegt. Der Translationsvektor besitzt ebenfalls drei Komponenten, so daß insgesamt das Modell der Starrkörperbewegung (2-1) über 6 Freiheitsgrade verfügt. Diese 6 Variablen, die als Funktionen der Zeit die Starrkörperbewegung beschreiben, sollen als *Bewegungsparameter* bezeichnet werden. Bei den differentiellen Betrachtungen der Bewegung werden die rotatorischen und translatorischen Freiheitsgrade durch Geschwindigkeiten beschrieben. Es ist deshalb üblich, diese 6 Geschwindigkeitskomponenten ebenfalls als Bewegungsparameter zu bezeichnen. Ergänzend zur Starrheitsannahme können Bahnmodelle explizit vorgegeben sein, durch die eine Bewegung weiter beschränkt wird, z. B. mit der Forderung nach Glattheit der Bewegung. Die Bahnmodellierung kann dazu führen, daß besonders für die Situation längerer Zeitabschnitte Bahnparameter alternativ oder ergänzend zu den Bewegungsparametern verwendet werden. Auch diese Parameter werden als Bewegungs- oder auch als kinematische Parameter aufgeführt.

Für gewisse Aufgaben vor allem im Bereich der Stereoanalyse (Korrespondenz, Rekonstruktion) ist lediglich die Änderung der Bewegungsparameter von Interesse. Mit der Vorgabe, daß zum Zeitpunkt t_0 das Bewegungsmodell (2-1) der Restriktion $\mathbf{X}(t_0) = \mathbf{U}(t_0)$ genügt, wird festgelegt, daß bewegtes und ortsfestes System zum Bezugszeitpunkt t_0 zusammenfallen. Rotations- und Translationsparameter sind zum Zeitpunkt t_0 mit Null anzunehmen, womit sich den differentiellen Betrachtungen eine geeignete Linearisierungsstelle anbietet. Genau dieser Vorteil läßt sich auch dadurch erzielen, daß z. B. in (2-3) \mathbf{U} formal eliminiert wird:

$$\mathbf{X}(t) = R(t, t_0) \mathbf{X}(t_0) + \mathbf{X}_0(t, t_0), \quad (2-4)$$

wobei

$$\begin{aligned} R(t, t_0) &= R^T(t) R(t_0) \\ \mathbf{X}_0(t, t_0) &= R^T(t) (\mathbf{U}_0(t_0) - \mathbf{U}_0(t)). \end{aligned}$$

Die Koordinaten $\mathbf{X}(t_0), \mathbf{X}_0(t, t_0)$ beziehen sich auf das bewegte System. Für $t \rightarrow t_0$ streben $R(t, t_0) \rightarrow I$, $\mathbf{X}_0(t, t_0) \rightarrow 0$. Damit wird die besondere Rolle des bewegten Bezugssystems für eine Linearisierung deutlich. Wir werden bei den differentiellen Betrachtungen im nächsten Abschnitt direkt davon Gebrauch machen. Soll das starre Objekt zum jeweils aktuellen Zeitpunkt t_i im bewegten System beschrieben werden, bietet (2-4) die Vorschrift für den Datumstransfer aus einem vorhergehenden Zeitpunkt. Von dieser Form des Bewegungsmodells wird deshalb bei Navigationsaufgaben oft Gebrauch gemacht. Wird die Bewegung zwischen den Zeitpunkten in Form von $R(t, t_0)$ und $\mathbf{X}_0(t, t_0)$ beobachtet oder rekonstruiert, kann die Aufdatierung des aktuellen Rotations- und Translationszustandes durch Umformung aus (2-4) gemäß

$$\begin{aligned} R(t) &= R(t_0) R^T(t, t_0) \\ \mathbf{U}_0(t) &= \mathbf{U}_0(t_0) - R(t) \mathbf{X}_0(t, t_0) \end{aligned}$$

direkt aus der Rotation und Translation des vorhergehenden Zeitpunktes und der Bewegung zwischen t_0 und t erfolgen. Übertragen auf zukünftige Zeitpunkte ist damit auch eine einfache Prädiktionsformel gegeben, die wiederum die Anwendung in der Navigation nahelegt.

2.1.2 Differentielle Bewegung

Mit der Differentiation des Bewegungsmodells (2-1) nach dem Parameter Zeit werden die rotatorischen und translatorischen Geschwindigkeiten in die Modellierung einbezogen. Die differentielle Betrachtung dient, zusammen mit der Beschreibung der projizierten Bewegung (Abschnitt 2.2), als Grundlage für den Rekonstruktionsprozeß der Bewegungsparameter (Teil III). Die Differentiation von (2-1) nach der Zeit t lautet

$$d_t \mathbf{U} = d_t R \mathbf{X} + R d_t \mathbf{X} + d_t \mathbf{U}_0, \quad (2-5)$$

wobei der Ableitungsoperator d_t gemäß $d_t := \frac{d}{dt}$ definiert ist. Setzen wir voraus, daß die *stationäre Szene im bewegten System* registriert wird, so gilt wegen $d_t \mathbf{U} = 0$

$$d_t \mathbf{X} = -R^T d_t R \mathbf{X} - d_t \mathbf{T}, \quad (2-6)$$

wobei die Geschwindigkeit des als Bezugspunkt gewählten Koordinatenursprungs sich gemäß $d_t \mathbf{T} = R^T d_t \mathbf{U}_0$ in den Koordinaten des bewegten Systems ausdrückt. Die Matrix

$$\Omega = R^T d_t R \quad (2-7)$$

ist eine schiefsymmetrische Matrix, was sich direkt durch Differentiation von $R^T R = I$

$$R^T d_t R = -d_t R^T R$$

mit $d_t R^T R = (R^T d_t R)^T$ zeigen läßt.

Mit der allgemeinen Definition einer schiefsymmetrischen Matrix

$$\Omega := \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \quad (2-8)$$

und der Definition der Abbildung des Vektors $\omega^T = (\omega_1 \ \omega_2 \ \omega_3)$ auf Ω

$$\Omega := [\omega]_{\times} \quad (2-9)$$

gilt für das Produkt von Ω mit einem Vektor \mathbf{p}

$$\Omega \mathbf{p} = \omega \times \mathbf{p}, \quad (2-10)$$

d. h. $\Omega \mathbf{p}$ kann in äquivalenter Form durch das Vektorprodukt $\omega \times \mathbf{p}$ dargestellt werden. Die vor allem in der Mechanik übliche Formulierung von (2-6) über das Vektorprodukt

$$d_t \mathbf{X}(t) = -\omega(t) \times \mathbf{X}(t) - d_t \mathbf{T}(t), \quad (2-11)$$

wird in der Literatur (z. B. ICCV1, 1987, ICCV2, 1988, ICCV3, 1990) häufig als Ausgangsmodell für die Bewegungsformulierung benützt. ω steht für die rotatorischen Geschwindigkeiten, $d_t \mathbf{T}$ für die translatorischen Geschwindigkeiten des bewegten Koordinatensystems. Zum besseren Verständnis von ω , $d_t \mathbf{T}$ wollen wir (2-11) etwas allgemeiner formulieren und dabei die Bedeutung und Wahl des Bezugspunktes diskutieren.

Allgemeine Starrkörperbewegung

Um die Starrkörperbewegung des bewegten Koordinatensystems anschaulich zu erklären, soll (2-11) als Spezialisierung der Bewegung eines starren Körpers dargestellt werden. Mit der Translationsgeschwindigkeit $d_t \mathbf{T}$ bezüglich eines willkürlich gewählten Referenzpunktes B und der Rotationsgeschwindigkeit ω um diesen Referenzpunkt lautet das Differentialgleichungssystem der momentanen Starrkörperbewegung

$$d_t \mathbf{X} = \omega \times (\mathbf{X} - \mathbf{X}_B) + d_t \mathbf{T}, \quad (2-12)$$

(vgl. z. B. Kanatani, 1990). \mathbf{X}_B gibt die 3D-Position des gewählten Referenzpunktes im Bezugssystem an. ω definiert eine Rotationsbewegung um eine Achse, die durch den Referenzpunkt (\mathbf{X}_B) geht und deren Richtung

durch $(\omega_1, \omega_2, \omega_3)$ festgelegt ist. Die Drehung um diese Achse hat die Winkelgeschwindigkeit $\sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2}$. Beziehung (2-12) läßt sich auch so interpretieren, daß aus den gegebenen Geschwindigkeiten $\omega, d_t\mathbf{T}$ bezogen auf den Referenzpunkt (\mathbf{X}_B) die Geschwindigkeiten $\omega', d_t\mathbf{T}'$ bezogen auf einen zweiten Referenzpunkt (\mathbf{X}) berechnet werden können, für die sich mit (2-12) ergibt: $\omega' = \omega, d_t\mathbf{T}' = d_t\mathbf{X}$. Der Ursprung O des Bezugssystems stellt eine Möglichkeit zur Wahl eines Referenzpunktes dar. Die Spezialisierung von (2-12) lautet mit $\mathbf{X}_B = 0$:

$$d_t\mathbf{X} = \omega \times \mathbf{X} + d_t\mathbf{T}, \quad (2-13)$$

Die Rotationsachse geht durch den Ursprung des Bezugssystems. Eine Rotation um die Rotationsachse $(\omega_1, \omega_2, \omega_3)$ mit Winkelgeschwindigkeit $|\omega|$ kann als Rotation mit den Winkelgeschwindigkeiten ω_i um die korrespondierenden Achsen X, Y, Z betrachtet werden (Kanatani, 1990). Diese Situation ist in Abbildung (2.2) graphisch

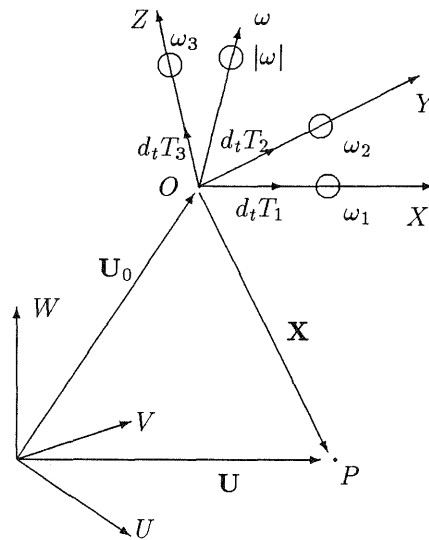


Abbildung 2.2: Interpretation der Bewegungsparameter: Translatorische und rotatorische Geschwindigkeit der Starrkörperbewegung sind mit Bezugspunkt O dargestellt (detaillierte Erläuterung im Text).

wiedergegeben (vgl. z. B. Waxman, 1987). Der direkte Vergleich von (2-13) mit (2-11) zeigt, daß die Bewegung des beobachtenden Sensors (repräsentiert durch das bewegte System) auch dadurch ausgedrückt werden kann, daß sich das Objekt mit den Geschwindigkeiten $(-\omega, -d_t\mathbf{T}; \text{Bezugspunkt } O)$ relativ zum Beobachter bewegt. Das bedeutet, daß jede Starrkörperbewegung sich auch als Eigenbewegung des Beobachters interpretieren läßt.

Mit (2-11) liegt das differentielle Bewegungsmodell in einer Form vor, in der die Rotationsmatrix $R(t)$ bzw. deren zeitliche Ableitung $d_t R(t)$ direkt nicht vorkommt. Deshalb ist es auch nicht nötig, $d_t R$ explizit nach Rotationsparametern zu entwickeln. Dieser Weg, der üblicherweise in den photogrammetrischen Lehrbüchern (z. B. Schwidofsky und Ackermann, 1976) aufgezeigt wird, ist wesentlich aufwendiger und auch hinsichtlich der Bewegungsschätzung umständlicher. Die Verwendung von ω bei der Parameterschätzung von Orientierungsaufgaben wurde von Pope (1970) vorgestellt und von Hinsken (1987) wieder aufgegriffen und numerisch untersucht.

Das Gleichungssystem (2-11) stellt eine Vektordifferentialgleichung dar. Bei der Bewegungsrekonstruktion werden wir auf (2-11) zurückgreifen und sehen, wie ω und $d_t\mathbf{T}$ in diskreter Form sehr vorteilhaft als zu schätzende Parameter eingeführt werden können.

2.2 Rotation

Im vorangehenden Abschnitt wurde die Rotationsmatrix vielfach verwendet, ohne dabei auf die Rotationsparameter einzugehen. Für die differentielle Bewegung war die Parametrisierung von $R(t)$ sogar irrelevant, denn

durch (2-11) ist ein Bewegungsmodell gegeben, in dem alle 6 Bewegungsparameter explizit vorkommen. Die rotatorischen Parameter sind dabei durch die Winkelgeschwindigkeiten ω repräsentiert. Für die diskrete Betrachtung der Bewegung werden die Parameter der Rotationsmatrix in expliziter Form benötigt, um die Bewegung durch 6 Freiheitsgrade vollständig beschreiben zu können.

Für die Entwicklung technischer System mit rotatorischen Freiheitsgraden werden in der Regel mechanisch leicht realisierbare Drehwinkel gewählt. Auch aus Gründen der Anschaulichkeit werden Drehungen bevorzugt, die sich aus Einzeldrehungen um die Achsen eines Bezugskordinatensystems zusammensetzen. In der Photogrammetrie ist es üblich, diese Drehwinkel mit ω, φ, κ zu bezeichnen¹. Weit verbreitet, z. B. in der Mechanik und in der Geodäsie, ist die Parametrisierung mit Euler-Winkeln. Einen Schritt hin zur Quaternionen-Repräsentation, auf die im folgenden detaillierter eingegangen wird, stellt die Rotationsmatrix dar, die als Drehung um eine im Raum festgelegte Achse definiert ist. Eine sehr gute Übersicht über die Repräsentation von 3D-Rotationen geben Fu et al. (1987) und Kanatani (1990).

In der dreidimensionalen orthonormalen Basis $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$, welche die Einheitsvektoren eines ortsfesten kartesischen Bezugssystems angibt, kann ein zweites rotierendes oder rotiertes System mit den Basisvektoren $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ durch die lineare Transformation²

$$\mathbf{r}_i = R \mathbf{b}_i \quad (i = 1, 2, 3)$$

festgelegt werden. Die Matrix

$$R = (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3) = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

definiert eine 3D-Rotation, wenn gilt

$$R^T R = I \quad \text{und} \quad \det R = 1, \quad (2-14)$$

d. h. wenn R eine orthogonale Matrix ist und eine Determinante $\det R = 1$ besitzt. Eine lineare Transformation des dreidimensionalen Raumes auf sich selbst $R^3 \mapsto R^3$, welche die Restriktionen (2-14) erfüllt, wird auch als $SO(3)$ - als Gruppe der speziellen, orthogonalen dreidimensionalen linearen Abbildungen - bezeichnet (Kanatani, 1990). Die Orthogonalität $R^T R = I$ resultiert bekanntlich in 6 unabhängigen Orthogonalitätsrestriktionen, z. B.,

$$\begin{aligned} \mathbf{r}_i^T \mathbf{r}_i &= 1 \quad (i = 1, 2, 3) \\ \mathbf{r}_i^T \mathbf{r}_j &= 0 \quad (\{i, j\} = \{1, 2\}, \{2, 3\}, \{3, 1\}), \end{aligned}$$

so daß für die Parametrisierung von R lediglich 3, allerdings frei wählbare Parameter festzulegen sind. Unabhängig von der gewählten Parametrisierung ist die Rotationsmatrix R *eindeutig*.

2.2.1 Kriterien für die Parametrisierung

Für die Festlegung der drei Freiheitsgrade von R sollen zwei Kriterien vorgegeben werden. Die Parametrisierung soll

1. regulär sein, d. h. es sollen keine kritischen (singulären) Situationen existieren, aus denen die Parameter nicht eindeutig bestimmbar sind.
2. zur Vereinfachung des Problems beitragen.

¹Da keine Gefahr der Verwechslung mit dem Vektor der Winkelgeschwindigkeiten ω besteht, wollen wir in diesem Abschnitt die in der Photogrammetrie übliche Bezeichnungsweise des Drehwinkels ω beibehalten

²Konsistent (allerdings umständlicher zu erläutern) mit (2-1) wäre die Formulierung $\mathbf{b}_i = R \mathbf{r}_i$. Ein Unterschied ergibt sich daraus nicht für die weitere Diskussion.

Die beiden Kriterien sind unterschiedlich präzise formuliert. Während das erste Kriterium relativ klar die Frage nach der voraussetzungslosen, eindeutigen Gewinnung der Rotationsparameter bei gegebener Rotationsmatrix stellt, ist das zweite Kriterium unscharf formuliert. Die Vereinfachung eines Problems hängt natürlich vom Problem selbst ab; d. h. das Kriterium kann lediglich beispielhaft für verschiedene Aufgaben die Vorteile einer Parametrisierung ansprechen. Eine solche Fragestellung ist z. B. (Kanatani, 1990): "Welche Euler-Winkel $\{\psi'', \theta'', \phi''\}$ ergeben sich für eine Rotation, die sich aus einer Folgerotation von $\{\psi', \theta', \phi'\}$ und $\{\psi, \theta, \phi\}$ zusammensetzt?". Die Frage zielt ab auf die Problematik, daß für die Winkelrepräsentation kein einfacher Ausdruck zur Berechnung der Parameter zusammengesetzter Rotationen existiert.

Es ist bekannt, daß z. B. die Parametrisierung mit den Euler-Winkeln ψ, θ, ϕ

$$R(\psi, \theta, \phi) := R_3(\phi)R_1(\theta)R_3(\psi)$$

für $\theta = n\pi$ ($n \in \mathbb{N}$) singularär ist, denn es gilt dann

$$R(\psi, n\pi, \phi) = R_3(\phi)R_3(\psi) = R_3(\phi + \psi).$$

Für die in der photogrammetrischen Literatur häufig benützte Parametrisierung $(\omega, \varphi, \kappa)$ der Drehung eines Punktes um ein festes Achssystem ergibt sich die zusammengesetzte Drehmatrix

$$R(\omega, \varphi, \kappa) := R_1(\omega)R_2(\varphi)R_3(\kappa)$$

zu

$$R(\omega, \varphi, \kappa) = \begin{pmatrix} \cos\varphi \cos\kappa & -\cos\varphi \sin\kappa & \sin\varphi \\ \cos\omega \sin\kappa + \sin\omega \sin\varphi \cos\kappa & \cos\omega \cos\kappa - \sin\omega \sin\varphi \sin\kappa & -\sin\omega \cos\varphi \\ \sin\omega \sin\kappa - \cos\omega \sin\varphi \cos\kappa & \sin\omega \cos\kappa + \cos\omega \sin\varphi \sin\kappa & \cos\omega \cos\varphi \end{pmatrix}. \quad (2-15)$$

Eine singularäre Situation erhält man für $\varphi = \frac{\pi}{2} + n\pi$ ($n \in \mathbb{N}$), in der $\omega \pm \kappa$ nur noch gemeinsam als Argument der Rotationsmatrix auftreten. Ganz allgemein gilt, daß keine auf drei Winkel beruhende Parametrisierung den Anspruch auf Regularität erfüllen kann. Die Ursache liegt darin, daß der dreidimensionale projektive Raum, der den Konfigurationsraum einer Rotation bildet, durch eine Parametrisierung mit drei Winkeln auf einen dreidimensionalen Torus beschränkt wird (Grafarend, 1983).

Neben den Singularitäten führt die Substitution von $(\omega, \varphi, \kappa)$ durch $(\pi + \omega, \pi - \varphi, \pi + \kappa)$ zur selben Rotationsmatrix, d. h. es gilt

$$R_1(\omega)R_2(\varphi)R_3(\kappa) = R_1(\pi + \omega)R_2(\pi - \varphi)R_3(\pi + \kappa).$$

Die Nichteindeutigkeit, die auch als Dualität bezeichnet wird, kann nur durch weitere Vorinformation, z. B. durch die Beschränkung auf $|\varphi| < \frac{\pi}{2}$ beseitigt werden.

2.2.2 Parametrisierung mit den Elementen des Einheitsquaternions

Als Alternative zur Winkelparametrisierung hat sich die Repräsentation einer 3D-Rotation mittels Quaternionen als äußerst vorteilhaft erwiesen. Quaternionen besitzen, ebenso wie Spinore, keine Singularitäten. Grafarend (1983) weist darauf hin, daß dieses Faktum in der Literatur (z. B. in Niemez, 1963, S. 368, Wittenburg, 1977, vgl. auch Hinsken, 1988) eingehend diskutiert wurde. Mit den skalaren, reellwertigen Elementen q_0, q_1, q_2, q_3 ist das (Hamilton) Quaternion $\overset{\circ}{q}$ gemäß

$$\overset{\circ}{q} := q_0 + q_1 j_1 + q_2 j_2 + q_3 j_3 \quad (2-16)$$

definiert, wobei die "Basis" des Quaternions die Beziehungen

$$j_i^2 = -1 \quad (i = 1, 2, 3) \quad (2-17)$$

$$j_1 j_2 = -j_2 j_1 = j_3, \quad j_2 j_3 = -j_3 j_2 = j_1, \quad j_3 j_1 = -j_1 j_3 = j_2$$

erfüllt. Zur Formulierung der Rotationsmatrix mit den Elementen q_i gemäß

$$R = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & -2(q_0q_3 - q_1q_2) & 2(q_0q_2 + q_1q_3) \\ 2(q_0q_3 + q_1q_2) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & -2(q_0q_1 - q_2q_3) \\ -2(q_0q_2 - q_1q_3) & 2(q_0q_1 + q_2q_3) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \quad (2-18)$$

wird vorausgesetzt, daß das Quaternion die Länge 1 besitzt

$$|\overset{\circ}{q}| = q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1 \quad (2-19)$$

Das bedeutet, daß R mit den Elementen eines Einheitsquaternions (2-16, 2-19) parametrisiert ist, wodurch für die Rotationsmatrix R (2-18, 2-19) auch nur drei unabhängige Freiheitsgrade verbleiben. Das Quaternion $\overset{\circ}{q}$ und seine Negation $-\overset{\circ}{q}$ repräsentieren dieselbe 3D-Rotation. Das bedeutet, daß die Parametrisierung mit (q_0, q_1, q_2, q_3) eine duale Lösung in $(-q_0, -q_1, -q_2, -q_3)$ besitzt, was sich durch Einsetzen in (2-18) direkt verifizieren läßt.

Anstelle von einer dualen Lösung wird in der mathematischen Literatur auch von einer 2—1 Korrespondenz zwischen den Einheitsquaternionen und den Elementen der Rotationsmatrix gesprochen. Bezüglich der Fragestellung nach einer globalen 1—1 Korrespondenz (weder Singularität noch Dualität) zwischen Parametrisierung und Rotationsmatrix diskutiert Stuelpnagel (1964) gruppentheoretische Arbeiten und zeigt, daß wenigstens 5 Parameter in die Parametrisierung einbezogen werden müssen. Ein von Stuelpnagel dargestellter Lösungsweg besteht darin, eine 6 dimensionale Parametrisierung (mit den Elementen der ersten beiden Spaltenvektoren einer Rotationsmatrix) auf 5 Parameter durch eine stereographische Projektion einzuschränken. Dieser Lösungsweg führt allerdings zu nichtlinearen Beziehungen zwischen den Parametern und den differentiellen Änderungen der Parameter bei Bewegung, womit er dem zweiten oben formulierten Kriterium entgegen steht.

Das zweite Kriterium fordert die Vereinfachung von rotationsbezogenen Problemstellungen durch eine geeignete Parametrisierung. Beispiel hierfür sind:

1. Thompson (1958/59), Schut (1961), Horn (1987, 1990), Sanso (1973) und Grafarend et al. (1989) nutzen Quaternionen bzw. die Quaternionen-Repräsentation der Rotationsmatrix, um direkte Lösungen für Orientierungsaufgaben zu entwickeln. So wird z. B. von Sanso ein Kleinster-Quadrate-Algorithmus zur räumlichen Ähnlichkeitstransformation in der Quaternionen-Algebra formuliert. Dieselbe Problemstellung mit nahezu identischem Lösungsweg hat Horn (1987) präsentiert, wobei in seiner Darstellung Matrix- und Quaternionen-Algebra kombiniert verwendet werden.
2. Da die Berechnung der Matrixelemente einer durch Drehwinkel parametrisierten Rotationsmatrix trigonometrische Funktionen beinhaltet, ist sie dem polynomialen Matrixtyp (2-18), dessen Elemente alle quadratisch sind, aus numerischen Gründen und hinsichtlich der Effizienz unterlegen (Schut, 1958/59, Thompson, 1959, Pope, 1970).
3. Ein Aspekt, auf den nachfolgend eingegangen wird, betrifft zusammengesetzte Drehungen. Hier bieten die Quaternionen die Möglichkeit einer einfachen, direkten Berechnung (Schut, 1958/59, S.158). Darüberhinaus geht die direkte Berechnung einher mit Effizienzvorteilen. Die Zahl der Rechenoperationen ist bei Multiplikation von Rotationsmatrizen in etwa doppelt so hoch wie die direkte Bestimmung der Rotationsparameter zusammengesetzter Rotationen in Quaternionen über die linearen Beziehungen (2-23). Numerisch bedingte Abweichungen von der Orthonormalität der Rotationsmatrix sind aufwendig zu korrigieren, wohingegen Abweichungen vom Einheitsquaternion durch eine simple Normierung über die Länge beseitigt werden können.
4. Die Formulierung rotatorischer Bewegung mit Euler-Winkeln und mit Quaternionen stellt Grafarend (1983) vergleichend gegenüber. Die Vorteile der Quaternionen-Repräsentation werden am Beispiel einer dynamischen Bewegungsformulierung besonders deutlich, da die Differentialgleichungen bei der Parametrisierung mit Winkeln nichtlinear sind, bei Verwendung der Quaternionen-Elemente aber linear werden. Gleiches gilt für die Differentialgleichungen bei rein kinematischer Betrachtung. Stuelpnagel (1964) betont, daß für die Bewegungsgleichungen rotatorischer Starrkörperbewegung bei einer Parametrisierung mit 5 Parametern noch keine linearen Beziehungen gefunden wurden, so daß er vor allem wegen der linearen Differentialgleichungen 6 oder 4 Parameter Methoden favorisiert. Praktische Gesichtspunkte sprechen vor allem für die 4 dimensionale Parametrisierung mit dem Einheitsquaternion.

5. Basis der Planung einer kollisionsfreien Bewegung eines Objekts durch einen mit Hindernissen versehenen Arbeitsraum ist die Problematik der Kollisionsdetektion. Canny (1986b) zeigt, daß durch die Quaternionen-Repräsentation der Rotation die Detektionsgleichungen eine einfache algebraische Form bekommen, die linear in den Positionskordinaten und quadratisch in den Rotationsparametern ist, wodurch die Berechnung der Kollisionspunkte erheblich vereinfacht wird.
6. Die diskrete rotatorische Bewegungsformulierung gemäß (2-6) in Kombination mit Quaternionen steht hinter der von Pope (1970) vorgeschlagenen Vereinfachung der linearisierten Kollinearitätsgleichung. Daß sich hierdurch eine deutliche Beschleunigung des gesamten Schätzprozesses erzielen läßt, wird von Hinsken (1987) als signifikanter Vorteil herausgestellt. Der Grund für die Vereinfachung liegt in der schiefsymmetrischen Matrix Ω , die sich aus der Folgerotation $R^T d_t R$ (2-7) zusammensetzt. Die Anwendung einer solchen Folgerotation wird von Wrobel und Klemm (1984) empfohlen, um Singularitäten der Winkelrepräsentation zu umgehen.
7. In experimentellen Untersuchungen hat Hinsken (1987) aufgezeigt, daß auch das Konvergenzverhalten iterativ gelöster Orientierungsaufgaben bei der Formulierung von R gemäß (2-18) in Kombination mit Ω (2-8) gegenüber Winkelparametrisierungen weiter verbessert wird, obgleich der Konvergenzradius und das Konvergenzverhalten bei der klassischen Winkelparametrisierung auch schon sehr groß bzw. gut ist. Der Konvergenzradius ist nach den Beobachtungen Hinskens auch bei der Auswertung photogrammetrischer Nahbereichsaufgaben so groß, daß zur Bestimmung der äußeren Orientierung des Bildes durch einen Räumlichen Rückwärtsschnitt keine Näherungswertbestimmung zusätzlich benötigt wird³.

Diese Beispiele sind keineswegs ein Beweis einer allgemeinen Überlegenheit der Quaternionen-Repräsentation gegenüber anderen Parametrisierungen; sie machen jedoch deutlich, daß sehr verschiedenartige Aspekte als Fürsprecher der Quaternionen-Repräsentation, vor allem auch in Hinblick auf algorithmische Lösungen zur Bewegungsschätzung zu werten sind.

2.2.3 Interpretation der Quaternionen-Parametrisierung

Die Quaternionen-Parametrisierung läßt sich dadurch interpretieren, daß die Beziehungen zu einer Rotationsmatrix aufgezeigt werden, welche durch Drehung um eine Drehachse $\mathbf{n}^T = (n_1, n_2, n_3)$, $|\mathbf{n}| = 1$ mit Drehwinkel θ definiert ist. Die Drehmatrix lautet

$$R = \cos\theta \ I + (1 - \cos\theta)\mathbf{nn}^T + \sin\theta[\mathbf{n}]_{\times} . \quad (2-20)$$

Die ersten beiden Terme sind symmetrisch, der dritte Term aufgrund (2-8,2-9) schiefsymmetrisch. Die Spur dieser Drehmatrix $\text{trace } R = 1 + 2\cos\theta$ und die Differenz $R - R^T = 2\sin\theta[\mathbf{n}]_{\times}$ lassen sich aus (2-20) direkt ablesen. Für die Spur der mit dem Einheitsquaternion repräsentierten Matrix R (2-18) und für die Differenz $R - R^T$ ergibt sich

$$\text{trace } R = 4q_0^2 - 1 \quad \text{und} \quad R - R^T = 4q_0 \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}_{\times} . \quad (2-21)$$

Das bedeutet, daß im Skalarteil $q_0 = \cos\frac{\theta}{2}$ des Einheits-Quaternions $\overset{\circ}{q}$ die Drehung um eine Achse zum Ausdruck kommt, die durch den Vektorteil $(q_1, q_2, q_3)^T = (\sin\frac{\theta}{2}\mathbf{n})$ festgelegt ist. Neben der geometrischen Interpretation ist damit auch die Vorschrift gegeben, um die beiden Parametrisierungsformen ineinander zu überführen.

Eine spezielle Quaternionen-Parametrisierung wurde in den Rechenprogrammen zur Triangulation mit unabhängigen Modellen von Ackermann et al. (1970) mit $\overset{\circ}{q} = (1, a/2, b/2, c/2)$ verwendet. Diese dreidimensionale

³Um keine zu optimistischen Erwartungen zu hegen, muß allerdings betont werden, daß zur Lösung linearisierter Orientierungsaufgaben grundsätzlich Näherungen vorzugeben sind. Werden diese Näherungen aus dem bei m Unbekannten m -dimensionalen Parameterraum willkürlich herausgegriffen, wird - abhängig von der Lösungsvielfalt (vgl. Abschnitt 2.4) - in der Regel die Lösung gefunden, die dem Startpunkt am nächsten liegt. Dabei ist keineswegs garantiert, daß hierbei die korrekte Lösung gefunden wird. Für die nichtlinearen Kleinste-Quadrate-Verfahren besteht zudem die Gefahr, in lokalen Minima hängen zu bleiben. Wir werden im Abschnitt 2.4 darauf zurückkommen.

Parametrisierung wird als Cayley-Parametrisierung (Stuelpnagel, 1964), die sich ergebende Rotationsmatrix zum Teil auch als Rodrigues-Matrix (Thompson, 1958/59) bezeichnet. Es gilt die bekannte Beziehung

$$R = (I - S)(I + S)^{-1} = I + \frac{2}{1 - (\text{trace } S^2/2)} (S^2 - S),$$

in der die schiefssymmetrische Matrix S die Komponenten a, b, c beinhaltet und S^2 für das Matrizenprodukt SS steht. Die Analogie zur Formulierung der Drehmatrix um eine vorgegebene Drehachse (2-20) ist erkennbar. Zu dem auf $q_0 = 1$ fixierten Skalarteil lautet der zugehörige Vektorteil $(q_1, q_2, q_3)^T = (\tan \frac{\theta}{2} \mathbf{n})$. Für $\theta = \pi$ divergieren die Elemente des Vektorteils, d. h. die Parameter a, b, c sind nicht mehr bestimmbar. Offensichtlich wird, daß dem Faktor $1/2$ in a, b, c (oder auch jedem anderen festen Faktor) dabei aus theoretischer Sicht keine Bedeutung zukommt. Entscheidend ist lediglich, daß der skalare Parameter des Quaternions festgehalten wird. Die Normierung eines Elements der Quaternionen-Repräsentation ist deshalb der auf das Einheitsquaternion bezogenen allgemeineren Form, für die keine solche kritische Situation existiert (s. o.), unterlegen (Shih, 1990).

2.2.4 Zusammengesetzte Rotationen

Für die Bestimmung der Rotationsparameter zusammengesetzter Drehungen bieten die Quaternionen eine sehr einfache Lösung. Mit der Bezeichnung $\overset{\circ}{q}$ für ein Quaternion, das eine Rotation R repräsentiert, und mit $\overset{\circ}{q}'$ für eine entsprechende zweite Rotation R' , ergibt sich die zusammengesetzte Rotation $R'' = R R'$ durch die Multiplikation der Quaternionen

$$\overset{\circ}{q}'' = \overset{\circ}{q} \overset{\circ}{q}', \quad (2-22)$$

beziehungsweise ausführlich

$$q''_0 + q''_1 j_1 + q''_2 j_2 + q''_3 j_3 = (q_0 + q_1 j_1 + q_2 j_2 + q_3 j_3)(q'_0 + q'_1 j_1 + q'_2 j_2 + q'_3 j_3).$$

Mit den in (2-16) formulierten Regeln folgt daraus das Gleichungssystem

$$\begin{pmatrix} q''_0 & q''_1 & q''_2 & q''_3 \end{pmatrix}^T = T_{\overset{\circ}{q}} \begin{pmatrix} q'_0 & q'_1 & q'_2 & q'_3 \end{pmatrix}^T, \quad (2-23)$$

mit der abbildenden Matrix

$$T_{\overset{\circ}{q}} = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix}. \quad (2-24)$$

Damit ist eine einfache Beziehung zur *direkten* Berechnung der Parameter zusammengesetzter Rotationen gefunden. Für die Winkelparmetrisierung existiert, wie schon eingangs erwähnt, eine solche Beziehung nicht. Die Elemente von $T_{\overset{\circ}{q}}$ sind Elemente des Einheitsquaternions. Die Matrix $T_{\overset{\circ}{q}}$ ist wie die Rotationsmatrix R orthogonal und besitzt eine Determinante $\det T_{\overset{\circ}{q}} = 1$.

Die Rotation eines Punktes \mathbf{p} kann mit dem Quaternion $\overset{\circ}{p}$ gemäß

$$\overset{\circ}{p}' = \overset{\circ}{q} \overset{\circ}{p} \overset{\circ}{q}^* \quad (2-25)$$

berechnet werden, wobei \mathbf{p} den Vektorteil des Quaternions $\overset{\circ}{p}$ darstellt, das mit einem Skalarteil $p_0 = 0$ ergänzt wurde. Das konjugierte Quaternion $\overset{\circ}{q}^* = (q_0 - q_1 j_1 - q_2 j_2 - q_3 j_3)$ repräsentiert eine inverse Rotation, denn es gilt: $\overset{\circ}{q} \overset{\circ}{q}^* = 1 + 0j_1 + 0j_2 + 0j_3$.

Die Übertragung von (2-25) auf die Matrixdarstellung (2-24) führt zu

$$\begin{pmatrix} 0 & p'_1 & p'_2 & p'_3 \end{pmatrix}^T = T_{\overset{\circ}{q}} T_{\overset{\circ}{q}^*} \begin{pmatrix} 0 & p_1 & p_2 & p_3 \end{pmatrix}^T.$$

In dem Produkt $T_q^0 T_q^0$ ist die Rotationsmatrix R (2-18) enthalten. Durch Ausmultiplizieren ergibt sich

$$T_q^0 T_q^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & R & & \\ 0 & & & \end{pmatrix},$$

womit die in (2-18) formulierte Rotationsmatrix R hergeleitet ist.

2.2.5 Ermittlung von Quaternionen-Parametern aus gegebener Rotationsmatrix

Die Bestimmung der Elemente des Einheitsquaternions aus gegebener Rotationsmatrix R ist mit (2-21) vorbereitet. Aus der Spur $\text{trace } R$ folgt q_0 , aus der schiefsymmetrischen Matrix $R - R^T$ werden die Elemente q_1, q_2, q_3 erhalten. Aufgrund der Dualität der Lösung kann $q_0 = \sqrt{(1 + \text{trace } R)/4}$ als positive Wurzel definiert werden. D. h. der Lösungsraum wird für die *Bestimmung* der Parameter aus gegebener Rotationsmatrix auf die positive q_0 -Achse (einschließlich $q_0 = 0$) beschränkt. Die Parametrisierung der Rotationsmatrix (2-18, 2-19) generell auf den bezüglich q_0 positiven Lösungsraum durch eine Forderung $q_0 \geq 0$ zu beschränken, ist nicht zulässig, da ansonsten die Beziehung (2-23) zur Berechnung zusammengesetzter Rotationen verloren geht. Ausführlich lautet die Beziehung zur Berechnung der Elemente q_1, q_2, q_3 ($q_0 \neq 0$ vorausgesetzt)

$$\begin{pmatrix} q_1 & q_2 & q_3 \end{pmatrix}^T = \frac{1}{2q_0} \begin{pmatrix} -(r_{23} - r_{32}) & (r_{13} - r_{31}) & -(r_{12} - r_{21}) \end{pmatrix}^T.$$

Für den Sonderfall $q_0 = 0$ (numerisch nahe 0) kann der Vektorteil nicht aus (2-21) berechnet werden. Hier bietet es sich an, aus den Nebendiagonalelementen von $R + R^T$

$$r_{ij} + r_{ji} = 4q_i q_j \quad \text{für} \quad i \neq j, \quad (i, j) \in \{1, 2, 3\} \quad (2-26)$$

die Parameter q_i ($i = 1, 2, 3$) zu bestimmen. Wegen der Dualität ist eine unter praktischen Gesichtspunkten nützliche Verallgemeinerung von $q_0 \geq 0$ hilfreich, z. B. die Vereinbarung, daß die erste oder die maximale Komponente $\{q_i, i = 0, 1, 2, 3\}$ positiv sein soll (Horn, 1990, S. 68).

2.2.6 Differentielle Rotationsparameter

Für die Formulierung des Bewegungsmodells (2-11) wurden Rotationsparameter ω bezogen auf das bewegte Koordinatensystem eingeführt. Die Beziehungen zu den Rotationsparametern des ortsfesten Systems ist implizit durch

$$\Omega = R^T d_t R$$

gegeben. Für die Parametrisierung von R mit den Elementen des Einheitsquaternions ergeben sich die linearen Beziehungen

$$\begin{pmatrix} 0 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = 2 \begin{pmatrix} q_0 & q_1 & q_2 & q_3 \\ -q_1 & q_0 & q_3 & -q_2 \\ -q_2 & -q_3 & q_0 & q_1 \\ -q_3 & q_2 & -q_1 & q_0 \end{pmatrix} \begin{pmatrix} d_t q_0 \\ d_t q_1 \\ d_t q_2 \\ d_t q_3 \end{pmatrix}, \quad (2-27)$$

die durch explizite Berechnung von $R^T d_t R$ und durch Koeffizientenvergleich mit Ω abgeleitet werden können. Der Vergleich mit (2-23) zeigt, daß (2-27) auch durch

$$\begin{pmatrix} 0 & \omega_1 & \omega_2 & \omega_3 \end{pmatrix}^T = 2 T_q^T \begin{pmatrix} d_t q_0 & d_t q_1 & d_t q_2 & d_t q_3 \end{pmatrix}^T,$$

formuliert werden kann, wobei T_q^T (2-24) mit der Rotationsmatrix R^T korrespondiert. Wegen der Orthogonalität von T_q^T kann die inverse Beziehung zu (2-27) mit

$$\begin{pmatrix} d_t q_0 & d_t q_1 & d_t q_2 & d_t q_3 \end{pmatrix}^T = \frac{1}{2} T_q \begin{pmatrix} 0 & \omega_1 & \omega_2 & \omega_3 \end{pmatrix}^T \quad (2-28)$$

direkt angegeben werden. Eine infinitesimale Rotation mit ω ist durch das Quaternion $\overset{\circ}{\omega} = 1 + \frac{1}{2}(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)$ darstellbar (vgl. z. B. Horn, 1990); d. h. (2-27) und (2-28) lassen sich in einfacher Weise mit den Beziehungen zur Ermittlung der Parameter zusammengesetzter Rotationen aus Quaternionen bestätigen. Für diskretes, endliches ω , wie es bei der Schätzung erhalten wird, müssen die transformierten Parameter $q'_i = q_i + d_t q_i$ ($i = 0, 3$)

$$\begin{pmatrix} q'_0 & q'_1 & q'_2 & q'_3 \end{pmatrix}^T = \frac{1}{2} T_q^{\circ} \begin{pmatrix} 2 & \omega_1 & \omega_2 & \omega_3 \end{pmatrix}^T \quad (2-29)$$

wieder entsprechend (2-19) auf 1 normiert werden. Die Normalisierung ist auch dadurch zu erzielen, daß das Quaternion $\overset{\circ}{\omega}$ durch eines der beiden Einheitsquaternionen

$$\overset{\circ}{\omega} = \sqrt{1 - \frac{1}{4}\omega^T\omega} + \frac{1}{2}(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)$$

oder

$$\overset{\circ}{\omega} = (1 + \frac{1}{2}(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)) / \sqrt{1 + \frac{1}{4}\omega^T\omega}$$

ersetzt wird.

Schießlich sollen noch die differentiellen Beziehungen zu den Winkeln ω, φ, κ angegeben werden. Trotz all der schon diskutierten Nachteile haben Winkel bislang doch zumindest einen Vorzug: Sie sind anschaulicher als algebraische Parameter und auch die Genauigkeitsangaben sind in Bezug auf Winkel vertrauter, so daß für die Darstellung von Ergebnissen klassische Winkelparametrisierungen bevorzugt werden. Die Berechnung von $R^T d_t R = \Omega$ führt zu

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} \cos\varphi \cos\kappa & \sin\kappa & 0 \\ -\cos\varphi \sin\kappa & \cos\kappa & 0 \\ \sin\varphi & 0 & 1 \end{pmatrix} \begin{pmatrix} d\omega \\ d\varphi \\ d\kappa \end{pmatrix}. \quad (2-30)$$

Aus der Inversion des Gleichungssystems folgt unter der Voraussetzung $\cos\varphi \neq 0$

$$\begin{pmatrix} d\omega \\ d\varphi \\ d\kappa \end{pmatrix} = \begin{pmatrix} \cos\kappa/\cos\varphi & -\sin\kappa/\cos\varphi & 0 \\ \sin\kappa & \cos\kappa & 0 \\ -\cos\kappa \tan\varphi & \sin\kappa \tan\varphi & 1 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}. \quad (2-31)$$

Bei der Berechnung von $(d\omega, d\varphi, d\kappa)$ aus $(\omega_1, \omega_2, \omega_3)$ darf der Winkel φ Werte um $\frac{\pi}{2} + n\pi$ ($n \in \mathbb{N}$) nicht annehmen. Diese Einschränkung der Parametrisierung $(\omega, \varphi, \kappa)$ wurde eingangs schon unter den Singularitäten der Winkelparametrisierungen diskutiert.

2.3 Projizierte Bewegung

2.3.1 Projektion

Mit der Perspektivabbildung wurde in Abschnitt 1.2.1 das geometrische Modell der Projektion des dreidimensionalen Raumes auf die Bildebene eingeführt. Gemäß Abbildung 2.3 soll die Bildebene senkrecht zur Z -Achse des im Projektionszentrum O gelagerten 3D-Koordinatensystems definiert sein. Die Zentralprojektion eines Punktes $P \mapsto P'$ führt zu den Abbildungsgleichungen (Kollinearitätsgleichungen)

$$\begin{pmatrix} x \\ y \\ c \end{pmatrix} = \lambda \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (2-32)$$

die, eliminiert man den ortsabhängigen Maßstabsfaktor $\lambda(\mathbf{X})$, äquivalent auch durch

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = c \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} \quad (2-33)$$

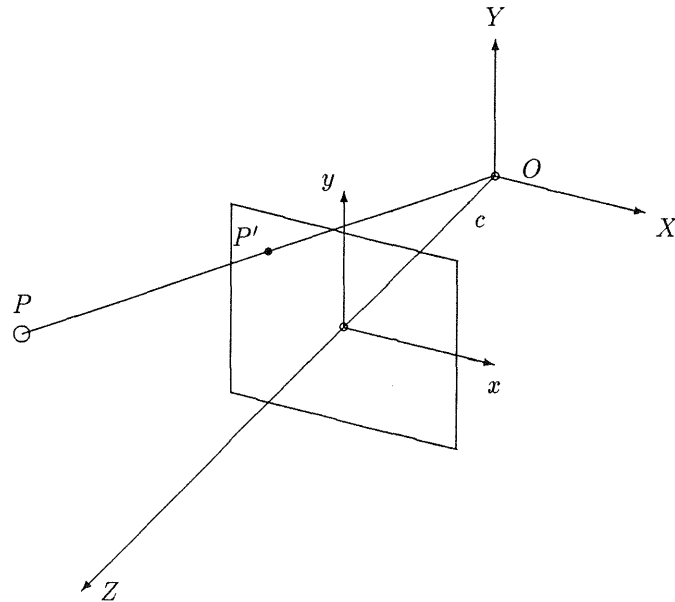


Abbildung 2.3: Kameramodell: Die Bildkoordinaten eines projizierten Punktes können in der Bildebene durch (x, y) oder im Kamerakoor­di­naten­system durch $(X = x, Y = y, Z = c)$ angegeben werden.

ausgedrückt werden können. Die Bildweite c wird auch mit der Kammerkonstanten bzw. Brennweite des optischen Systems identifiziert, da in der Regel die Tiefe $Z \gg c$ ist. Sieht man von Aufgaben im extremen Nahbereich ab, so gilt diese für die Luftbildaufnahme typische Situation auch für nahezu alle mit Digitalkameras (CCD-Sensoren) aufgenommenen Szenen. Die Kalibrierung der Kamera ist in (2-32) vorausgesetzt, so daß die Bildkoordinaten (x, y, c) bereits mit Korrekturen der inneren Orientierung (Hauptpunkt, Objektivverzeichnung, Kalibrierung der Bildweite) versehen sind.

Das Kameramodell (2-32) bildet zusammen mit dem Bewegungsmodell (2-1) die bekannten Grundgleichungen der analytischen Photogrammetrie. Um die Querverbindungen zur Literatur insbesondere im Bereich *Rechner Sehen* aufzuzeigen, ist es hilfreich, diese beiden Modelle nicht schon im Ansatz zu vermischen. Ein Übergang auf *verallgemeinerte* Bildkoordinaten durch Normierung gemäß $x/c \rightarrow x$, $y/c \rightarrow y$ ist zwar in sehr vielen Bereichen der Bildverarbeitung üblich (vgl. z. B. Jähne, 1989); hier wird allerdings darauf verzichtet, um leichter lesbar zu bleiben.

2.3.2 Differentielle projizierte Bewegung

Durch die Identifikation des Kamerakoor­di­naten­systems mit dem bewegten System (Abschnitt 2.1.1) wird $\mathbf{X} = \mathbf{X}(t)$ eine Funktion der Zeit, d. h. mit $\mathbf{X}(t)$ gemäß (2-3) eine Funktion der Bewegung des Kamerasystems. Die Projektion dieser Bewegung ist durch das Abbildungsmodell (2-32) festgelegt. Mit der Differentiation der Abbildungsgleichungen nach der Zeit ergibt sich

$$\begin{aligned} d_t x &= \frac{c}{Z} \left(d_t X - \frac{X}{Z} d_t Z \right) \\ d_t y &= \frac{c}{Z} \left(d_t Y - \frac{Y}{Z} d_t Z \right) \end{aligned}$$

bzw. in Vektorschreibweise

$$d_t \mathbf{x} = L d_t \mathbf{X}, \quad (2-34)$$

wobei die Matrix L

$$L = \lambda \begin{pmatrix} 1 & 0 & -X/Z \\ 0 & 1 & -Y/Z \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} c & 0 & -x \\ 0 & c & -y \end{pmatrix} \quad (2-35)$$

die Projektion der 3D-Geschwindigkeiten in die Bildebene angibt. Die Möglichkeiten zur Bestimmung der projizierten Geschwindigkeiten werden im Teil II (Bildzuordnung) ausführlich diskutiert. In (2-34) ist ein Differentialgleichungssystem gewonnen, das die Bewegungen im Bild- und Objektraum zueinander in Beziehung setzt. Interpretieren wir (2-34) punktbezogen, so handelt es sich um zwei Gleichungen mit drei, bei unbekannter Tiefe Z mit vier unbekanntenen Größen des Objektraumes. Die Inversion dieses unterbestimmten Systems wird also ohne weitere Annahmen über Bewegungs- oder Deformationsfreiheitsgrade des Objekts, über dessen Form, über die Anzahl der diskreten Beobachtungszeitpunkte o. ä. nicht gelingen.

In den nächsten Abschnitten wird die Starrheitsannahme zur Regularisierung des unterbestimmten Systems einbezogen und daraus die projizierte Starrkörperbewegung abgeleitet. Weitere zusätzliche Annahmen über die Form des Objektes tragen dazu bei, die Dimension des Lösungsraumes weiter einzuschränken.

2.3.3 Starrheitsannahme

Die Starrheitsannahme ist der Herleitung der differentiellen Bewegung (2-11) zugrunde gelegt, so daß sich direkt durch Einsetzen in (2-34) die projizierten Geschwindigkeiten in der Bildebene ergeben

$$d_t \mathbf{x} = L(-\boldsymbol{\omega} \times \mathbf{X} - d_t \mathbf{T}) . \quad (2-36)$$

Mit der Abbildung des Vektors \mathbf{X} auf die Matrix V entsprechend der Definition (2-9)

$$V = [\mathbf{X}]_{\times}$$

kann (2-36) auch als

$$d_t \mathbf{x} = L V \boldsymbol{\omega} - L d_t \mathbf{T} \quad (2-37)$$

formuliert werden, wobei die Elemente des Matrizenprodukts $L V$

$$L V = \frac{1}{c} \begin{pmatrix} xy & -c^2 - x^2 & cy \\ -c^2 - x^2 & xy & -cx \end{pmatrix} \quad (2-38)$$

ausschließlich Funktionen der Bildkoordinaten und der Bildweite sind. Eine weitere Darstellungsform von (2-36), die sich in starkem Maße auf das Vektorprodukt stützt, wird z. B. von Negahdaripour und Horn (1987, 1989) u. a. verwendet

$$d_t \mathbf{x}_* = -e_3 \times (\mathbf{x}_* \times (\mathbf{x}_* \times \boldsymbol{\omega} - \frac{1}{Z} d_t \mathbf{T})) .$$

Abweichend von der bisherigen Bezeichnung wird hier der Bildkoordinatenvektor dreidimensional mit $\mathbf{x}_* = (x, y, 1)^T$ entsprechend (2-32) allerdings in der Form verallgemeinerter Bildkoordinaten (siehe Abschnitt 2.3.1) eingeführt. e_3 steht hierbei für den Einheitsvektor $e_3^T = (0, 0, 1)$.

$d_t \mathbf{x}$ ist das momentane Geschwindigkeitsfeld der projizierten Bewegung in der Bildebene. Anstelle des Begriffes 'momentanes Geschwindigkeitsfeld' sind eine ganze Vielzahl von weiteren Bezeichnungen üblich, wie z. B. Bewegungsfluß, Bewegungsfeld, Bildgeschwindigkeit, Bildflußfeld, Bewegungsvektorfeld, etc.. Die Bezeichnung Vektorfeld deutet in der Regel dabei an, daß $d_t \mathbf{x}$ diskretisiert in Raum und Zeit in einem dichten, oft bildelementbezogenen Feld von Verschiebungsvektoren vorliegt. Begriffe wie optischer Fluß oder Bildfluß werden zum Teil auch für die ausschließlich geometrische Beschreibung der Bewegung in der Bildebene verwendet. Üblicher ist es aber, durch die begriffliche Unterscheidung zwischen Bewegungsfluß und optischem Fluß deutlich zu machen, daß die geometrisch definierten Geschwindigkeiten nicht notwendigerweise mit den tatsächlich visuell sichtbaren und damit beobachtbaren Geschwindigkeiten identisch sein müssen. Wir werden darauf im 3. Kapitel bei der Schätzung des optischen Flusses näher eingehen.

Die explizite parametrische Darstellung des Bewegungsflusses in Komponentenform lautet schließlich

$$\begin{aligned} d_t x &= c \left(\frac{xy}{c^2} \omega_1 - \left(1 + \frac{x^2}{c^2} \right) \omega_2 + \frac{y}{c} \omega_3 \right) + \frac{c}{Z} (-d_t T_1 + \frac{x}{c} d_t T_3) \\ d_t y &= c \left(\left(1 + \frac{y^2}{c^2} \right) \omega_1 - \frac{xy}{c^2} \omega_2 - \frac{x}{c} \omega_3 \right) + \frac{c}{Z} (-d_t T_2 + \frac{y}{c} d_t T_3) \end{aligned} \quad (2-39)$$

Aus (2-39) ist ersichtlich, daß neben den Bewegungsparametern $(\omega, d_t \mathbf{T})$ als weitere Größe des Objektraumes die Tiefe Z des Punktes eingeht. Die Projektionsgleichungen sind in den Objektraumparametern nichtlinear, da die Tiefe eines Punktes als Quotient von $d_t \mathbf{T}$ auftritt. Die erstmalige Darstellung des Bewegungsflusses in der Form (2-39) wird im Bereich Computer Vision Longuet-Higgins und Prazdny (1980, S. 389) zugeschrieben⁴. Für festes $d_t \mathbf{T}$ oder festes Z sind die Projektionsgleichungen linear, was aus der Darstellung von (2-37) gemäß

$$d_t \mathbf{x}(x, y) = L_V(x, y) \omega + L_Z(x, y) \frac{1}{Z} d_t \mathbf{T}$$

mit $L_V = L V$ und $L_Z = -L Z$ ersichtlich ist. Da die Lösung eines solchen Gleichungssystems sehr verschiedenartig angegangen werden kann (siehe Abschnitt 2.4), sind hierzu eine Fülle von Vorschlägen erarbeitet worden.

Expansions- bzw. Kontraktionspunkt

Die Einführung des Expansionspunktes gelingt anschaulich leichter, wenn vorausgesetzt wird, daß der Sensor sich ausschließlich translatorisch bewegt. In der Bildebene gibt es einen ausgezeichneten Punkt, den Fluchtpunkt einer Parallelenschar zur Bewegungsrichtung mit der Bildposition (x_F, y_F) , an dem der Bewegungsfluß identisch Null ist. Mit $d_t \mathbf{x} = 0$ ($\omega = 0$ vorausgesetzt) folgt aus (2-39)

$$(x_F, y_F) = c \left(\frac{d_t T_1}{d_t T_3}, \frac{d_t T_2}{d_t T_3} \right),$$

d. h. die momentane translatorische Bewegungsrichtung (x_F, y_F, c) wird im Bild durch die Position des Fluchtpunktes festgelegt. Für alle anderen Punkte des Bildes gilt

$$d_t \mathbf{x} = \frac{d_t T_3}{Z} \begin{pmatrix} x - x_F \\ y - y_F \end{pmatrix}.$$

Bei Bewegung auf das Objekt zu ($d_t T_3 > 0$) divergieren alle projizierten Punkte der beobachteten stationären Szene radial vom Fluchtpunkt weg, weshalb der Fluchtpunkt auch als Focus of Expansion (FOE) bezeichnet wird. Die entgegengesetzte Bewegung, weg vom Objekt ($d_t T_3 < 0$), führt zu einer Vorzeichenumkehr für die Bildgeschwindigkeiten $d_t \mathbf{x}$, d. h. der Fluchtpunkt wird zu einem Focus of Contraction (FOC).

Liegen die Bildgeschwindigkeiten $d_t \mathbf{x}$ in wenigstens zwei Punkten vor, kann der FOE als Schnittpunkt der Geschwindigkeitsvektoren ermittelt werden. Bei bekanntem FOE läßt sich ferner die relative Tiefe ($Z/d_t T_3$) in diesen Bildpunkten ableiten. Eine weitere Größe, die insbesondere im Hinblick auf die Navigation von Bedeutung ist, ist die Zeit bis zum Zusammentreffen (Time to Contact, TTC) von Sensor und Objekt, die sich gemäß

$$TTC = \frac{Z}{d_t T_3} = \frac{x - x_F}{d_t x} = \frac{y - y_F}{d_t y}$$

berechnen läßt.

Wird in die Betrachtungen die rotatorische Bewegung wieder miteinbezogen, so zeigt (2-39), daß die FOE-bezogene radiale Komponente des Bewegungsflusses von dem Effekt der Rotation überlagert wird. Ist ω bekannt, läßt sich der rotatorische Anteil in der projizierten Bildgeschwindigkeit eliminieren, so daß die oben ausgeführten Betrachtungen entsprechend für die reduzierten Geschwindigkeiten $d_t \mathbf{x} - L V \omega$ gelten. Ein Verfahren zur Schätzung des FOE bei unbekannter translatorischer und rotatorischer Bewegung wird in Abschnitt 4.2.5 diskutiert.

2.3.4 Objektmodell Ebene

Über die Form der projizierten dreidimensionalen Szene wurde bislang keine Annahme getroffen. Für die Bildzuordnung werden oft heuristisch einfache Objektformen, die zudem in ihrer Ausdehnung lokal begrenzt sind,

⁴In ähnlicher Form sind die Gleichungen in photogrammetrischen Lehrbüchern zu finden, z. B. als Parallaxgleichungen zur Bestimmung der relativen Orientierung oder als linearisierte Abbildungsgleichungen, die spezialisiert für den Fall der Senkrechtaufnahme dargestellt sind (Schwidersky und Ackermann, 1976, S.37).

vorausgesetzt. Für die punktbezogene Rekonstruktion der dreidimensionalen Szene (Kapitel 7) soll aber auf weitere Annahmen über die Form der beobachteten Welt verzichtet werden. D. h. die Interpretation der projizierten beobachteten Starrkörperbewegung besteht in den Bewegungsparametern und der punktabhängigen Tiefe.

Die Modellierung der Objekte z. B. durch Drahtmodelle, durch Oberflächenmodelle bestehend aus verknüpften Bilinearfacetten, als Polyeder, generalisierte Zylinder (vergleiche das "Zylindermännchen" in Abbildung 2.1) muß erfolgen, wenn ein abgebildetes dreidimensionales Objekt lokalisiert und identifiziert werden soll. Auch die Rekonstruktion von Oberflächenmodellen impliziert einen entsprechenden Modellansatz. Die Modellierung der Objektwelt stellt eine weitere Information dar, die als Vorinformation in den Rekonstruktions- und auch Zuordnungsprozeß eingebracht werden kann. Für die Formulierung von Zuordnungsproblemen wird als lokale Approximation eines beobachteten stationären Szenenausschnittes häufig eine ebene Fläche

$$Z = r_1 X + r_2 Y + r_3 \quad (2-40)$$

verwendet. Die Frage nach dem Bewegungsfluß dieser projizierten Ebene läßt sich dadurch beantworten, daß die Restriktion (2-40) direkt in (2-36) eingesetzt wird. Zu einer eleganteren Formulierung gelangen wir, wenn der Bezugspunkt für die rotatorischen und translatorischen Geschwindigkeiten nicht ins Projektionszentrum, sondern in den Schnittpunkt der Oberfläche mit der Tiefenachse des 3D-Bildkoordinatensystems gelegt wird (Kanatani, 1990). D. h. die Position des Bezugspunktes ist $(0, 0, r_3)$. Die Bedeutung des Bezugspunktes für die Bewegungsparameter wurde im Abschnitt 2.1.2 diskutiert. Dabei wurde dargelegt, wie auf das Projektionszentrum als Bezugspunkt transformiert werden kann. Für den Bezugspunkt $\mathbf{X}_0 = (0, 0, r_3)^T$ kann mit $\mathbf{X} - \mathbf{X}_0 = (X, Y, r_1 X + r_2 Y)^T$ die differentielle Bewegung der Ebene gemäß (2-11, 2-12) direkt angegeben werden. Die Projektion ins Bildsystem gemäß (2-34) führt schließlich zu den Differentialgleichungen

$$\begin{aligned} d_t x &= a_0 + a_1 x + a_2 y + (a_6 x + a_7 y)x \\ d_t y &= a_3 + a_4 x + a_5 y + (a_6 x + a_7 y)y \end{aligned} \quad (2-41)$$

des Bewegungsflusses der perspektiv projizierten Ebene (z. B. Tsai et al., 1982, Longuet-Higgins, 1984). Die 8 Parameter (a_0, \dots, a_7) sind als Funktionen der 6 Bewegungsparameter und der 3 Ebenenparameter z. B. in Kanatani (1990) angegeben.

Wird der Bezugspunkt ins Projektionszentrum gelegt, so ergeben sich die Parameter (a_0, \dots, a_7) zu

$$\begin{aligned} a_0 &= c(-\omega_2 - \frac{1}{r_3} d_t T_1) \\ a_1 &= \frac{1}{r_3} (+r_1 d_t T_1 + d_t T_3) \\ a_2 &= \omega_3 + \frac{r_2}{r_3} d_t T_1 \\ a_3 &= c(\omega_1 - \frac{1}{r_3} d_t T_2) \\ a_4 &= -\omega_3 + \frac{r_1}{r_3} d_t T_2 \\ a_5 &= \frac{1}{r_3} (+r_2 d_t T_2 + d_t T_3) \\ a_6 &= -\frac{1}{c} (\omega_2 + \frac{r_1}{r_3} d_t T_3) \\ a_7 &= \frac{1}{c} (\omega_1 - \frac{r_2}{r_3} d_t T_3) \end{aligned} \quad (2-42)$$

Der Beweis kann durch Einsetzen der Ebenengleichung (2-40) in (2-39) erfolgen, wobei die Ebenengleichung als Funktion der Bildkoordinaten lautet:

$$Z = r_3 (1 - r_1 \frac{x}{c} - r_2 \frac{y}{c})^{-1} .$$

Geeignetes Zusammenfassen führt zu den Beziehungen (2-41, 2-42). Das Geschwindigkeitsfeld ist in den Bildkoordinaten ein Polynom zweiter Ordnung. Beziehung (2-41) hat für ein ebenes Objekt globale Gültigkeit. Darüberhinaus ist sie für schwach gekrümmte Oberflächen eine lokale Approximation.

Approximiert man zur Beschreibung des Bewegungsflusses einer Ebene die Perspektivabbildung durch eine orthographische Projektion ($x = X, y = Y$), so reduziert sich (2-41) zu

$$\begin{aligned} d_t x &= a_0 + a_1 x + a_2 y \\ d_t y &= a_3 + a_4 x + a_5 y. \end{aligned} \quad (2-43)$$

Die Parameter des affinen Modells lauten als Funktion der Bewegungsparameter

$$\begin{aligned} a_0 &= -\omega_2 r_3 - d_t T_1 \\ a_1 &= -\omega_2 r_1 \\ a_2 &= -\omega_2 r_2 + \omega_3 \\ a_3 &= -\omega_1 r_3 - d_t T_2 \\ a_4 &= \omega_1 r_1 - \omega_3 \\ a_5 &= -\omega_1 r_2. \end{aligned} \quad (2-44)$$

Aus (2-42) ist ersichtlich, daß bei der perspektiven Projektion die translatorischen Geschwindigkeiten $d_t T_i$ und die Tiefe r_3 der Ebene mit demselben Faktor multipliziert werden können, ohne daß dies einen Einfluß auf die Beziehung zu den Parametern a_i hat. D. h. es besteht eine Maßstabsunbestimmtheit, auf die im nächsten Abschnitt bei der Diskussion der Rekonstruktionsverfahren, z. B. der relative Orientierung, weiter eingegangen wird.

Bei der orthographischen Projektion ist die Unterbestimmtheit noch wesentlich stärker ausgeprägt. Zum Beispiel kommt die Tiefenkomponente $d_t T_3$ in (2-44) explizit nicht vor. Um dennoch Bewegung und Struktur zu bestimmen, müssen wenigstens drei Bilder einbezogen werden. Bei diskreter punktbasierter Betrachtung hat Ullman (1979) gezeigt, daß bei 4 Punktkorrespondenzen über 3 orthographische Bilder Bewegungs- und Strukturparameter bestimmbar sind.

Im Gegensatz zu Ullman interpretieren Negahdaripour und Lee (1992) die Beziehungen (2-43) als Approximation von (2-41), wobei die quadratischen Glieder vernachlässigt wurden. Liegen Messungen des Bewegungsflusses $d_t \mathbf{x}$ innerhalb eines regelmäßig gerasterten Bildausschnittes diskret vor, so zeigen Negahdaripour und Lee, daß die Parameter a_6 und a_7 gegenüber den Parametern ($a_0 - a_5$) um ein Vielfaches ungenauer geschätzt werden. Sie verzichten deshalb für die Rekonstruktion der Bewegungs- und Strukturparameter aufgrund von (2-42) auf a_6 und a_7 und schlagen stattdessen vor, die 6 Parameter ($a_0 - a_5$) in zwei Bildregionen unabhängig zu ermitteln. Sind die beiden Regionen Abbildungen verschiedener Objektebenen mit sich unterscheidenden Orientierungen, so stehen für die Ermittlung der 12 auf den Objektraum bezogenen Parameter (6 Bewegungs-, 2×3 Strukturparameter) die 12 "beobachteten" (2×6) Parameter des Bewegungsflusses gegenüber. Wegen der bereits angesprochenen generellen Maßstabsunbestimmtheit verbleiben 11 objektraumbezogene Parameter, die, von singulären Fällen abgesehen, im allgemeinen eindeutig bestimmt werden können. Negahdaripour und Lee (1992) haben hierfür eine direkte Lösung aufgezeigt.

Die Transformation (2-43) hat, da es sich um ein sehr einfaches lineares Modell handelt, für die geometrische Modellierung bei der Bildzuordnung eine weite Verbreitung gefunden.

2.3.5 Lokal quadratische Oberflächen

Eine entsprechende Vorgehensweise, wie sie im vorhergehenden Abschnitt für die Ebene vorgestellt wurde, haben Waxman et al. (1987) angewendet, um den Einfluß lokaler quadratischer Oberflächenform für den Bewegungsfluß zu diskutieren. Wird die Fläche durch

$$Z = r_0 + r_1 X + r_2 Y + \frac{1}{2} r_3 X^2 + \frac{1}{2} r_4 Y^2 + r_5 XY$$

dargestellt, erfolgt durch die Normierung dieser Gleichung mit Z gemäß (2-33) der Übergang auf Bildkoordinaten. Dabei erscheint die Tiefe Z erneut in den Gliedern, die quadratisch in den Bildkoordinaten sind. Die Substitution von Z als Funktion der Bildkoordinaten führt zu

$$Z^{-1} = r_0^{-1} \left(1 - r_1 \frac{x}{c} - r_2 \frac{y}{c} - \frac{1}{2} r_3^* \left(\frac{x}{c} \right)^2 - \frac{1}{2} r_4^* \left(\frac{y}{c} \right)^2 - r_5^* \left(\frac{xy}{c^2} \right) \right) + \mathcal{O}(x^3, \dots),$$

wobei Glieder 3. und höherer Ordnung in $\mathcal{O}(x^3, \dots)$ zusammengefaßt sind. Die normalisierten Krümmungen (r_j^*) sind durch

$$(r_3^*, r_4^*, r_5^*) = r_0 (r_3, r_4, r_5)$$

definiert. Durch die Vernachlässigung der Glieder 3. und höherer Ordnung liegt die Formbeschreibung der Oberfläche durch ein quadratisches Modell in den Bildkoordinaten vor. Wird dieses approximierende Modell der Oberfläche in die Beziehungen für den Bewegungsfluß (2-39) eingesetzt, ergeben sich

$$\begin{aligned} d_t x &= a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \\ d_t y &= a_6 + a_7 x + a_8 y + a_9 x^2 + a_{10} xy + a_{11} y^2 . \end{aligned} \quad (2-45)$$

Dabei wurden erneut Glieder 3. Ordnung vernachlässigt.

Während (2-41) den Bewegungsfluß einer perspektiv projizierten Ebene exakt beschreibt, ist in (2-45) eine Approximation des Bewegungsflusses bis zum 2. Grad gegeben, der von einer lokalen quadratischen Fläche zweiter Ordnung erzeugt wird. D. h. im Gegensatz zu (2-41) besitzt (2-45) nur lokale Gültigkeit, was mit der lokalen Approximation der quadratischen Fläche beim Übergang auf Bildkoordinaten einhergeht. Auf die Angabe der Beziehungen zwischen den 12 Parametern des Bewegungsflusses ($a_0 - a_{12}$) und den 3D-Struktur- und Bewegungsparametern ($r_0 - r_5, \omega, d_t \mathbf{T}$) kann an dieser Stelle verzichtet werden, da im weiteren kein Gebrauch davon gemacht wird. Stattdessen sei auf Waxman et al. (1987) verwiesen. Direkte Lösungsverfahren zur Berechnung der 3D-Struktur und Bewegung werden von Waxman sowohl für die Ebene als auch für die lokal quadratische Fläche aufgezeigt.

Schon die Darstellung der quadratischen Fläche (via Z^{-1}) als Funktion der Bildkoordinaten legt nahe, daß die gekrümmte Fläche durch eine Ebene approximiert werden kann, wenn die normalisierten Krümmungen klein sind, d. h. das Produkt aus lokaler Krümmung und Objektdistanz klein ist. Ansonsten ist sind aber die Voraussetzungen nicht gegeben, den Bewegungsfluß durch weniger als 12 Parameter zu approximieren (Waxman et al., 1987).

2.4 Rekonstruktionsansätze

In den vorangehenden Abschnitten wurde primär das Ziel verfolgt, Bewegung bis hin zur projizierten Starrkörperbewegung zu beschreiben. Die dabei abgeleiteten Beziehungen umfassen im wesentlichen alle Gleichungen, die als Grundlage für die Rekonstruktion von Belang sind.

Die Formulierung von Rekonstruktionsansätzen ist hauptsächlich davon abhängig

1. welche beobachteten Daten zur Rekonstruktion zur Verfügung stehen
2. welche Annahmen über die Form der Szene getroffen werden
3. welche Anwendung verfolgt wird, bzw. welche Beschränkungen bzgl. der Bewegungsfreiheitsgrade getroffen werden
4. wieviel Aufnahmen die Sequenz umfaßt .

Diese Kriterien sind nicht unabhängig voneinander, so daß sie im folgenden auch zusammenhängend diskutiert werden sollen. Die Gemeinsamkeit aller Verfahren besteht im vorgegebenen Ziel, d. h. in der Bestimmung von rotatorischen und translatorischen Bewegungsparametern des sich bewegenden Sensors und - wenn auch nicht in allen Ansätzen - der Rekonstruktion von 3D-Struktur der abgebildeten Szene. Werden über die Stationarität hinaus keine weiteren Annahmen über die Form der Szene getroffen (z. B. durch Objektmodelle, Oberflächenmodelle), so besteht die zu rekonstruierende Strukturinformation vorwiegend in einer 3D-Punktmenge, die auch als Tiefenkarte bezeichnet wird, wenn sie im bewegten Sensorsystem bildelementbezogen vorliegt.

2.4.1 Die beobachteten Daten

Die primär erfaßten Daten einer Sequenz von Bildern sind die Intensitätswerte in der aufgerasterten Bildebene, einschließlich ihrer zeitlichen Dimension. Naheliegender wäre deshalb, die Intensitätswerte direkt als Beobachtungen zur 3D-Rekonstruktion von Bewegung und Struktur heranzuziehen. Ein solcher Weg ist im Bereich der Stereorekonstruktion z. B. von Smith (1986) zur Rekonstruktion von Oberflächenprofilen aus Stereobildern (allerdings bei bekannter Orientierung) und Wrobel (1987) zur Rekonstruktion von Oberflächenmodellen aus Bildverbänden vorgeschlagen worden. Wrobel bezieht neben geometrischen Objektmodellen auch die Modellierung physikalischer Parameter (Beleuchtung, Reflexion) mit ein, was eine verfeinerte Modellierung der Bildentstehung, aber auch eine deutliche Erhöhung des Rekonstruktionsaufwandes impliziert.

Entsprechende Entwicklungen im Bereich der Bildsequenzen sind von Horn und Weldon (1988), Negahdari-pour und Horn (1987, 1989), Taalebinezhad (1990), Michael (1990) vorgeschlagen worden. Im Gegensatz zu den auf Korrespondenz oder auf optischem Fluß beruhenden Verfahren werden bei diesen "direkten" Methoden Bewegungs- und Strukturparameter direkt aus den Bildintensitäten der Sequenz rekonstruiert. Neben der Starrkörperhypothese liegt den direkten Verfahren die Restriktion zugrunde, daß zwischen den Aufnahmezeitpunkten einer Bewegungseinheit keine signifikanten Beleuchtungsänderungen vorkommen dürfen. Der temporale Bildgradient läßt sich dadurch zu den räumlichen Bildgradienten und zur Bewegung des Bildpunktes im Bild bzw. zu den 3D-Bewegungs- und Strukturparameter in Beziehung setzen. Im Kapitel 3 wird diese Restriktion näher betrachtet.

Die Paradigmen der Bildanalyse, wie sie vor allem Ende der 70er Jahre (Marr, 1978, 1982, Barrow und Tennenbaum, 1978) formuliert wurden und bis heute noch breite Zustimmung erfahren⁵, haben sich in starkem Maße auch in den Arbeiten im Bereich Bildfolgenanalyse niedergeschlagen. Daraus leitet sich die Anschauung her, daß die Schätzung der Bewegung in der Bildebene als eigenständiges Problem zuerst zu lösen ist und dann in einem zweiten Schritt dieses Zuordnungsergebnis als Beobachtung für die Rekonstruktion verwendet wird. Ein Vorteil einer solchen modularen Betrachtung ist, daß die Einzelaufgaben leichter zu durchdringen sind und deshalb gut studiert werden können. Darüberhinaus beinhaltet die Modularisierung auch die Möglichkeit, durch die Auswahl bzw. die Formulierung der Einzelkomponenten Einflüsse weitgehend zu eliminieren, die z. B. von Reflexions- oder Beleuchtungsunterschieden herrühren. Dies kann durch den Übergang von der ikonischen zur symbolischen Bildrepräsentation erfolgen oder auch dadurch, daß diese physikalischen Effekte als Bias-Parameter approximiert und im Zuordnungsmodell mitgeschätzt werden. In beiden Fällen kann die nachfolgende 3D-Rekonstruktion ausschließlich auf die Geometrie beschränkt bleiben.

Unabhängig davon, ob die Zuordnung auf ikonischer oder symbolischer Ebene durchgeführt wird, besteht das Ziel der Zuordnung darin, projizierte Geschwindigkeit für den Rekonstruktionsprozeß zu erfassen. Wegen der Diskretisierung in Raum (Bildebene) und Zeit treten an die Stelle der Geschwindigkeiten Verschiebungsvektoren bzw. -vektorfelder. Die Verfahren zur Schätzung des optischen Flusses (Kapitel 3) geben in der Regel ein dichtes, pixelbezogenes Verschiebungsvektorfeld an. Als beobachtete Meßdaten gehen in der Mehrzahl der Ansätze die räumlichen und zeitlichen Intensitätsgradienten in den Schätzprozeß ein. Findet die Zuordnung auf der symbolischen Ebene (Kapitel 4) statt, müssen zunächst Merkmale aus dem Bild extrahiert werden, die dann als beobachtete Daten in die Korrespondenz eingehen. Ein zunehmender Komplexitätsgrad der Merkmale hat in der Regel eine geringere Zahl von Merkmalen und daher auch von Verschiebungsvektoren zur Folge. Die überwiegende Zahl der vorgeschlagenen korrespondenzbasierten Verfahren verwendet Punktmerkmale und

⁵Die als monolithische Verfahren bezeichneten Methoden, bei denen das Ergebnis (die Interpretation) direkt aus den primär erfaßten Daten berechnet wird, haben die Forschung im Bereich Computer Vision bis in die 70er Jahre dominiert. Danach kam es zu einer Neuorientierung insbesondere durch die Arbeiten von Marr, aber auch Barrow und Tennenbaum u. a., für die auch der Begriff des signals-to-symbols Paradigmas steht. Diese Methoden, die in starkem Maße durch biologische Betrachtungen motiviert sind, sehen eine Zerlegung der gesamten Aufgabe in Teile vor, die mit einer Hierarchie von semantisch bedeutungsvollen Repräsentationsformen einhergeht. Seit Mitte der 80er Jahre deutet sich ein Trend zurück zu den monolithischen Verfahren an (Fischler und Firschein, 1987, S. 512), was sich im Bereich der Bildsequenzen vor allem in den Arbeiten der MIT-Gruppe um Prof. Horn (s.o.) bestätigt. Die Methodologie hinter Marr's Arbeiten besteht in dem allgemeinen Prinzip zum Entwurf großer komplexer Informationssysteme. Zunächst wird das System in funktionale Komponenten unterteilt und jede Komponente analysiert. Danach muß die Repräsentation von Information in den Komponenten und der Informationsaustausch zwischen den Komponenten gewählt werden. Schließlich kann die Überprüfung des Gesamtsystems in seinen Details erfolgen. Der Trend heute zur Vereinheitlichung und Integration der einzelnen Komponenten kann als Systemverbesserung (Robustifizierung) betrachtet werden, bei der die Integration die direkte Beziehung von den Primärdaten hin zur Interpretation herstellt (Aloimonos, 1989).

deren zugehörige Verschiebungsvektoren. Einige Ansätze nützen als beobachtete Merkmale korrespondierende Linien; auch Ecken, die durch orthogonalen Kanten definiert sind, konische Kreisbögen, Kombinationen von Punkten und Linien und einige weitere Kombinationen von Merkmalen sind in einzelnen Arbeiten zur Rekonstruktion bislang einbezogen worden. Entsprechend den Merkmalen ändern sich auch die beobachteten Größen. Zum Beispiel treten an die Stelle der Position von Punktmerkmalen die Richtungskoeffizienten bei Geraden. Die Wahl des Merkmalstyps hat insbesondere Auswirkungen auf die Strukturparameter. Anstelle der Punktposition oder Tiefe werden bei Geraden z. B. 4 Geradenparameter, alternativ dazu auch die Richtung in 3D und ein Punkt der Geraden (5 Parameter) als Strukturparameter verwendet oder es werden bewegungsinvariante Größen wie der Abstand und der Winkel zwischen zwei Linien des Objekts gewählt. Eine Übersicht über die korrespondenzbasierten Ansätze gibt Huang (1990).

2.4.2 Lösungswege und Anwendungen bei geringer Anzahl von Bildern

Vorschläge zur Lösung von Rekonstruktionsaufgaben sind hauptsächlich für das Monobild und für die Zwei- und Dreibildeinheit ausgearbeitet und untersucht worden. Charakteristisch für die Situation des Monobildes ist, daß die Korrespondenz von Merkmalen im 3D-Objektraum und deren Projektion in der 2D-Bildebene als gegeben vorausgesetzt werden. Die 3D-Merkmale sind Paßpunkte, -linien oder -flächen, die allgemeiner z. B. in Form komplexer Objektmodelle vorliegen können. Sind Modelle von Objekten gegeben, so besteht die Aufgabenstellung in der Regel darin, diese Objekte zu identifizieren und zu lokalisieren. Im Kontext der passiven Navigation kann der Lokalisierschritt in der Rekonstruktion von Orientierungsstartwerten für eine Bewegungsbestimmung bestehen. Die Startwerte legen den Bezug zu einem übergeordneten Referenzkoordinatensystem fest, was auch als Datumsgebung⁶ bezeichnet wird. Das Datum kann bei vielen Anwendungen frei festgelegt werden; in der Regel dadurch, daß das Bezugssystem bezüglich Position und Richtung mit dem Bildkoordinatensystem des ersten Bildes der Sequenz identisch gewählt wird und die Bewegungsbasis zum zweiten Bild die Länge 1 erhält. Eine weitere Anwendung der passiven Navigation ist gegeben, wenn diese sich ausschließlich auf Paßinformation des Objektraumes stützt. In einer solchen durch Paßinformation bekannten Szene reicht im Prinzip eine sukzessive Monobildauswertung, um die Bewegung des Sensors bezüglich Position und Richtung zu bestimmen.

Monobild

Für die Berechnung der äußeren Orientierung eines Bildes aus 3D-2D Punktkorrespondenzen hat sich in der Photogrammetrie der räumliche Rückwärtsschnitt als iterative Lösung eines überbestimmten, linearisierten, bei Winkelparametrisierungen transzendenten Gleichungssystems etabliert (Schwidefsky und Ackermann, 1976). Der Nachteil dieser Verfahren ist, daß in der Regel geeignete Näherungswerte vorausgesetzt werden müssen. Als Alternative, die heute mehr als Komplement zur Näherungswertbestimmung für die nichtlinearen Verfahren betrachtet werden, sind direkte Lösungen (Fischler und Bolles, 1981, Grafarend et al., 1989) für 3 Punkte in allgemeiner Lage vorgeschlagen worden. Die Besonderheit des von Fischler und Bolles dargestellten Verfahrens ist dessen Einbindung in ein robustes Konzept, das als RANSAC (random sample consensus) bezeichnet wurde. Hinter RANSAC steht die Idee, mit einem Minimum an Korrespondenzen die Lösung zu berechnen. Die direkte Lösung wird solange wiederholt und jeweils auf Konsistenz mit den in der aktuellen Berechnung nicht involvierten Punkte geprüft, bis sich eine stabilisierte Aussage einstellt. Eine weitere direkte Lösung, die DLT (Direct Linear Transform Solution), wurde von Marzan und Karara (1975) vorgeschlagen. Sie überführen die nichtlinearen 3D-2D Beziehungen in ein homogenes Gleichungssystem, das linear in den 9 Elementen r_{ij} der Rotationsmatrix und den 3 Positionskoordinaten ist. Zu diesen 12 Unbekannten trägt jede Punktkorrespondenz zwei Gleichungen bei, so daß zur Lösung mindestens 6 Punktkorrespondenzen benötigt werden. Während bei der direkten Lösung mit 3 Punkten bis zu 4 reelle Lösungen existieren, ist die direkte lineare Lösung eindeutig, wenn auch nicht optimal, da die nichtlinearen Orthogonalitätsrestriktionen (2-14) auf die Rotationsmatrix nicht einbezogen sind. Lösungswege für 3D-2D Linienkorrespondenzen sind z. B. von Liu et al. (1988) aufgezeigt worden. Rotation und Translation lassen sich dabei leicht separieren, so daß nichtlineare Lösungen mit 3, lineare

⁶Da eine Serie von Bildern bzw. ein Bildverband nach der Korrespondenz geometrisch einem Netz von Richtungsbündeln entspricht, ist die Datumsgebung (d. h. die Lagerung des Netzes) wie bei 3D-Richtungsnetzen durch 3 Translations-, 3 Rotations- und 1 Maßstabsparameter festgelegt. Zur Bestimmung der Startwerte für die Translation und Rotation ist ein Bild ausreichend, für die Maßstabsübertragung in das Richtungsnetz ein zweites Bild notwendig.

Lösungen mit 8 Linienkorrespondenzen erhalten werden.

Zwei- und Dreibildeinheit

Die größten Anstrengungen sind bislang eindeutig auf die Zwei- und Dreibildeinheit konzentriert. Der einfachere Fall dabei ist die 3D-3D Korrespondenz, d. h. die Merkmale sind hinsichtlich ihrer Position in einem dreidimensionalen Raum gegeben (z. B. aus einer Stereoeinheit erfaßt). Für die Schätzung der Rotation und Translation zwischen den Bezugszeitpunkten existieren direkte Lösungen bei der Parametrisierung der Rotationsmatrix mit Quaternionen (vgl. z. B. Faugeras und Hebert, 1983). Sieht man von einem Maßstabsfaktor ab, entspricht die Aufgabenstellung der absoluten Orientierung in der Photogrammetrie, für die ebenfalls direkte Lösungen vorgeschlagen wurden (z. B. Sanso, 1973, Horn, 1987, Umeyama, 1991). Im allgemeinen deutlich anspruchsvoller ist die Situation, wenn die Korrespondenz sich auf 2D-Merkmale beschränkt. Für 2D-Punktkorrespondenzen sind der Translationsvektor und die Distanzen zum Objekt bis auf einen Maßstabsfaktor bestimmbar, wenn wenigstens 5 Punkte in allgemeiner Lage vorliegen. Photogrammetrisch ist die Aufgabenstellung als relative Orientierung bekannt. Der Begriff der relativen Orientierung ist in letzter Zeit auch im Bereich Rechner Sehen aufgegriffen worden (Horn, 1990). Die Anzahl der vorgeschlagenen Verfahren ist beträchtlich. Dies liegt zum einen daran, daß mit der Entwicklung möglichst effizienter Algorithmen Fragen der Existenz, der Eindeutigkeit der Lösung oder der Obergrenze für die Lösungsvielfalt der Verfahren nicht leicht beantwortet werden können. Daneben ist die Qualität der Lösung, d. h. die Sensitivität der Lösung bezüglich dem Meßrauschen eine Schwerpunktfrage. Versucht man die Rekonstruktionsansätze zusammenzufassen, kann dies nach verschiedenen Gesichtspunkten erfolgen (Huang und Netravali, 1990):

(1) “structure first” vs. “motion first”

Vorschläge die vorsehen, die Strukturparameter zuerst zu bestimmen, sind nur wenige unterbreitet worden. Zuerst liegt die Idee, daß die Strecken zwischen den Positionen der Objektpunkte bewegungsinvariant sind, so daß hieraus Restriktionen für die Rekonstruktion formuliert werden können. Die große Mehrzahl aller Verfahren versucht zuerst die Bewegungsparameter zu bestimmen; einige wenige Ansätze sehen eine simultane Rekonstruktion von Struktur und Bewegung vor. Der im nachfolgenden Abschnitt (2.4.3) formulierte Rekonstruktionsansatz (2-50) fällt in diese dritte Kategorie.

(2) Transzendente, polynomische oder lineare Gleichungen

Der mathematische Hintergrund praktisch aller Verfahren besteht in der Lösung eines transzendenten, eines polynomischen (algebraischen) oder eines linearen Gleichungssystems. Dies gilt neben den Punktmerkmalen auch für die anderen Merkmalstypen und darüberhinaus für die 3D-2D und 3D-3D Merkmalskorrespondenzen.

Transzendente und polynomische Ansätze

Aufgrund der Bewegungsgleichungen (2-4) und der perspektiven Projektion der Punkte gemäß (2-32) ergeben sich die Beziehungen

$$\begin{aligned} \frac{x_2}{c} &= \frac{r_{11}x_1 + r_{12}y_1 + r_{13}c + T_1 \frac{c}{Z_1}}{r_{31}x_1 + r_{32}y_1 + r_{33}c + T_3 \frac{c}{Z_1}} \\ \frac{y_2}{c} &= \frac{r_{21}x_1 + r_{22}y_1 + r_{23}c + T_2 \frac{c}{Z_1}}{r_{31}x_1 + r_{32}y_1 + r_{33}c + T_3 \frac{c}{Z_1}} \end{aligned} \quad (2-46)$$

Die Bildkoordinaten der Punktkorrespondenzen sind zum Zeitpunkt t_1 durch $\mathbf{x}_1 = (x_1, y_1)^T$, zum Zeitpunkt t_2 durch $\mathbf{x}_2 = (x_2, y_2)^T$ dargestellt. $\mathbf{T} = (T_1, T_2, T_3)^T$ steht für die Translation zwischen den Bezugszeitpunkten (in Gl.(2-4) durch $\mathbf{X}_0(t_0, t_1)$ ausgedrückt) und Z_1 ist die Tiefenkomponente zum Bildpunkt \mathbf{x}_1 . Die Elimination der Tiefe aus (2-46) führt zu einer Gleichung für jeden korrespondierenden Punkt

$$\begin{aligned} (x_2 T_3 - c T_1)(r_{21}x_1 + r_{22}y_1 + r_{23}c - \frac{y_2}{c}(r_{31}x_1 + r_{32}y_1 + r_{33}c)) = \\ (y_2 T_3 - c T_2)(r_{11}x_1 + r_{12}y_1 + r_{13}c - \frac{x_2}{c}(r_{31}x_1 + r_{32}y_1 + r_{33}c)) , \end{aligned}$$

die bei Winkelparametrisierungen transzendent in den Winkeln und linear und homogen in den Translationsparametern ist. In den 9 unbekanntenen Elementen r_{ij} der Rotationsmatrix R und den 3, bzw. wegen des freien

Maßstabes 2, unbekanntem Translationen ist die Gleichung ein Polynom zweiten Grades. Nimmt man die 6 quadratischen Restriktionen auf die Rotationsmatrix (aus 2-14) hinzu, sind wiederum 5 Punktkorrespondenzen nötig, um ein nicht unterbestimmtes Gleichungssystem zu erhalten. Zur Lösung dieser nichtlinearen Gleichungen sind iterative Methoden entwickelt worden (Huang, 1990). Ziemlich ungeklärt sind dabei die Auswirkungen der Stochastizität der beobachteten Größen unter den entsprechenden Minimierungskriterien auf die ermittelten Parameter und damit die Frage nach der Optimalität der Lösung bzw. deren Sensitivität gegen Fehler in den Beobachtungen.

Hinter den polynomischen Ansätzen steht die Idee, die Anzahl der unbekanntem Parameter zu reduzieren, was impliziert, daß aus den quadratischen Polynomen Polynome höherer Ordnung in weniger, im günstigsten Fall einer Unbekanntem werden. Jerian und Jain (1988) leiten aus 3 Punktkorrespondenzen ein Polynom 4. Grades in 3 Rotationsparametern her, wobei die spezielle Quaternionen-Parametrisierung mit fixiertem Skalarteil $q_0 = 1$ (Abschnitt 2.2) verwendet wurde. Mit 5 Punkten liegen dann 3 Gleichungen in 3 Unbekanntem vor. Eine weitere Einschränkung des dreidimensionalen Suchraumes bei Vorgabe einer Variablen gelingt Jerian und Jain, indem sie zwei Polynome 4. Grades in 2 Parametern auf ein Polynom 16. Grades in einer Unbekanntem reduzieren. Ein Argument gegen die transzendenten oder polynomischen Gleichungssysteme ist, daß die Anzahl der Lösungsmehrdeutigkeiten nur schwer ermittelt, oft nur sehr grob nach oben abgeschätzt werden kann (z. B. bei Jerian und Jain auf $\leq 4 \cdot 16 = 64$). Deshalb war das Bestreben groß, auch für die Stereokorrespondenz lineare Algorithmen zu entwickeln, womit eindeutige Lösungen gewonnen werden.

Lineare Ansätze

Die linearen Algorithmen beruhen auf der Koplanaritätsbedingung, deren zentrale Bedeutung für die relativen Orientierungsverfahren der Photogrammetrie schon sehr früh (z. B. Schut, 1957/58) erkannt wurde. Wird ein Punkt im bewegten 3D-Koordinatensystem zum Zeitpunkt t_1 mit \mathbf{X}_1 , zum Zeitpunkt t_2 mit \mathbf{X}_2 , die Rotation und Translation zwischen den Bezugszeitpunkten mit R und \mathbf{T} bezeichnet, entsprechend den Vereinbarungen für die nichtlinearen Ansätze (2-46), so lautet die Koplanaritätsbedingung

$$\mathbf{X}_1^T (\mathbf{T} \times R\mathbf{X}_2) = 0 . \quad (2-47)$$

Mit der Abbildung von \mathbf{T} auf die schiefsymmetrische Matrix $[\mathbf{T}]_{\times}$ gemäß (2-9) folgt

$$\mathbf{X}_1^T [\mathbf{T}]_{\times} R\mathbf{X}_2 = \mathbf{X}_1^T E\mathbf{X}_2 = 0 . \quad (2-48)$$

Die Gleichung ist homogen, so daß durch Normierung mit dem jeweiligen ortsabhängigen Maßstabsfaktor (2-32) der Übergang auf die Bildkoordinaten erfolgt

$$\begin{pmatrix} x_1 \\ y_1 \\ c \end{pmatrix}^T E \begin{pmatrix} x_2 \\ y_2 \\ c \end{pmatrix} = 0 . \quad (2-49)$$

Die Matrix E hat die Dimension 3×3 , d. h. sie besteht aus 9 unabhängigen Elementen. In diesen 9 Unbekanntem ist (2-49) linear und homogen, d. h. diese Gleichung ist im allgemeinen bei wenigstens 8 Punktkorrespondenzen (bis auf einen Maßstabsfaktor) eindeutig⁷ bestimmbar. Für den überbestimmten Fall ergibt sich bei n Punktkorrespondenzen aus (2-49) ein redundantes Gleichungssystem

$$F\mathbf{e} = \mathbf{v} ,$$

mit

$$F = F(\{\mathbf{x}_1, \mathbf{x}_2\}_j, j = \{1, n\}) \quad \text{und} \quad \text{vec} E = \mathbf{e} ,$$

das durch Kleinste-Quadrate-Minimierung ($\mathbf{v}^T \mathbf{v} = \text{Min}$) zur direkten Schätzung der Elemente e_i der E -Matrix führt. Ist E bestimmt, können R und \mathbf{T} durch eine Eigenwertzerlegung von E bestimmt werden (Weng et al.,

⁷Die wegen der Homogenität des Gleichungssystems bis auf einen Maßstabsfaktor eindeutige Bestimmbarkeit der Elemente der E -Matrix wird auch als "essentielle" Schätzbarkeit bezeichnet, die Elemente von E deshalb als essentielle Parameter (Tsai und Huang, 1984).

1989), womit eine direkt gelöste Kleinste-Quadrate-Anpassung von R und \mathbf{T} an die (stochastisch) geschätzten Elemente der E -Matrix erzielt wird (Philip, 1991). Die direkte lineare Lösung wurde durch Longuet-Higgins (1981) im Bereich Rechner Sehen eingeführt und ist heute unter dem Begriff der Bestimmung der E -Matrix sehr populär. In der Photogrammetrie hat Stefanovic (1973) praktisch denselben Lösungsweg aufgezeigt, ohne damit allerdings große Resonanz zu erzielen. Schon zuvor hat Rinner (1963) den linearen Ansatz vorgeschlagen, diesen dann um die Orthogonalitätsrestriktionen erweitert, womit sich ein polynomisches Gleichungssystem ergibt. Polynomische Restriktionen in der Komponenten der E -Matrix und ihre Effekte für die relative Orientierung werden z. B. von Huang und Faugeras (1989) diskutiert. Jedoch wird zunehmend häufiger Kritik an den direkten Lösungen geübt. Die linearen Methoden arbeiten in der Regel gut bei geringem Rauschen und einer großen Anzahl von Punktkorrespondenzen (Weng et al., 1989). Sie sind allerdings gegenüber groben Fehlern sehr empfindlich, d. h. die linearen Verfahren besitzen eine eher geringe Zuverlässigkeit. Insbesondere ist die Wirkung der Meßfehler auf die Strukturparameter nahezu unabhängig von der Punktzahl. Die geschätzten Bewegungsparameter sind weniger korrekt bei kleiner Punktzahl und vor allem auch bei kleiner Bewegungsbasis, da dann das lineare Gleichungssystem schlecht konditioniert ist.

Ansätze zur relativen Orientierung bei Linienkorrespondenzen über 3 Bilder sind bislang nur wenige vorgestellt worden. Eine geschlossene, lineare Lösung, die auf die Bestimmung essentieller Parameter zurückgeführt wird, haben sowohl Spetsakis und Aloimonos (1990) als auch Weng et al. (1992) präsentiert.

Wir wollen an dieser Stelle die Rekonstruktionsansätze bei gegebenen (beobachteten) Merkmalskorrespondenzen nicht weiter vertiefen. Beschränkt auf korrespondierende Punkte und einfache Luftbildkonfiguration (Senkrechtaufnahme, übliche Brennweiten und Stereoüberdeckung) sind Verfahren der relativen Orientierung oder des räumlichen Rückwärtsschnittes längst photogrammetrische Praxis. Da die Mono- oder Stereo-Orientierungsverfahren oft nur zur automatischen Näherungswertgewinnung für Bündelblocktriangulationen benützt werden, sind Fragen nach der Optimalität der Lösung, der Fehlersensitivität, etc., nur von untergeordneter Bedeutung. Dies gilt allerdings schon nicht mehr für die Gewinnung von Näherungswerten zur Rekonstruktion bei langen Bildsequenzen. Wenig kritisch sind die Bewegungsparameter, die sich bei Bildsequenzen aufgrund der hohen Abtastrate zwischen zwei Bezugszeitpunkten in der Praxis nicht beliebig ändern können. Die Schwierigkeit liegt vor allem in den Strukturparametern, d. h. den Punktkoordinaten von Verknüpfungspunkten, die bei kleiner Bewegungsbasis eine relativ große Distanz aufweisen können, weshalb die linearen Verfahren zur relativen Orientierung oft sehr unzureichende Näherungswerte liefern.

Zur Vertiefung der Rekonstruktionsansätze aus Merkmalskorrespondenzen sei auf die Übersichtsartikel von Huang und Netravali (1990) und Aggarwal (1986) verwiesen. In deren Arbeiten werden verschiedene Verfahren aufgezeigt und dabei gelöste und ungelöste Fragen bezüglich der Existenz bzw. der Singularitäten, der Eindeutigkeit und der Sensitivität der Lösung bezüglich Fehlern besprochen. Diese Fragen werden umso wichtiger, je weniger Vorwissen z. B. über die Aufnahmekonfiguration vorliegt⁸.

2.4.3 Rekonstruktionsansatz für die lange Bildsequenz

Die Formulierung des Rekonstruktionsansatzes für die lange Sequenz soll auf markante Punkte konzentriert werden. Die Position des Punktes und der zugehörige Verschiebungsvektor sind dabei gleichbedeutend mit der Lage der korrespondierenden Punkte im entsprechenden Bildpaar. Für die Verallgemeinerung von der Stereoeinheit auf eine längere Bildsequenz ist es vorteilhaft, die korrespondierenden Punkte in allen Bildern, in denen sie auftreten, durch ihre entsprechende Position zu beschreiben und die Verknüpfung durch eine Identifikationsnumerierung anzugeben. Diese Standardvorgehensweise, die in der Photogrammetrie mit der Entwicklung der Bildtriangulationsverfahren schon eine lange Tradition hat, schlägt sich auch direkt in den Parametern zur Beschreibung der Bewegung nieder. An die Stelle der rotatorischen und translatorischen Geschwindigkeiten, bzw. deren diskrete Form, treten die Rotations- und Translationsparameter des sich bewegenden Systems, die photogrammetrisch als äußere Orientierung der Kamera bezeichnet werden.

Für die Rekonstruktion sind damit sowohl die beobachteten Daten, als auch die zu schätzenden Bewegungsparameter dem jeweiligen diskreten Bezugszeitpunkt zugeordnet. Die Strukturparameter im ortsfesten System sind

⁸So wurde kürzlich von Faugeras und Maybank (1990) nachgewiesen, daß für die relative Orientierung mit 5 Punktkorrespondenzen bis zu 10 reelle Lösungen existieren.

wegen der Stationarität der Szene keine Funktion der Zeit. Sind die Strukturparameter bekannt (zumindest in Form stochastischer Vorinformation), können im Prinzip die 3D-2D Korrespondenzansätze zur Rekonstruktion der Bewegung herangezogen werden. Aus den vorliegenden 2D-2D Punktkorrespondenzen läßt sich wiederum Information über neue Strukturen gewinnen. Der rekursive 3D-Filter, der im Kapitel 7 entwickelt wird, vereint diese beiden Aspekte der Rekonstruktion, wobei die Integrationswirkung der langen Sequenz ganz wesentlich über die Rekonstruktion der Struktur stattfindet.

Die diskreten und kontinuierlichen Ansätze werden in der Computer Vision Literatur streng getrennt behandelt (siehe z. B. Daniilides und Nagel, 1990). So benützen Heeger und Jepson (1992) zur Klassifizierung der Rekonstruktionsansätze hauptsächlich das Unterscheidungskriterium "kontinuierlich versus diskret" in der Zeit. Das Verschiebungsvektorfeld, aber auch die Rotation und Translation zwischen zwei Bezugszeitpunkten, kann als Approximation entsprechender Geschwindigkeiten nur unter der Voraussetzung hinreichend kurzer Zeitintervalle betrachtet werden. Da die Bilder einer Bildfolge aber schon diskret in der Zeit erfaßt werden, gilt umgekehrt, daß alle kontinuierlichen (differenziellen) Verfahren spätestens bei der algorithmischen Lösung mit diskreter Approximation arbeiten müssen. Für viele der kontinuierlichen Ansätze hat das zur Folge, daß die algorithmischen Lösungen nicht mehr korrekt arbeiten, wenn die Bewegung zwischen zwei Zeitpunkten zu groß wird (Heeger und Jepson, 1992).

Übergang von der kontinuierlichen zur diskreten Argumentation

Der oben angesprochene Übergang von der kontinuierlichen zur diskreten Argumentation ist deshalb ein sehr wichtiger Schritt bei der Formulierung des nachfolgenden Rekonstruktionsansatzes. Nicht die Geschwindigkeiten sollen ermittelt werden, auch nicht deren diskrete Approximation zwischen den Bezugszeitpunkten, sondern die Bewegungsparameter bezüglich eines festen übergeordneten Koordinatensystems zu den Bezugszeitpunkten selbst. Dies berührt die Frage der Datumsgebung für die 3D-Bewegungsparameter, d. h. die Frage nach der Festlegung von Startwerten, die bereits im vorhergehenden Abschnitt diskutiert wurde. Ferner treten anstelle der Geschwindigkeiten bzw. Verschiebungsvektoren die Positionen der Primitive im jeweiligen Bild. Um die kontinuierlichen, differenziellen Darstellungen der vorhergehenden Kapitel auf diskrete, feste Bezugszeitpunkte zu übertragen, muß in erster Linie die Betrachtungsweise, nicht so sehr der Formalismus geändert werden. Wird für einen Moment den Parameter Zeit gänzlich außer Acht gelassen, so findet die Differentiation (Operator d_t) nicht nach der Zeit, sondern nach den zu schätzenden Parametern statt. Es folgt daraus eine Linearisierung der nichtlinearen Beziehungen an einer im Prinzip frei wählbaren, geeigneten Näherungsstelle. Dies hat unmittelbar zur Konsequenz, daß die ermittelten Parameter im allgemeinen nicht mehr als Approximation der rotatorischen und translatorischen Geschwindigkeiten des sich bewegenden Sensors zu interpretieren sind, sondern lediglich Korrekturen der Näherungswerte darstellen.

Rekonstruktionsansatz

Die Modelle für die Rekonstruktion setzen sich aus den perspektiven Abbildungsgleichungen (2-32) und der Beziehung für die Starrkörperbewegungen (2-1) zusammen. Beobachtete Größen sind die Bildkoordinaten der Punkte in den einzelnen Bildern. Die zu ermittelnden Unbekannten sind die Parameter der Rotation R , die Position des Sensors \mathbf{U}_0 und die Koordinaten \mathbf{U}_i der 3D-Punktmenge, die in der stationären Szene beobachtet wurden. Da \mathbf{U}_i keine Funktion der Zeit ist, wurde bei der Beschreibung der differentiellen Bewegung die Restriktion $d_t \mathbf{U}_i = 0$ berücksichtigt. Für die Rekonstruktion der Struktur gilt es nun aber die Positionen \mathbf{U}_i entsprechend mitzubestimmen, wobei für die Position der Punkte ebenso wie für die Bewegungsparameter Näherungswerte vorliegen.

Der Rekonstruktionsansatz setzt sich dann aus der Perspektive gemäß (2-34)

$$d\mathbf{x} = L d\mathbf{X}$$

und der Starrkörperbewegung entsprechend (2-11,2-37)

$$d\mathbf{X} = V \boldsymbol{\omega} + R^T (d\mathbf{U} - d\mathbf{U}_0)$$

zusammen. Die translatorische Bewegung $d_t \mathbf{T}$ (2-11), die sich auf das bewegte Koordinatensystem bezieht, wurde hierbei ersetzt durch die Koordinierung des Projektionszentrums im übergeordneten System. Zusammengefaßt ergibt sich der Rekonstruktionsansatz

$$d\mathbf{x}_i = L V \omega - L R^T d\mathbf{U}_0 + L R^T d\mathbf{U}_i \quad (2-50)$$

der die beiden beobachteten Bildkoordinaten \mathbf{x}_i eines Punktes in Beziehung zu den 6 Parametern (ω bzw. $q(\omega)$, \mathbf{U}_0) der äußeren Orientierung und zu den 3 Koordinaten des Punktes (\mathbf{U}_i) im übergeordneten Referenzsystem setzt. Die Hilfsfunktion, die ω bei der Parameterschätzung übernimmt, wurde im Abschnitt 2.2 diskutiert. Der Übergang von ω zu den Rotationsparametern q_i der Quaternionen-Parametrisierung ist mit der linearen Transformation (2-29) angegeben. Mit den Näherungen für die Rotation R , für die Translation \mathbf{U}_0 und für die Position der Objektpunkte \mathbf{U}_i können die Designmatrizen L (2-35), LV (2-38) an der entsprechenden Linearisierungsstelle in einfacher Weise aufgestellt werden. Spezielle Aspekte des Rekonstruktionsverfahrens, wie z. B. das Einbringen von Paßpunktinformation, sind im Prinzip auch unter dem Begriff der 3D-2D Korrespondenz angesprochen. Auf sie wird beim Entwurf eines dreidimensionalen Kalman-Filters zurückgekommen.

Im Gegensatz zur kurzen Sequenz, die aus 2 oder 3 Bildern besteht, liegt bei der langen Sequenz das Hauptaugenmerk auf dem Integrationsaspekt über den Parameter Zeit. Zur Gewinnung von Näherungswerten können im Prinzip die linearen Verfahren aus der Stereoeinheit vorteilhaft verwendet werden. Alternativ dazu bietet die lange Sequenz die Möglichkeit zur Prädiktion der Zustandsgrößen, in denen die Bewegungsparameter und ein Teil der Strukturparameter enthalten sind. Das Problem der Lösungsmehrdeutigkeiten der Stereoalgorithmen bei geringer Bildanzahl ist bei der langen Sequenz nicht vorhanden. Aber die Problematik aller nichtlinearen Kleinste-Quadrate-Verfahren, daß das Erreichen der optimalen Lösung oft entscheidend von guten Startwerten abhängig ist, bleibt existent. Eine Forderung an den Prädiktionsschritt besteht deshalb darin, daß hinreichend gute Näherungswerte durch die Prädiktion gefunden werden.

Querbezüge zu andern Arbeiten mit langen Sequenzen

Der Rekonstruktion aus der langen Sequenz wird, vergleicht man sie mit der Anzahl der Arbeiten über die Zwei- oder Dreibildeinheit, bislang sehr wenig Aufmerksamkeit geschenkt. Die unter experimentellen Gesichtspunkten herausragende Entwicklung der letzten 10 Jahre ist die eines autonom navigierenden Fahrzeugs, mit dem auf einer für die Öffentlichkeit noch nicht freigegebenen Autobahn Geschwindigkeiten von 100 km/h erreicht und dabei eine Strecke von mehr als 20 km zurückgelegt wurde (Dickmanns und Graefe, 1988, Dickmanns und Mysliwetz, 1992). Wie für diese Arbeit, so gilt auch für viele andere anwendungsspezifisch formulierte Ansätze, daß von den Bewegungs- und Strukturparametern in (2-50) nur ein Teil rekonstruiert wird. Kuan, Phipps und Hsueh (1987) verzichten bei der Modellierung zur autonomen Fahrzeugführung auf Straßen gänzlich auf eine Bewegungsbestimmung, Wünsche (1988), Lin und Zhu (1990) beschränken die Bewegungsparameter auf zwei translatorische und einen rotatorischen Freiheitsgrad (Bewegung in einer Ebene). Sawhney et al. (1990) lassen nur Bewegungen auf einer Kreisbahn zu. In weiteren Arbeiten, vor allem im Robotikbereich, werden die Bewegungsfreiheitsgrade ausschließlich auf Rotationen oder auf Translationen beschränkt oder es wird der Bewegung die Restriktion auferlegt, daß rotatorische und translatorische Geschwindigkeiten für die gesamte Zeitdauer der Bewegung konstant sind. Mit unabhängig erfaßten Bewegungsparametern (Odometer, Inertialmeßsystem, GPS, Roboter) haben Novak (1991), Thorpe et al. (1992) und Roberts und Bahnu (1990) die Schätzung der Bewegungsparameter zusätzlich unterstützt bzw. auch ersetzt.

Den experimentellen Verfahren stehen methodische Entwicklungen gegenüber, welche die Ermittlung aller 6 Bewegungsfreiheitsgrade anstreben. Weitere signifikante Unterscheidungsmerkmale für diese Arbeiten liegen darin, ob

1. Objektmodelle explizit gegeben sind oder über die Starrheit hinaus keine weiteren Informationen über das Objekt vorliegen;
2. die Schätzung rekursiv (sequentiell, inkrementell) oder simultan (batch-Lösung) erfolgt;
3. es sich um Mono- oder Stereosequenzen handelt;

4. eine dynamische oder eine kinematische Modellierung der Bewegung vorgesehen ist;
5. die 3D-Struktur durch Punkte oder Linien beschrieben wird;
6. neben der funktionalen Beziehungen auch eine stochastische Modellbildung ausgearbeitet wird;
7. sich die Rekonstruktion auf ein lokal mitbewegtes Koordinatensystem oder ein übergeordnetes Referenzsystem bezieht;
8. ein oder mehrere bewegte Objekte verfolgt werden.

Gegebene Objektmodelle schränken die Dimension bei der Bestimmung von Strukturparametern über das Objekt beträchtlich ein und die monobildtypische Aufgabenstellung der Objektlokalisierung wird bei Bildsequenzen üblicherweise als Objektverfolgung bezeichnet. Eine Filterlösung, in der neben der Position und Richtung auch die translatorischen und rotatorischen Geschwindigkeiten polyedrischer Objekte rekursiv über die Sequenz hinweg bestimmt werden, wurde von Gennery (1992) vorgestellt. Mit dem Modell einer Objektebene arbeiten Barron et al. (1990), die vor allem betonen, daß für die Ermittlung der rotatorischen und translatorischen Geschwindigkeiten und des Normalenvektors der Ebene die zu verschiedenen Zeitpunkten gemessenen Bildpunktgeschwindigkeiten sich nicht auf denselben 3D-Punkt beziehen, sondern lediglich von derselben Ebene herrühren müssen. Korrespondenz wird hierbei nicht für die einzelnen Messungen in der Ebene zu verschiedenen Zeitpunkten gefordert, sondern lediglich in abgeschwächter Form durch die Zugehörigkeit zur selben Ebene vorausgesetzt. Dies impliziert aber die Gruppierung der zur selben Ebene gehörenden Messungen.

Burger und Bhanu (1990) konzentrieren sich auf die Bestimmung der Rotation und des "fuzzy" Focus of Expansion, der einer stochastischen Beschreibung der Bewegungsrichtung gleichkommt. Sie verzichten auf die Berechnung von Strukturinformation und nützen die berechneten Bewegungsparameter eines Bildpaares lediglich als Näherungsinformation zur Beschränkung des Suchraumes für die nachfolgende Stereoeinheit. Wie Burger und Bhanu vermeiden auch Spetsakis und Aloimonos (1991) die Berechnung von Struktur. Sie erweitern die linearen Ansätze aus der Stereoeinheit (2-49), in dem sie eine simultane Lösung über N Bilder vorsehen.

Matthies et al. (1988, 1989) greifen die Arbeiten von Moravec (1980) auf, der den Einsatz von Stereobildsequenzen zur autonomen Navigation in einer unbekanntem Umgebung als einer der ersten untersuchte. Unter der Voraussetzung, daß die Basis kalibriert vorliegt, wird dadurch in einfacher Weise die Maßstabsunbestimmtheit beseitigt. Grundsätzlich ist die Bestimmung der Bewegungs- und vor allem der Strukturparameter aufgrund der meist besseren Schnittgeometrie zuverlässiger und auch genauer; ein Vorteil, der allerdings durch den zusätzlichen Aufwand der Korrespondenzbildung in der Stereoeinheit erkauft wird. Das Hauptinteresse gilt der Struktur und deren stochastischen Beschreibung, die Matthies in Form von Tiefenkarten über Filtergleichungen rekursiv ableitet und fortführt.

Eine über die Starrkörperbewegung hinausgehende Modellierung der Trajektorie haben Broida und Chellappa (1986, 1991) und Weng et al. (1987) vorgeschlagen. Die Heuristik an die Bewegungsform ist, daß diese durch eine "glatte" Bewegung beschrieben werden kann. Broida und Chellappa approximieren die translatorische Bewegung durch ein Vektorpolynom N -ten Grades in der Zeit, d. h. durch Position, Geschwindigkeit, Beschleunigung, etc.. Das Modell für die Rotation nützt die Differentialgleichungen (2-28) für die Quaternion-Parametrisierung. Die Winkelgeschwindigkeiten ω werden dann ebenfalls durch ein Polynom in der Zeit approximiert. Naheliegender für die Bestimmung der Polynom-Parameter ist die Berechnung in einer Simultanlösung, obwohl Broida und Chellappa (1986) auch eine rekursive Prozedur vorgeschlagen haben. Aus der Sicht der Dynamik argumentieren Weng et al. (1987), die als Rotationsmodell eine Präzession (d. h. der Körper rotiert um eine körperfeste Achse, die wiederum um eine raumfeste Achse rotiert) vorsehen. Glattheit für die Translationsbewegung des Rotationszentrums wird durch ein Vektorpolynom in der Zeit eingebracht. Eine weitere Besonderheit in Weng's Arbeit ist, daß zunächst Bewegungsparameter aus den Punktkorrespondenzen zweier Aufnahmezeitpunkte ermittelt und dann die rotatorischen und translatorischen Parameter der Trajektorie abgeleitet werden. Dabei spielt es nur eine untergeordnete Rolle, ob dabei Mono- oder Stereosequenzen ausgewertet werden.

Zhang und Faugeras (1992) arbeiten ebenso wie Matthies mit Stereobildsequenzen, verwenden als Primitive aber Linien bzw. Liniensegmente. Ähnlich wie bei Weng erfolgt ein frühzeitiger Übergang in den Objektraum. Die Linien werden als 3D-Primitive zwischen den Bezugszeitpunkten verfolgt und daraus deren Position und

Bewegung ermittelt. Als Bewegungsparameter für die einzelnen 3D-Primitive werden neben der Winkelgeschwindigkeit auch Translationsgeschwindigkeit und -beschleunigung (und im Prinzip auch höhere Ableitungen) in einen Kalman-Filter als Zustandsvektor eingeführt. In einem zweiten Schritt können diese Primitive aufgrund ihrer Bewegung zu Objekten gruppiert werden, wodurch eine Objektsegmentierung erreicht wird. Ein ähnliches Konzept wie Zhang und Faugeras haben Crowley und Stelmazk (1990) vorgestellt.

In den Arbeiten von v. Brandt und Zaig (1991) und von Cui et al. (1990) wird versucht, die nichtlinearen Beziehungen (2-50) bezüglich Bewegungs- und Strukturbestimmung möglichst zu entkoppeln. Bei v. Brandt und Zaig wird die Tiefenkomponente durch Kalman-Filterung aufdatiert und die Schätzung der Eigenbewegung unter Nutzung dieser (stochastisch) bekannten 3D-Punkte bestimmt. Durch die bekannte Tiefe vereinfacht sich entsprechend (2-39) der Schätzprozeß zur Bewegungsschätzung zu einem linearen Problem. Bewegung und Struktur werden jeweils im Wechsel über die Sequenz hinweg ermittelt bzw. fortgeführt. Durch die Vernachlässigung der Korrelation unter den Strukturparametern handelt es sich bei jedem 3D-Punkt um ein separates Kalman-Filter mit einer skalaren Zustandsgröße. Cui et al. (1990) sehen ebenfalls vor, durch einen linearen Algorithmus Bewegungsparameter zu ermitteln, setzen diese allerdings nur als Näherung für die nichtlineare Parameterschätzung an. Die Vernachlässigung der Korrelation zwischen Struktur- und Bewegungsparameter wird zur Formulierung einer zweistufigen Prozedur vorausgesetzt, in der wechselweise Bewegung und Struktur geschätzt werden, bis sich eine stabile Lösung einstellt. Da alle Strukturparameter zu einem Bezugszeitpunkt simultan geschätzt werden, zwischen den Bezugszeitpunkten für Bewegung und Struktur eine Kalman-Filterlösung vorgesehen ist, bezeichnen Cui et al. ihren Ansatz als "rekursiven-batch" Typ.

Zusammenfassend soll noch einmal hervorgehoben werden, daß die meisten der diskutierten Arbeiten mit langen Bildsequenzen sich vom Rekonstruktionsansatz (2-50) zunächst dadurch unterscheiden, daß sie nicht in einem übergeordneten stationären Koordinatensystem, sondern in dem mit der Kamera mitbewegten System rekonstruieren. Dies impliziert zu jedem Bezugszeitpunkt einen Datumwechsel für die rekonstruierten Strukturparameter und vor allem deren Genauigkeitsbeschreibung durch Kovarianzmatrizen, was in der Regel kaum beachtet wird. Weitere Vorschläge, wie sie insbesondere in Arbeiten unterbreitet werden, die sich auf die Bewegungstereoeinheit konzentrieren, haben die rekursive Anwendung der Stereorekonstruktion zur Grundlage. Der Schlüssel, auch bei der rekursiven Rekonstruktion aus der Monosequenz, liegt nun aber darin, daß die Information aus den zurückliegenden Schätzungen zum aktuellen Rekonstruktionszeitpunkt voll zur Verfügung steht, was in eleganter Weise durch 3D-Kalman-Filterung erfolgen kann.

Die meisten Verfahren legen die Bewegung in einer unbekanntem Umgebung der Bildsequenzanalyse als Maxime zugrunde, weshalb Vorwissen über das Objekt in Form spezifischer Objektmodelle nur geringe Beachtung findet. Dies gilt eingeschränkt auch für die sehr anwendungsorientierten Arbeiten. Zum Beispiel verwenden Dickmanns und Mitarbeiter ein generisches Straßenmodell, um damit wesentlich flexibler zu bleiben als dies mit spezifischen (statischen) Objektmodellen, z. B. Karten, möglich ist. Das Thema Bildzuordnung bzw. Korrespondenz, das wir in den nächsten Kapiteln besprechen, wird bei Arbeiten mit langen Sequenzen vielfach ausgespart und durch interaktive Messung oder Simulation ersetzt.

Teil II

Bildzuordnung

Die Zuordnung von Bildern stellt einen Meßprozeß dar, in dem die Messung auf den Intensitätsstrukturen eines Bildes oder auf einer daraus abgeleiteten symbolischen Repräsentation des Bildinhaltes beruht. Mit der Erfassung einer dynamischen oder auch stationären Umgebung durch eine Kamera, die sich im Falle eines navigierenden Fahrzeugs in der dreidimensionalen Welt bewegt, entsteht die für Bildsequenzen typische Meßaufgabe: *die Bestimmung der in den Bildern erkennbaren Bewegungen*¹. *Bildzuordnung* steht dann für die Aufgabe, mithilfe eines Modells diese auf den Bildraum bezogenen Bewegungen bestmöglich zu ermitteln. Für die in der Bildebene und in der Zeit diskret vorliegenden Bilddaten kann die Messung der Bewegung durch die Verschiebung der Bildpunkte zwischen zwei Bezugszeitpunkten repräsentiert werden. Bei hinreichend hoher zeitlicher Abtastung approximiert der Verschiebungsvektor eines Bildpunktes die Geschwindigkeit der Intensitätsstruktur in diesem Punkt. Dies äußert sich in Definitionen des Begriffes *optischer Fluß*

- *Optical flow is the apparent motion of the brightness pattern* (Horn, 1986, S. 280)
- *Optical flow is the instantaneous velocity of the brightness pattern* (Horn, 1986, S. 401)
- *Optical flow is the distribution of the apparent velocities of the movement of brightness patterns in an image* (Horn und Schunk, 1981, S. 185),

von denen noch viele weitere mit ähnlicher Zielrichtung existieren². Bildpunktbezogen gibt der optische Fluß den Geschwindigkeitsvektor einer Bildstruktur zu einem Zeitpunkt t an, der in einer endlichen Zeiteinheit einem zweidimensionalen Bewegungsvektor oder Parallaxenvektor entspricht. Die diskrete räumliche und zeitliche Erfassung der Bilder, für die in der Praxis hohe zeitliche (der räumlichen Auflösung adäquate) Abtastraten meist nicht gegeben sind, zwingt in jedem Fall dem Meßprozeß eine diskrete Struktur auf. D. h. bestimmt werden bild-elementbezogen Verschiebungsvektoren oder Verschiebungsraten (Nagel, 1989). Bei der diskreten Berechnung wird dennoch häufig von Geschwindigkeiten gesprochen. Auf die gesamte Bildebene bezogen ist der optische Fluß eine vektorwertige Funktion, die die momentane Geschwindigkeit in der Bildebene angibt. In der entsprechenden diskreten Betrachtung tritt an deren Stelle ein Verschiebungsvektorfeld. Dieses Vektorfeld kann flächenhaft dicht vorliegen, d. h. mit Verschiebungsvektoren in allen Bildpunkten, oder auch auf Konturen beschränkt bleiben. Im Hinblick auf eine symbolische Bildrepräsentation durch punktbezogene Primitive wird der Begriff des optischen Flusses auch zur Beschreibung der Verschiebungsvektoren, z. B. prominenter Grauwertstrukturen verwendet (Nagel, 1987).

Die Messung der Bewegung im Bild muß sich daran orientieren, in welcher Art und Weise das Resultat des Meßprozesses weiter verwendet werden soll. Beispiele sind dichte Verschiebungsvektorfelder, die zu jedem Bildpunkt des Bildes einen Verschiebungsvektor angeben und deshalb im Prinzip eine sehr detaillierte Analyse der Geometrie oder Bewegung von Objekten in der realen Welt gestatten. Verschiebungsvektorfelder, die ebenfalls bildpunktbezogen dicht, aber auf den Konturverlauf³ beschränkt sind, legen nahe, daß entsprechende Konturlinien in 3D als markanteste Information der Objekte analysiert werden. Eine weitere beträchtliche Informationsreduktion bezüglich des Objektraumes liegt vor, wenn die Bewegungsmessung auf prominente Grauwertstrukturen oder Bildprimitive beschränkt bleibt. Im Gegensatz zu den flächenhaften oder konturorientierten Meßverfahren wird der einzelne Bildpunkt nicht mehr als Meßposition betrachtet. Stattdessen beinhaltet der Transfer, insbesondere zu höherwertigen Primitiven, eine Analyse der Bildstruktur und damit der Bildpunkte

¹In vielen Ansätzen zur Bewegungsschätzung wird räumliche und zeitliche Kontinuität im Sinne des in Abb. 1.1 dargestellten Bildblocks zugrundegelegt und dabei von der Bewegung im Bild oder in der Bildebene gesprochen. Zeitdiskret ist selbstverständlich Bewegung in einer Momentaufnahme per Definition nicht sichtbar oder bestimmbar.

²Wir haben hier ausschließlich Definitionen des optischen Flusses von Horn zitiert, obwohl noch viele weitere Autoren ähnliche Definitionen vorgenommen haben. Diese Definitionen des Begriffes haben in der Literatur eine breite Zustimmung gefunden haben. Geprägt wurde der Begriff von Gibson (Gibson, 1950, Gibson et al., 1959). Manche Autoren, wie z. B. Kanatani (1990), identifizieren mit dem optischem Fluß aber auch die rein geometrisch definierten projizierten Geschwindigkeiten (siehe Abschnitt 2.3), die wir mit Bewegungsfluß bezeichnet haben.

³Konturlinien im Bild sind z. B. durch die Nulldurchgänge eines LoG-gefilterten Bildes feststellbar (vgl. Abschnitt 3.2.4).

in ihrer (lokalen) räumlichen Anordnung. Für die Stereomessung auf dieser symbolischen Bilddatenebene ist insbesondere charakteristisch, daß die Position oder andere Parameter eines Punktes, einer Linie oder Fläche in jedem Bild separat erfaßt werden und die Bewegung sich dann in der Differenz dieser Parameter äußert.

Das Thema Bildzuordnung wollen wir wiederum an der Gesamtzielsetzung dieser Arbeit ausrichten, wobei wir die Maxime einer unbekannt, stationären Umgebung zugrunde legen (vgl. Abschnitt 1.8). Für die Zuordnungsproblematik innerhalb der Monobildsequenz steht als kleinste Einheit die Stereokonfiguration (Bewegungsstereo). Die Bewegung im Bild zwischen zwei Aufnahmezeitpunkten haben wir nach der Abstraktionsebene der Bilddaten in intensitätsbasierte Verfahren und in merkmalsbasierte Verfahren unterteilt (vgl. Abschnitt 1.5). Kennzeichnend für die intensitätsbasierten Ansätze ist vorwiegend ein auf das ganze Bild oder auf Regionen beschränktes, dichtes Bewegungsvektorfeld, wohingegen die merkmalsbasierten Verfahren meist abhängig vom Abstraktionsgrad der Primitive eine eher geringere Anzahl gemessener Daten der Bewegung im Bild liefern.

Eine andere Möglichkeit ist die Kategorisierung der Meßprozesse nach der mathematischen Formulierung der Optimierungsaufgabe. Die Gruppe der gradientenbasierten oder differentiellen Verfahren beruht auf der Differentiation der Bildfunktion $I(x, y, t)$ nach Ort und Zeit, wobei die räumlichen und zeitlichen Gradienten zueinander in Verbindung gesetzt werden. Zur Regularisierung des Problems wird zusätzlich Glattheit der Bewegung in der Bildebene gefordert und die Optimierungsaufgabe zur Bestimmung des optischen Flusses mit Techniken der Variationsrechnung gelöst. Die Besonderheit in diesen Lösungswegen ist, daß erst beim Entwurf der algorithmischen Lösung auf eine Diskretisierung übergegangen wird. Eine zweite Gruppe bilden die Kleinste-Quadrate-Verfahren, die schon im Ansatz die Bildintensitäten als diskrete Meßdaten einbeziehen. Sie werden in der Regel fensterorientiert modelliert und fordern, daß die Bildbewegung innerhalb des Fensters durch eine einfache geometrische Transformation beschreibbar ist. Diese geometrischen Transformationsmodelle stehen in enger Beziehung zur Formbeschreibung von Objekten (siehe Abschnitte 2.3.4 und 2.3.5).

Zu dieser zweiten Gruppe können auch die Korrelationsverfahren gezählt werden, die fensterorientiert fordern, daß die Bildbewegung innerhalb des Korrelationsfensters durch eine lokale Translation beschrieben wird. Das klassische Korrelieren, in dem durch Maximierung des Korrelationskoeffizienten die Zuordnung hergestellt wird, stand am Anfang der Methodenentwicklung zur Bildzuordnung. In Kombination mit globalen Glattheitsrestriktionen hat Anandan (1987a) daraus ein Verfahren zur dichten Bewegungsschätzung vorgeschlagen.

Die dritte Gruppe bilden die diskreten Verfahren, die durch die Extraktion von Bildprimitiven bei der Formulierung des Zuordnungsansatzes diskrete Attribute, z. B. Punktpositions- oder Geradenparameter zur Bewegungsschätzung heranziehen. Die Zuordnungsverfahren auf der symbolischen Ebene der Bildbeschreibung setzen einen zweistufigen Meßprozeß voraus, in dem zuerst aus der Analyse der Bildstruktur signifikante Information bezüglich der Position oder anderer Ortsparameter im Bild erfaßt und dann über Such- oder Relaxationsverfahren korrespondierende Primitive einander zugeordnet werden, was in geometrischer Hinsicht letztlich der Bewegungsmessung dieser Primitive gleichkommt.

Die klassische Betrachtungsweise, daß die intensitätsbasierten Verfahren mehr mit dichteren Bildfolgen und die merkmalsbasierten Verfahren mehr mit den losen Bildfolgen in Verbindung gebracht werden, reflektiert in direkter Weise die differentielle bzw. diskrete Problemformulierung bezüglich Zuordnung und Rekonstruktion in der dreidimensionalen Welt. Für den Meßprozeß haben wir bereits ausgeführt, daß die differentiellen Ansätze in diskretisierte, algorithmische Lösungen überführt werden müssen. Die Erweiterung der Ansätze zu Mehrebenenprozeduren durch die Nutzung von Bildpyramiden ermöglicht die Gewinnung von Näherungsinformationen. Die differentiellen Verfahren sind so umzuformulieren, daß Näherungsinformationen über die Bewegung im Bild berücksichtigt werden können. Mit einer hierarchischen Vorgehensweisen lassen sich dann wesentlich größere Konvergenzradien erzielen, so daß entsprechende Verfahren auch bei losen Bildsequenzen einsetzbar sind. Eine weitere Möglichkeit zur Gewinnung von Näherungswerten ergibt sich aus der rekursiven Parameterschätzung und findet vor allem bei der Verfolgung von Bildprimitiven Anwendung. Die bereits zu früheren Zeitpunkten zugeordneten Primitive können über die Bewegungs- und Strukturrekonstruktion z. B. innerhalb eines Kalman-Filters in das aktuelle Bild prädiert werden und sind dadurch bezüglich ihres Ortes näherungsweise bekannt, wobei die mitprädierte Unsicherheit auch einen Hinweis auf den notwendigen Suchbereich gibt. Dennoch werden auch hier größere Konvergenzradien intensitätsbasiert und merkmalsbasiert hauptsächlich durch die Entwicklung hierarchischer Prozeduren gewonnen.

Eine Sonderrolle nehmen eine Klasse von intensitätsbasierten Verfahren ein, die erst in jüngerer Zeit vorgeschlagen wurden. Setzt man eine der in der Bildebene vorliegenden räumlichen Auflösung adäquat dichte zeitliche Diskretisierung voraus, wie sie etwa in dem Raum-Zeit-Bildblock in Abbildung 1.1 vorliegt, lassen sich Raum-Zeit Filtertechniken (Korrelationstechniken) zur Bestimmung der Bewegung im Bildraum einsetzen. Lineare Filterung (Adelson und Bergen, 1985), die Schätzung lokaler Geschwindigkeiten mit sog. Gabor-Filtern durch Analyse des Raum-Zeit Powerspektrums (Heeger, 1986, 1988), die Kantendetektion im Raum-Zeit Bildblock (Buxton und Buxton, 1984) und die konzeptionelle Erweiterung der Epipolarlinien zu Epipolarebenen (Bolles et al., 1987) sind eine Reihe von Ansätzen, die insbesondere auf die Kontinuität⁴ im Bild-Zeit Raum aufzubauen. Diese Operationen beziehen mehrere, z. T. sehr viele Bilder mit ein, d. h. sie sind ein erster Schritt zu einer über den Parameter Zeit integrierten Analyse der Bildfolgen. Ein großer Vorteil der hohen zeitlichen Abtastrate ist, daß ein aufwendiges Verfolgen von Bildtextur bzw. Primitiven praktisch entfallen kann. Der entscheidende Nachteil liegt aber in der aus der hohen zeitlichen Abtastrate resultierenden geringen Bewegungsbasis, die eine Schätzung von Bewegung und Struktur im Objektraum erst mit einer sehr großen Anzahl von Bildern gestattet. Von den oben genannten Arbeiten hat auch nur Bolles die Rekonstruktion von 3D-Struktur angesprochen. Wir wollen im folgenden diese Ansätze nicht weiter vertiefen und den Schwerpunkt auf die Zuordnungsproblematik legen, die insbesondere auch lose Bildfolgen einbezieht.

Übersicht über Teil II

Im nächsten Kapitel führen wir in die Schätzung des optischen Flusses ein, wobei wir am Beispiel des sehr populären Aufsatzes von Horn und Schunk (1981) detaillierter auch den algorithmischen Lösungsweg aufzeigen werden. Die meisten Verfahren zur Bestimmung des optischen Flusses beruhen auf der Minimierung sogenannter quadratischer Funktionale. Das mathematische Rüstzeug zur Lösung solcher Optimierungsaufgaben stellt die Variationsrechnung dar, die darüber hinaus zur Formulierung weiterer Rekonstruktionsaufgaben im Bereich Computer Vision, wie beispielsweise der Oberflächenrekonstruktion, verwendet wird.

Eine Alternative zu diesen kontinuierlichen Ansätzen bilden die Kleinste-Quadrate-Verfahren, bei denen die Meßdaten und die zu bestimmenden Größen (in Form von Parametern) diskret in die Formulierung des Ansatzes eingehen. Verschiedene Beobachtungstypen, wie z. B. auch konturorientierte Beobachtungen, und vor allem ein breites Modellierungsspektrum, wie beispielsweise "starre Szene", "schwach gekrümmte Oberfläche" oder "differentialgeometrische Modellierung der Bildfunktion" führen zu einer Vielzahl von Ansätzen zur Bewegungsschätzung in örtlich begrenzten Bildausschnitten. Analogien zwischen Variationslösungen und Kleinste-Quadrate-Zuordnungsverfahren, auch in Bezug auf die Arbeiten in der Photogrammetrie, werden diskutiert, um ein vertieftes Verständnis für diese für Bildfolgen typischen Arbeitsschwerpunkte im Zuordnungsbereich zu gewinnen.

Die Probleme bei der Ermittlung des optischen Flusses in Regionen mit Bewegungsdiskontinuitäten sind bislang noch nicht befriedigend gelöst, zumal aus theoretischer Sicht zuverlässige Schätzungen der Bildgeschwindigkeiten nur in Bildbereichen mit starken Intensitätsdiskontinuitäten erzielt (Verri und Poggio, 1989) werden können. Dies legt nahe Verfahren einzubeziehen, wenn nicht gar zu favorisieren, die auf der symbolischen Beschreibungsebene angesiedelt sind. Unter den möglichen Bildprimitiven wie Punkten, Linien und Flächen werden wir uns in dieser Arbeit auf Punkte beschränken (wobei darüberhinaus konturorientierte Verfahren zur Schätzung des optischen Flusses in den Variations- und Kleinste-Quadrate-Ansätzen diskutiert sind). Im 4. Kapitel wird durch Punktmerkmale eine Form der symbolischen Repräsentation des Bildinhaltes dargestellt und ihre Funktion bei der Zuordnung bzw. Verfolgung dieser Punkte über die Sequenz hinweg diskutiert. Eine weitere wichtige Dimension für die Bildzuordnung besteht in hierarchischen Zuordnungsstrategien, durch die Näherungswerte mittels Bildpyramiden gewonnen werden. Eine spezielle pyramidale Bildrepräsentation bildet der Scale-Space, auf den wir im 5. Kapitel vertieft eingehen. Mit der funktionalen Integration des Scale-Space Parameters in die Formulierung eines intensitätsbasierten Zuordnungsansatzes werden wir eine neue Theorie zur Lösung des Zuordnungsproblems entwickeln. An Beispielen wird die Überlegenheit bezüglich Konvergenzradius und Effizienz dieses Verfahrens gegenüber bestehenden Methoden deutlich. Auch ein Schwachpunkt merkmalsbasierter

⁴Die kontinuierliche funktionale Beschreibung im Bild-Zeit Raum als Folge der Bewegung im Objektraum ist Teil des Prinzips, das besagt, daß Kohärenz (Grundsatz vom Zusammenhang; die Natur produziert kohärente Strukturen viel eher als zufällige) in der Welt Kontinuität im Bild erzeugt (Fischler und Firschein, 1987, S. 765).

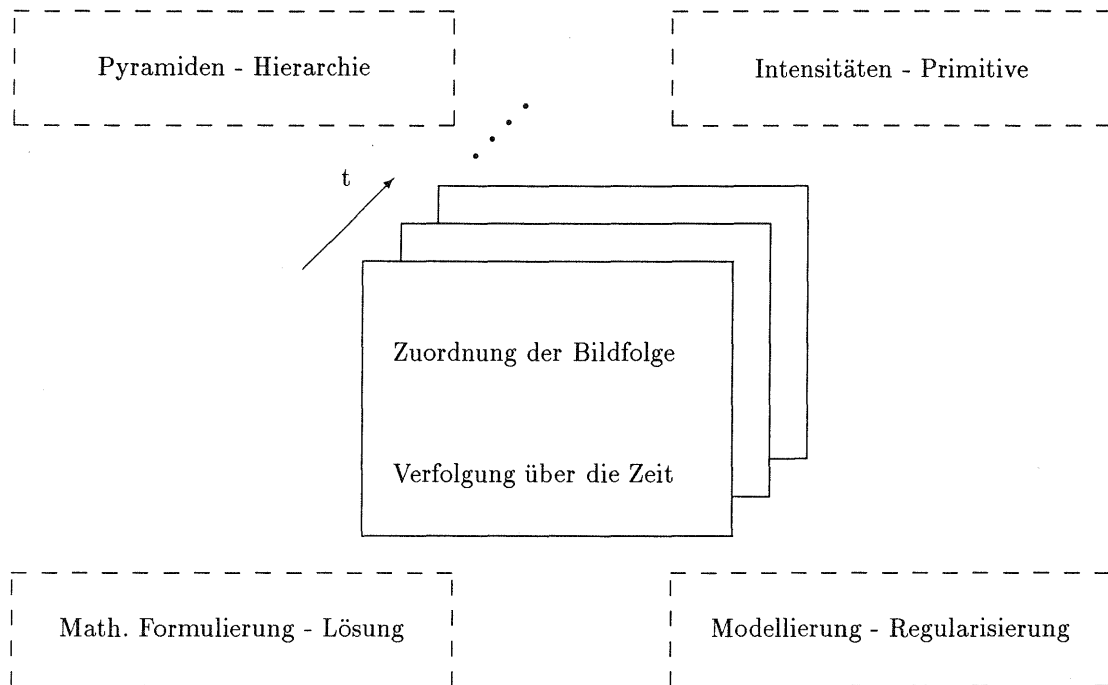


Abbildung 3.1: Aspekte der Zuordnungsproblematik

Verfahren, daß nur sehr wenige der extrahierten Primitive über einen längeren Zeitabschnitt in der Sequenz detektiert werden, kann durch die Verfolgung der Punkte mit diesem Zuordnungsansatz beträchtlich verbessert werden. Von Punkten, die lange in der Sequenz sind, profitiert vor allem die 3D-Rekonstruktion. Der dafür ausgearbeitete Kalman-Filteransatz hat eine Integrationswirkung für die Bewegungs- und Strukturparameter, so daß mit fortschreitender Zeit bzw. damit wachsender Zahl verarbeiteter Bilder dies zur Qualitätssteigerung bezüglich der Genauigkeits- und Zuverlässigkeitsparameter führt. Eine weitere unmittelbare Konsequenz ist, daß auch die Kontrolle des Meßprozesses über den Rekonstruktionsansatz verstärkt, d. h. die Lokalisierung von Zuordnungsfehlern verbessert wird. Die verschiedenen Aspekte und Komponenten der Bildzuordnung sind schematisch in Abbildung 3.1 graphisch wiedergegeben.

3 Schätzung des optischen Flusses

Wissenschaftlicher Beitrag der Arbeit in diesem Kapitel:

Für die Bestimmung des optischen Flusses ist eine große Vielfalt an Verfahren vorgeschlagen worden, wobei grob nach kontinuierlichen und diskreten Ansätzen unterteilt werden kann. Da diese meist zu sehr verschiedenen Lösungswegen (Variationsrechnung vs. Kleinste-Quadrate-Schätzung) führen, sind Vergleiche zwischen den Lösungswegen bislang kaum angestellt worden. Hier will die Arbeit einen Weg aufzeigen, wie solche für die theoretische Durchdringung und Vereinheitlichung der Verfahren wichtige Vergleiche angestellt werden können. Dies erfolgt an einem Beispiel in einer Gegenüberstellung und Diskussion der Unterschiede bzw. Gemeinsamkeiten bezüglich der Zielfunktion, der analytischen sowie der algorithmischen Lösung. Die Aufarbeitung von herausragenden Arbeiten erfolgt zusammen mit einer ausführlichen Diskussion der Querbezüge zwischen den Arbeiten und vor dem Hintergrund, die Eigenschaften der Verfahren bezüglich des Nutzens für die Zuordnung der Bildsequenz aufzuzeigen. Dabei wird deutlich, daß kein Verfahren als das "Beste" für eine solche Aufgabe identifiziert werden kann. Die Ursache liegt darin, daß schon bei wenig komplexen Szenen eine einzelne Prozedur die notwendige Interpretation z. B. bei Tiefensprüngen oder Verdeckungen nicht leisten kann. Diese Erkenntnisse, die meist in den Bereich paradigmatischer Arbeiten führen, helfen die Möglichkeiten und Grenzen der Bildzuordnung zu erkennen und besser einzuschätzen.

Ein grundlegendes Problem der Bildzuordnungsaufgaben⁵ ist, daß die primär beobachteten Daten - vorwiegend die Bildintensitäten - die Bestimmung des gewünschten Resultates - die Bewegung im Bild - nicht hinreichend beschränken. Diese sogenannten inversen⁶ Probleme werden als schlecht gestellte Probleme bezeichnet. Allgemein gilt als gut gestelltes Problem eine Aufgabenstellung, für die (a) eine Lösung existiert, (b) diese Lösung eindeutig ist und (c) diese Lösung kontinuierlich von den ursprünglichen Daten abhängt (vgl. z. B. Poggio et al., 1985). Die Bildzuordnungsaufgabe verletzt schon das erste dieser drei Kriterien, denn es gilt grob gesprochen für jedes Bildelement aus einer beobachteten Grauwertdifferenz einen zweidimensionalen Verschiebungsvektor zu bestimmen. Um zu einer Lösung zu gelangen, wird die Lösungsvielfalt durch Regularisierung in ein gut gestelltes Problem überführt. Die Regularisierung stellt eine Form von a priori Information dar, die den Zuordnungsprozeß dann zum Erfolg führt, wenn sie die physikalische Ursache hinreichend beschreibt. Verschiedene Formen von a priori Informationen haben wir im Abschnitt 1.4 unter den Heuristiken der Zuordnungsverfahren, insbesondere der Glattheitsannahme für die Bewegung im Bild, bereits in seiner grundsätzlichen Bedeutung diskutiert. In der Glattheitsrestriktion spiegelt sich wieder, daß, sieht man von Bewegungsdiskontinuitäten ab, benachbarte Bildpunkte eines sich bewegenden bzw. deformierenden Objekts ähnliche Geschwindigkeiten aufweisen.

Bevor wir auf die komplette Modellierung zur Schätzung des optischen Flusses eingehen, wollen wir uns im nächsten Abschnitt der Regularisierung und finite Elemente Modellierung zuwenden.

3.1 Regularisierung und finite Elemente Analyse

Da die beobachtungsspezifische Modellierung (Abschnitt 3.2) nur lokalen Charakter hat, kann, wenn es die lokale Bildstruktur nicht zuläßt (z. B. bei homogenen Graufächen), im ungünstigsten Fall aus der zeitlichen Variation keinerlei Information über die Bewegung in diesem Bereich gewonnen werden. Es werden zusätzlich Glattheitsrestriktionen benötigt, die einen globaleren Charakter haben, d. h. die Konsistenz eines Bewegungsvektors mit seiner Umgebung modellieren.

Die Regularisierung ist eine mathematische Standardtechnik zur Lösung schlecht gestellter Probleme. Die Aufgabenstellung besteht darin, durch schwache Glättungsrestriktionen (Tikhonov und Arsenin, 1977, vgl. Szeliski, 1990) die Probleme in gut gestellte Probleme zu überführen. "Schwach" steht dafür, daß im stochastischen Sinne keine fehlerfreie, sondern eine approximierende Restriktion einzuhalten ist (vgl. z. B. die Diskussion des schwachen geodätischen Datums in Schaffrin, 1984). Aus den beobachteten Daten soll eine regularisierte Lösung bestimmt werden. Für die Interpolation eines Höhenmodells ist die Lösung eine zweidimensionale Funktion (Höhe) über der Lagebezugsfläche $u(x, y)$, für die Stereoproblematik in Epipolargeometrie ist die Parallaxenfunktion $u(x, y)$ die zu bestimmende Größe, die ebenfalls eine skalare Funktion zweier Variablen (Bildkoordinaten) ist. Der optische Fluß ist eine vektorwertige Funktion $\mathbf{u}(x, y, t)$, die sich als Funktion der Bildkoordinaten und der Zeit darstellen läßt und die auf dem diskreten Bildraster eines Aufnahmezeitpunktes auch als Verschiebungsvektorfeld bezeichnet wird.

Die Regularisierung in fast allen Arbeiten im Bereich Computer Vision wird über die (partiellen) Ableitungen bis zum n-ten Grad einer Funktion formuliert. Wählen wir als Beispiel eine skalare Funktion $u(x, y)$ zweier Variablen x und y , so kann die Glattheitsrestriktion als quadratische Norm

$$E(u) = \| H(\partial u / \partial x, \partial u / \partial y, \partial^2 u / (\partial x)^2, \partial^2 u / \partial x \partial y, \partial^2 u / (\partial y)^2, \dots \partial^n u / (\partial x)^n \dots) \|^2 \quad (3-1)$$

dargestellt werden, in der $E(u)$ auch als Stabilisierungsfunktional bezeichnet wird. Der Operator H ist eine lineare Funktion der ersten, zweiten oder noch höheren Ableitung der zu bestimmenden Funktion u . Die quadratische Norm - auch als L_2 -Norm bezeichnet - kann als Integral mit quadratischen Argumenten dargestellt werden. Zwei typische Stabilisierungsfunktionale beruhen auf den ersten und zweiten partiellen Ableitungen,

⁵Nicht nur die Bildzuordnung, sondern darüber hinaus eine Vielzahl weiterer sog. "early vision"-Prozesse sind schlecht gestellte Probleme. Dazu zählen die Rekonstruktion von Bewegung und Struktur, andere "structure from x"-Verfahren oder auch die Oberflächenrekonstruktion (vgl. z. B. Poggio et al., 1985).

⁶Inversion steht für die Umkehrung des Prozesses. Ist die Bewegung im Bild bekannt, können die Bildintensitäten leicht entsprechend der gegebenen Bewegung an die neue Bildposition transformiert werden. Umgekehrt ist die Aufgabenstellung wesentlich komplizierter.

wobei⁷

$$E_1(u) = \int_B (u_x^2 + u_y^2) dx dy \quad (3-2)$$

weil es die Energie einer Membran zum Ausdruck bringt, auch als Membranmodell bezeichnet wird. Die Tiefstellung von x und y ist als partielle Ableitung der Funktion zu verstehen, d. h. $u_x = \frac{\partial}{\partial x}u$, etc. Die Energie einer flexiblen dünnen Platte als Maß zur Modellierung von Glattheit lautet

$$E_2(u) = \int_B (u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2) dx dy. \quad (3-3)$$

Die Kombination von E_1 und E_2 nach dem Vorschlag von Terzopoulos(1986)

$$E_{12}(u) = \int_B \varrho(x, y) \{ (1 - \tau(x, y))(u_x^2 + u_y^2) + \tau(x, y)(u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2) \} dx dy \quad (3-4)$$

mit den Gewichtsfunktionen $\varrho, \tau \in [0, 1]$ gestattet es, bei der Oberflächeninterpolation Diskontinuitäten der Oberfläche $u(x, y)$ bezüglich der Höhe und der Orientierung mit zu berücksichtigen. Die Wahl von $\varrho(x, y) = 0$ hebt die sog. C^0 -Kontinuität der Oberfläche auf und gestattet Höhengsprünge entlang der Höhendiskontinuität. Die Vorgabe von $\varrho(x, y) = 1$ und $\tau(x, y) = 0$ führt zu (3-2) und bewahrt lediglich C^0 -Kontinuität, d. h. erlaubt Bruchkanten (Orientierungsdiskontinuitäten) entlang der Diskontinuität der Oberflächenorientierung. Die Festlegung $\tau(x, y) = 1$ führt zum Modell (3-3) mit einer sog. C^1 -Oberfläche, d. h. einer kontinuierlichen Oberfläche mit kontinuierlichen 1. Ableitungen. Die zweiten und höheren Ableitungen müssen hierbei keine Kontinuität besitzen. Damit sind die sehr glatten Regionen des Objekts angesprochen, die weder in der Tiefe noch in der Orientierung Diskontinuitäten aufweisen. Die gewichtete Addition von E_1 und E_2 wird auch als Standardstabilisierer nach Tikhonov benannt und bewahrt C^0 - oder C^1 -Kontinuität der Oberfläche. Der Vorschlag von Terzopoulos ist demgegenüber allgemeiner, da er auch Tiefendiskontinuitäten einbezieht.

Für die Schätzung des optischen Flusses haben Horn und Schunk (1981, S. 188) die Gradienten des Geschwindigkeitsfeldes $\mathbf{u}(x, y, t)$ zur Regularisierung gemäß

$$E_3(\mathbf{u}) = \int_B (u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy \quad (3-5)$$

herangezogen. Bei der Glättung der Geschwindigkeiten u in x -Richtung und v in y -Richtung sind keine gemischten Komponenten integriert, d. h. $E_3(\mathbf{u}) = E_1(u) + E_1(v)$. Der Vergleich mit (3-2) zeigt, daß zur Stabilisierung in beiden Achsrichtungen jeweils ein Membranmodell angesetzt wird. Eine Alternative zu E_3 , die der Regularisierung (3-3) näherkommt, schlagen Horn und Schunk durch die Verwendung des Laplace-Operators $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ gemäß

$$E_4(\mathbf{u}) = \int_B (u_{xx}^2 + u_{yy}^2 + v_{xx}^2 + v_{yy}^2) dx dy \quad (3-6)$$

vor. Auch hier ist direkt ersichtlich, daß $E_4(\mathbf{u})$ in Energiekomponenten separierbar ist, die nur von u bzw. v abhängen. Eine systematische, mathematisch sehr gründliche Aufarbeitung der Glättungsrestriktionen des optischen Flusses, die quadratisch in den ersten Ableitungen und quadratisch in den ersten und zweiten Ableitungen der Bildfunktion sind, ist in Snyder (1989, 1991) dargestellt. Als wichtigste Eigenschaft von (3-5) wird die Invarianz bezüglich der Wahl des Koordinatensystems und die positive Definitheit des Glättungsoperators herausgestellt, die garantiert, daß $E(\mathbf{u})$ ein eindeutiges Minimum besitzt. Die Analogie zur Argumentation von Terzopoulos besteht für E_3, E_4 darin, eine (3-4) entsprechende Verallgemeinerung auszuarbeiten, wodurch Diskontinuitäten des optischen Flusses und dessen erster Ableitung explizit berücksichtigt werden können.

Ohne näher darauf einzugehen haben wir bislang angenommen, daß eine *Funktion* geschätzt werden soll, die im Fall des optischen Flusses darüber hinaus vektorwertig ist. Die diskrete Berechnung, zu der die Regularisierungsoperatoren diskret approximiert werden müssen, benötigt auch eine Diskretisierung der zu schätzenden Funktion. Der optische Fluß $\mathbf{u}(x, y, t)$ stellt dann ein zweidimensionales Vektorfeld zum Bezugszeitpunkt t dar,

⁷Wir wollen Doppelintegrale bzw. Doppelsummen verkürzt mit Angabe des Integrations- bzw. Summationsbereichs B folgendermaßen darstellen: \int_B steht für $\int \int_B$ und \sum_B für $\sum \sum_B$.

wobei durch eine Interpolationsvorschrift innerhalb des Vektorfeldes die explizite Beziehung zur Funktion vorliegt. Die Berechnung auf der Basis des Variationskalküls einschließlich der Diskretisierung zur Bestimmung des optischen Flusses auf diesem Lösungsweg wird im nächsten Abschnitt verfolgt.

Eine alternative Standardtechnik zur diskreten Lösung der kontinuierlich formulierten Optimierungsaufgaben ist die finite Elemente Analyse (Terzopoulos, 1988, Szeliski, 1990). Bei quadratischer Rasterung mit Maschenabstand $h = 1$ lautet die Energiegleichung für das Membranmodell entsprechend (3-2) in diskreter Differenzendarstellung

$$E_5(u) = \sum_B \{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2\}, \quad (3-7)$$

wobei sich (i, j) auf die Position der Knotenpunkte bezieht. Die Abweichungen von der Glättungsrestriktion "dünne Platte" werden durch

$$E_6(u) = \sum_B \{(u_{i+1,j} - 2u_{i,j} + u_{i-1,j})^2 + 2(u_{i+1,j+1} - u_{i,j+1} - u_{i+1,j} + u_{i,j})^2 + (u_{i,j+1} - 2u_{i,j} + u_{i,j-1})^2\} \quad (3-8)$$

ausgedrückt. Die Membran hat Terzopoulos dabei finit durch ein Dreieckselement, die dünne Platte durch ein quadratisches Element beschrieben (Terzopoulos, 1986). Die Variablen in den Knotenpunkten treten an die Stelle der zu schätzenden Funktion. Die zugrundegelegte Maschenweite muß sich nicht an der Rasterung im Bild orientieren. Im Gegenteil, um in der topologischen Festlegung (Dreieck, Rechteck) mit einer simplen Rasterstruktur flexibel zu sein, ist es günstiger, die Festlegung der Lage der Knotenpunkte unabhängig von der Position der beobachteten Daten zu treffen. Das macht deutlich, daß explizit ein Interpolator angegeben werden muß. Tatsächlich hat Terzopoulos bei der diskreten Formulierung des finite Elementansatzes eine entsprechende Formbeschreibung bzw. Interpolationsvorschrift für Membran und dünne Platte zugrunde gelegt. Ist das Verhältnis zwischen Anzahl der Messungen und Anzahl der Knotenpunkte klein, wird durch die Interpolation z. B. beim optischen Fluß aus einer geringen Zahl von Meßstützstellen ein dichtes Vektorfeld ermittelt. Liegen umgekehrt viel mehr Meßdaten als Knotenpunkte vor, so besteht die Funktion des Interpolators vorwiegend in einer Approximationsaufgabe zur Flächenbeschreibung zwischen den Knotenpunkten eines relativ grob diskretisierten finiten Elements. Die Interpolationsvorschrift bei skalaren Funktionen auf regelmäßigem Gitter wird, z. B. beim Resampling von Bildern, oft durch Nächster Nachbar, bilineare oder bikubische Interpolation festgelegt. Eine diskrete Approximation für das Membranmodell (3-2,3-7) hat Terzopoulos durch ein lineares Polynom der Form

$$u(x, y) = ax + by + c$$

beschrieben, das damit die lokale Interpolationsvorschrift innerhalb eines finiten Dreieckselements angibt. Für die diskrete Approximation des Energiefunktionals einer dünnen Platte (3-3,3-8) wählt Terzopoulos ein quadratisches Polynom vom Typ

$$u(x, y) = ax^2 + by^2 + cxy + dx + ey + f.$$

Die Parameter a und b des linearen Polynoms sind durch die diskreten 1. Ableitungen in (3-7), die Parameter a , b und c des quadratischen Polynoms durch die diskreten 2. Ableitungen in (3-8) festgelegt. Das bedeutet, daß bei der Interpolation innerhalb des finiten Elements der dünnen Platte die zweiten partiellen Ableitungen a , b , c konstant, innerhalb des Dreieckselements die ersten Ableitungen a , b konstant sind.

Bei der Rekonstruktion von Geländemodellen aus digitalisierten Luftbildern hat sich im photogrammetrischen Bereich als Interpolator zwischen den Knotenpunkten eines rechteckigen, finiten Elements die Bilinearfläche durchgesetzt (Ebner et al., 1980, Wrobel, 1987, Hahn, 1989). In Experimenten zur Genauigkeitsoptimierung des Gesamtprozesses der Oberflächenerfassung hat sich gezeigt, daß bezogen auf das Bild ein solches finites Element eine Fläche von 30×30 oder 50×50 Bildpunkten abdeckt, wobei in der Regel 20 – 50 korrespondierende Punkte innerhalb dieses Bereichs erfaßt wurden (Hahn und Förstner, 1988). Die explizite Form des Oberflächenmodells innerhalb eines finiten Elements stellt aufgrund der Approximationssituation einen ganz wesentlichen integrierenden und lokal glättenden Aspekt dar, zumal der Bilinearinterpolator Tiefpaßeigenschaften aufweist (Tempfli, 1982). Hinreichend lokale Redundanz vorausgesetzt, haben die gewichtet eingeführten

global wirksamen Glattheitsrestriktionen entsprechend (3-7, 3-8) lokal eine eher geringere Bedeutung für das produzierte Geländemodell. Fehlt allerdings die lokale Information, werden dann über die Regularisierung aus einer auch weitgefaßten Umgebung die Höhen interpoliert. Kehrt sich das hoch redundante Verhältnis zwischen der Anzahl der Beobachtungen und der Anzahl der unbekanntenen Rasterhöhen ins Gegenteil um, so bekommen die Glättungsrestriktion für die gesamte Interpolationaufgabe eine auch lokal sehr gewichtige Bedeutung. Die Regularisierung in den zu schätzenden Knotenpunkten bietet dann neben der Glättung (3-7, 3-8) auch wirksam⁸ die Möglichkeit, entsprechend der Formulierung E_{12} (3-4) Diskontinuitäten bei der Rekonstruktion des Modells zu berücksichtigen.

3.2 Variationsformulierungen zur Schätzung des optischen Flusses

Für die Schätzung des optischen Flusses gilt, wie für viele andere Aufgaben der Computer Vision auch, daß sie als Optimierungsaufgaben dargestellt werden. Zu deren Lösung wird eine Zielfunktion formuliert, die maximiert oder minimiert werden soll. Diese Optimierungsaufgaben zur Bestimmung einer unbekanntenen *Funktion*, z. B. des optischen Flusses, werden als Variationsprobleme bezeichnet und meist mithilfe des Variationskalküls zu lösen versucht (Fischler und Firschein, 1987).

Klassische geodätische und photogrammetrische Aufgaben werden in der Regel ebenfalls als Minimierungsprobleme mit einer entsprechenden Zielfunktion dargestellt. Beobachtet wird im allgemeinen diskret und redundant, und es gilt einen *Satz von Parametern* zu schätzen, die als Funktion der Beobachtungen darstellbar sind. Diese Technik wird als Ausgleichsrechnung bezeichnet und die geschätzten Parameter korrespondieren mit dem Minimum der Zielfunktion.

Bei der Variationsrechnung wird die Schätzproblematik von Parametern um die von Funktionen erweitert. Mehr als die Parameter interessieren die Funktionen, die als Extrema der Minimierungsfunktion zu bestimmen sind. Hierin besteht eine gegenüber Ausgleichsproblemen signifikante Verallgemeinerung der Schätzproblematik. Im folgenden sollen Grundkenntnisse der Ausgleichsrechnung und der Variationsrechnung vorausgesetzt werden, wie sie z. B. den Lehrbüchern von Koch (1980) und Weinstock (1974) zu entnehmen sind. Eine sehr komprimierte Übersicht über das Variationskalkül ist in Büchern zur Bildverarbeitung z. B. in Horn (1986) oder Bässmann und Besslich (1991) zu finden. Die zum weiteren Verständnis wichtigen Beziehungen aus der Variationsrechnung werden kurz erläutert.

3.2.1 Variationsprobleme und Euler'sche Gleichung

Für viele Variationsprobleme ist typisch, daß sie sich in der Form

$$\int_B L dx \rightarrow \min \quad (3-9)$$

darstellen lassen, wobei zum Beispiel

$$L = L(u, u_x, u_y, x, y) \quad (3-10)$$

den fachspezifischen Sachverhalt beschreibt. L ist in diesem Beispiel abhängig von der unbekanntenen Funktion u und deren partiellen Ableitungen u_x, u_y nach den Parametern x bzw. y .

Das Variationsproblem (3-9,3-10) läßt sich auch gemäß⁹

$$E(u(x, y)) = \int_B L(u, u_x, u_y, x, y) dx dy \rightarrow \min \quad (3-11)$$

formulieren, wobei E als *Funktional* bezeichnet wird. Im Gegensatz zu Kleinste-Quadrate-Standardproblemen ist das Zielfunktional $E \rightarrow \min$ primär nicht von Parametern, sondern von der unbekanntenen Funktion $u(x, y)$

⁸Bei der Approximationssituation, wie sie oben dargestellt wurde, mit relativ grob gerasterten Höhenknoten könnte eine ähnliche Wirksamkeit durch eine lokale Verdichtung des Rasterpunktabstandes entlang der Diskontinuität erzielt werden, was aber impliziert, daß dem Schätzprozeß eine entsprechende Datenstruktur zugrundeliegt.

⁹vgl. die Konventionsvereinbarungen im Abschnitt 3.1, Fußnote 7

abhängig. Dieser Funktion u ist eine Zahl - der Wert des Integrals - zugeordnet. Gesucht ist die Funktion, für die das Integral einen minimalen Wert liefert. Zur Berechnung von (3-11) wird das Funktional $E(u)$ in der Umgebung der Lösung u mit einer Funktion η 'variiert', d. h. $\epsilon\eta$ zu u hinzuaddiert. Durch Differentiation nach ϵ ergibt sich aus den notwendigen Bedingungen für das Extremum des Funktionals (3-11) die *Euler'sche Gleichung*

$$L_u - \frac{\partial}{\partial x} L_{u_x} - \frac{\partial}{\partial y} L_{u_y} = 0, \quad (3-12)$$

die in dieser Form eine partielle Differentialgleichung darstellt. Verallgemeinerungen von (3-10) werden hinsichtlich mehrerer Funktionen und höherer Ableitungen benötigt. Ist der Integrand

$$L = L(u, v, u_x, v_x, \dots x) \quad (3-13)$$

von zwei bzw. mehreren Funktionen abhängig, ergeben sich entsprechend zwei bzw. mehrere Euler'sche Gleichungen

$$\begin{aligned} L_u - \frac{d}{dx} L_{u_x} &= 0 \\ L_v - \frac{d}{dx} L_{v_x} &= 0. \end{aligned} \quad (3-14)$$

Die Erweiterung hinsichtlich höherer Ableitungen

$$L = L(u, u_x, u_{xx}, \dots x) \quad (3-15)$$

führt zur Euler'schen Gleichung der Form

$$L_u - \frac{d}{dx} L_{u_x} + \frac{d^2}{dx^2} L_{u_{xx}} - \dots = 0. \quad (3-16)$$

Die Stabilisierungsfunktionale decken einen Aspekt der Problemformulierung (3-11) ab. Ein zweiter ergibt sich aus der beobachtungsspezifischen Modellierung, so daß sich das Gesamtproblem häufig aus zwei Teilen

$$\int_B L \, dx \, dy = \int_B G \, dx \, dy + \lambda \int_B H \, dx \, dy. \quad (3-17)$$

zusammensetzt. G repräsentiert dabei einen aufgabenspezifischen und beobachtungsbezogenen Teil, der wegen der "Schlechtgestelltheit" ("ill-posedness") des Problems durch den regularisierenden Teil H ergänzt wurde. Durch H wird Vorinformation in den Prozeß eingebracht und dieser dadurch in ein gut gestelltes Problem überführt. Der Parameter λ ist der Regularisierungsparameter, der die Rolle der relativen Gewichtung zwischen beobachtendem Term und regularisierendem Term übernimmt. λ hat dabei auch Einfluß auf die Stabilisierung der numerischen Lösung. Da λ nur relativ zu sehen, ist es auch üblich, die beiden Funktionale gemäß $(1 - \lambda) \int_B G \, dx \, dy + \lambda \int_B H \, dx \, dy$ zusammenzusetzen.

3.2.2 Beobachtungsmodelle des optischen Flusses

Die beobachteten Daten für die Rekonstruktion des optischen Flusses sind die in der Bildebene registrierten Intensitäten. Bei den rein geometrischen Betrachtungen zur Beschreibung der projizierten Bewegung als Funktion der Bewegung im Objektraum (Abschnitt 2.3) wurde noch keine Aussage über die Intensitäten selbst getroffen. Die Aufgabe des Beobachtungsmodells für den optischen Fluß ist es nun, die Beziehungen zwischen den erfaßten Intensitätsbildern der Sequenz und der projizierten Bewegung zu formulieren.

Die überwiegende Anzahl der Verfahren zur Schätzung des optischen Flusses beruht auf der Annahme, daß das von einem Punkt des Objektraumes reflektierte Licht innerhalb eines kurzen Zeitintervalls δt konstant bleibt und zu keiner Änderung der Intensität dieses Punktes führt, wenngleich sich die Position des Punktes im Bild infolge der Bewegung im Objektraum verändert. Mathematisch läßt sich diese Annahme als Beobachtungsmodell gemäß

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (3-18)$$

darstellen. Die Reihenentwicklung der rechten Seite von (3-18) bis zur ersten Ordnung (mit $\delta t \rightarrow 0$) oder gleichbedeutend die Differentiation von $I(x, y, t)$ nach der Zeit führt zur Restriktion

$$I_x u + I_y v + I_t = 0. \quad (3-19)$$

I_x, I_y sind die räumlichen Gradienten, I_t die zeitliche Variation der Intensität im Bildpunkt (x, y) zum Bezugszeitpunkt t . (3-19) wird von Horn (1986) als *optical flow constraint equation* bezeichnet, wobei aber auch *intensity-constancy constraint*, *motion constraint equation* oder ähnliche Begriffe in der Literatur verwendet werden. Sie stellt für zwei Bezugszeitpunkte das Beobachtungsmodell dar, das die Änderung der beobachteten Bildintensitäten (I_t) an einem Bildpunkt in Beziehung zu der unbekanntem Verschiebung (u, v) dieses Punktes setzt. Aus der punktbezogenen Interpretation wird offensichtlich, daß diese eine Beobachtungsgleichung es sicherlich nicht gestattet, zwei Parameter zu bestimmen. Geometrisch stellt (3-19) eine Geradengleichung in der (u, v) -Ebene dar (vgl. Abbildung 3.2). Die Geschwindigkeiten (u, v) müssen auf einer Geraden liegen,

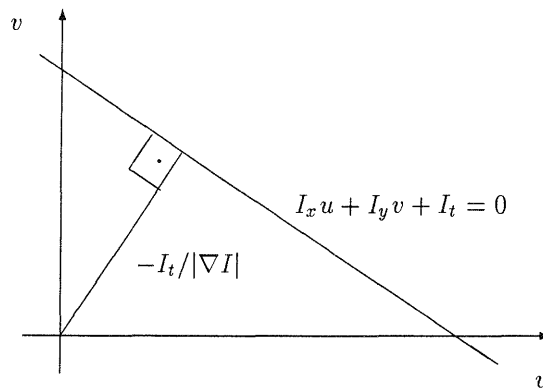


Abbildung 3.2: Geometrische Interpretation der Beobachtungsgleichung

die senkrecht zum Gradientenvektor (I_x, I_y) verläuft. Bezeichnet u_\circ die Bildgeschwindigkeit in Richtung des Grauwertgradienten $\nabla I = (I_x, I_y)^T$ und v_\circ die zu u_\circ orthogonale Komponente der Bildgeschwindigkeit im Bildpunkt (x, y) , so kann (3-19) auch äquivalent durch

$$|\nabla I| u_\circ + I_t = 0 \quad (3-20)$$

dargestellt werden. Bestimmbar ist offensichtlich nur die Komponente der Bewegung in Richtung des Grauwertgradienten ∇I , die sich zu $u_\circ = -I_t/|\nabla I|$ ergibt. Daß keine Information über die Bewegung v_\circ orthogonal zum Bildgradienten, d. h. keine Meßgröße in Richtung des Konturverlaufes gewonnen werden kann, ist als Apertur- oder Blendenproblem (siehe Abschnitt 1.4) bekannt. Wird mit l_\circ die lokal aus (3-20) oder auf andere Art und Weise ermittelte Bewegungskomponente symbolisiert, die auch als *normal flow estimate* bezeichnet wird, mit \mathbf{n}_C der auf Länge 1 normierten Normalenvektor der Kontur im entsprechenden Konturpunkt, so kann eine konturorientierte Beobachtungsgleichung gemäß

$$u_\circ - l_\circ = \mathbf{n}_C^T \mathbf{u} - l_\circ = 0 \quad (3-21)$$

aufgestellt werden, wie sie von Hildreth (1984) zur Schätzung des konturbasierten optischen Flusses verwendet wird.

Konturorientierte Verfahren zur Bestimmung von l_\circ , bei denen die Normalenrichtung nicht intensitätsbasiert über den Bildgradienten festgelegt wird, werden z. B. von Wohn et al. (1991) diskutiert. Wird die Korrespondenz von Konturlinien zweier diskreter Bezugszeitpunkte als gegeben vorausgesetzt, kann l_\circ als Abstand eines Punktes einer Kontur $C(t_1)$ zur korrespondierenden Kontur $C(t_2)$ in Richtung der Konturnormalen $\mathbf{n}_{C(t_1)}$ festgelegt werden. Eine etwas aufwendigere Alternative ist die Festlegung von l_\circ durch Bestimmung des nächstliegenden Punktes unter Berücksichtigung des Normalenvektors der zweiten Kontur $C(t_2)$. Vor- und Nachteile beider Methoden unter Einbeziehung von Kurvenapproximationen für Kontursegmente, insbesondere auch Betrachtungen zum Fehlerverhalten, werden in Wohn et al. (1991) diskutiert.

Kritik an der Beobachtungsgleichung (3-19) wird in letzter Zeit vermehrt geäußert (z. B. Verri und Poggio, 1987, Faugeras, 1990). Sehr eindrucksvoll werden in Horn (1986) Beispiele diskutiert, die zeigen, daß grundsätzlich die Beleuchtung, die Reflexionseigenschaften der Oberfläche und die lokale Orientierung der Oberfläche berücksichtigt werden müssen. Eines der Beispiele beschreibt das Szenario einer rotierenden untexturierten Kugel (mit konstantem Albedo) bei einer ortsfesten Beleuchtungsquelle. Die zeitliche Variation I_t ist identisch 0, da sich die Intensitäten in der Bildebene nicht ändern. Die aufgrund der Abbildungsgleichungen geometrisch definierten projizierten Geschwindigkeiten (Abschnitt 2.3), die wir zur Unterscheidung vom optischen Fluß \mathbf{u} mit $d_t \mathbf{x}$ bezeichnet haben, und die Intensitätsgradienten ∇I sind nahezu im ganzen Bild ungleich 0, so daß die Annahme $\mathbf{u} = d_t \mathbf{x}$ für (3-19) völlig unzutreffend ist. Die umgekehrte Situation, d. h. $\nabla I^T d_t \mathbf{x} = 0$ aber $I_t \neq 0$ ergibt sich, wenn die untexturierte Kugel fest steht und die Beleuchtung sich bewegt. Verri und Poggio (1987) diskutieren die Diskrepanz von Bewegungsfluß und optischem Fluß für Lambert'sche und spiegelnde Oberflächen unter rotatorischer und translatorischer Bewegung. Die wesentliche Erkenntnis aus dieser Arbeit ist, daß optischer Fluß und Bewegungsfluß an solchen Stellen im Bild nah beisammen liegen, an denen ein großer Bildgradient vorliegt. Stark texturierte Punkte im Bild sind z. B. Eckpunkte oder Kantenelemente, die oft auch direkte Konsequenzen einer Diskontinuität im Objektraum sind.

Um die Abhängigkeit von der Beleuchtung und lokalen Oberflächenorientierung mit zu berücksichtigen, schlagen Cornelius und Kanade (1983) vor, (3-19) gemäß

$$q + I_x u + I_y v + I_t = 0 . \quad (3-22)$$

zu erweitern. $q(x, y, t)$ steht hier als zusätzlicher zeitabhängiger Kompensationsterm. Daß dieser zusätzliche Term die Unterbestimmtheit (Aperturproblem) der inversen Aufgabenstellung weiter steigert, ist offensichtlich. Cornelius und Kanade haben deshalb zusätzlich zu den Glättungstermen für u und v (3-5) eine isotrope Glättung bezüglich q hinzugenommen.

Gennert (1988) führt ein multiplikatives Modell zur Transformation der Bildintensitäten zweier diskreter Aufnahmezeitpunkte ein, das kontinuierlich gemäß

$$I(x, y, q, t) = q(x, y, t)I(x, y, t)$$

dargestellt werden kann. Er führt aus, daß der Ansatz dann gerechtfertigt ist, wenn Änderungen in der Albedo signifikanter sind als Änderungen in der nur von der Oberflächenorientierung abhängigen Reflexion¹⁰. Die grundsätzlichen Probleme zur Bestimmung des multiplikativen Terms sind vergleichbar mit denen des additiven Terms in (3-22). Im Unterschied zu Cornelius und Kanade setzt Gennert in seiner Arbeit allerdings epipolare Bilder voraus, so daß hier für die Parallaxen nur eine eindimensionale Suche notwendig ist.

Für die Bewegungsschätzung in kleinen Bildausschnitten wird oft ein Modell verwendet, das eine additive und eine multiplikative Änderung der Bildintensitäten gestattet. Gegenüber den beiden vorhergehenden Modellen wird der Kompensationsterm aber unabhängig vom Ort (innerhalb des Bildfensters) nur als Funktion der Zeit angesetzt:

$$I(x, y, b_0, b_1, t) = b_0(t) + b_1(t) I(x, y, t) . \quad (3-23)$$

Die Differentiation entsprechend dem Vorgehen in (3-18,3-19) führt zur Restriktion

$$d_t b_0 + I d_t b_1 + I_x u + I_y v + I_t = 0 ,$$

wobei zum aktuellen Betrachtungszeitpunkt ($t_0 = t$) die Funktionswerte $b_0(t_0) = 0$, $b_1(t_0) = 1$ gewählt wurden. Bezogen auf zwei diskrete Bezugszeitpunkte sind b_0 , b_1 radiometrische Kompensationsparameter, die oft auch als *offset* und *gain* oder als radiometrische Verschiebungs- und Maßstabsparameter bezeichnet werden.

¹⁰Werden die Bildintensitäten $I(x, y)$ als Produkt eines Reflexionsfaktors $\alpha(x, y)$ - Albedo - und einer Reflexionskarte $R(\mathbf{n}(x, y))$ dargestellt, so ist die Reflexionskarte nur von der Oberflächenorientierung \mathbf{n} des abgebildeten Punktes (x, y) abhängig. Alle anderen - wie z. B. die materialspezifischen - Effekte sind dann der Albedo zugeordnet.

3.2.3 Variationslösung von Horn und Schunk

Da die Beobachtungsgleichung (3-19) ausschließlich auf lokaler Bildinformation beruht und zudem noch unterbestimmt ist, sind verschiedene zusätzliche Restriktionen mit globaler Wirkung vorgeschlagen worden. Besonders populär ist der Vorschlag von Horn und Schunk nicht zuletzt auch deshalb, weil neben der Lösung durch die Euler'sche Gleichung ein sehr einfaches Relaxationsverfahren für die diskret beobachteten Bilder ausgearbeitet wurde.

Das Variationsproblem lautet in dem von Horn und Schunk gewählten Ansatz

$$E(\mathbf{u}) = \lambda \int_B (\nabla I^T \mathbf{u} + I_t)^2 dx dy + \int_B (|\nabla u|^2 + |\nabla v|^2) dx dy \quad (3-24)$$

in dem die Integranden die Funktionen $u(x, y)$, $v(x, y)$ und deren erste Ableitungen nach den Bildkoordinaten enthalten, d. h. das Funktional E ist vom Typ $L = L(u, v, u_x, u_y, v_x, v_y, x, y)$. Der erste Term in (3-24) drückt den Fehler im Beobachtungsmodell aus, da allein schon wegen des Rauschens in den Bildintensitäten die Beobachtungsgleichungen nicht eingehalten werden können. Die Residuenquadrate, die im zweiten Term aufintegriert werden, sind die Abweichungen von der globalen Forderung nach Glattheit (3-5), die zur Regularisierung des Problems führen. (3-24) liegt damit in einer Form entsprechend (3-17) vor, in der λ die beiden quadratischen Fehlerterme relativ zueinander gewichtet.

Die Eulerschen Gleichungen mit $L = L(u, v, u_x, u_y, v_x, v_y, x, y)$ lauten aufgrund von (3-12, 3-14) entsprechend

$$\begin{aligned} L_u - \frac{\partial}{\partial x} L_{u_x} - \frac{\partial}{\partial y} L_{u_y} &= 0 \\ L_v - \frac{\partial}{\partial x} L_{v_x} - \frac{\partial}{\partial y} L_{v_y} &= 0. \end{aligned} \quad (3-25)$$

Zur expliziten Berechnung der Euler'schen Gleichung werden die Ableitungen L_u, L_v benötigt

$$\begin{aligned} L_u &= 2I_x (\nabla I^T \mathbf{u} + I_t) \\ L_v &= 2I_y (\nabla I^T \mathbf{u} + I_t), \end{aligned}$$

die ausschließlich den beobachtungsspezifischen Term betreffen. In den partiellen Ableitungen $\frac{\partial}{\partial x} L_{u_x}, \dots, \frac{\partial}{\partial y} L_{v_y}$

$$\begin{aligned} \frac{\partial}{\partial x} L_{u_x} &= 2\lambda u_{xx}, & \frac{\partial}{\partial y} L_{u_y} &= 2\lambda u_{yy}, \\ \frac{\partial}{\partial x} L_{v_x} &= 2\lambda v_{xx}, & \frac{\partial}{\partial y} L_{v_y} &= 2\lambda v_{yy} \end{aligned}$$

spiegelt sich insbesondere der regularisierende Term wieder. Zusammengefügt ergeben sich die beiden Euler'schen Differentialgleichungen

$$\begin{aligned} \lambda \nabla^2 u &= I_x^2 u + I_x I_y v + I_x I_t \\ \lambda \nabla^2 v &= I_x I_y u + I_y^2 v + I_y I_t, \end{aligned} \quad (3-26)$$

wobei ∇^2 der Laplace-Operator ist. Die Wirkung dieser Gleichungen wollen wir in Abhängigkeit von der lokalen Bildinformation näher betrachten. In Regionen mit einheitlicher Intensität, d.h. in Regionen ohne jeglichen Kontrast, sind die Intensitätsgradienten gleich Null, womit auch die rechten Seiten dieser Gleichungen zu Null werden. Der Beobachtungsterm liefert keinen Beitrag zur Bestimmung des optischen Flusses, so daß die lokale Bildgeschwindigkeit durch den regularisierenden Laplace-Operator in beiden Richtungen aus der Umgebung des Punktes interpoliert wird. Entlang Intensitätskanten ist die Bewegung der Kante in Richtung der Intensitätsgradienten entsprechend (3-20) bestimmbar. Die Komponente in Richtung des Kantenverlaufs wird durch die Glattheitsrestriktion wiederum aus der Umgebung entlang des Kantenverlaufs interpoliert. Die lokale Umgebung einer Grauwertecke schließlich gestattet die Bestimmung des optischen Flusses in u und v , so daß hier die zuverlässigste Information für die globale Glättung über das Bild vorliegt. Die Glattheitsrestriktion, das zeigt (3-26) auch, ist für jeden Bildpunkt identisch modelliert, d. h. Glättung findet ebenso in Eckpunkten wie auch quer zum Kantenverlauf statt.

Um zu einem Lösungsverfahren auf dem diskreten Bild zu gelangen, muß noch der Übergang von der kontinuierlichen Formulierung zur Differenzendarstellung ausgearbeitet werden. Für die Entwicklung einer Iterationsformel soll die Differenzgleichung in eine Form $u = f(u)$ überführt werden, woraus sich ein Iterationsschema $u^{(n+1)} = f(u^n)$ entwerfen läßt. Diese Überführung einer Operatorgleichung in ein (iteratives) Lösungsschema ist ein sehr wichtiger Aspekt für die Lösung der gestellten Variationsaufgabe.

Horn und Schunk approximieren den Laplace-Operator durch einen Differenzenoperator gemäß

$$\nabla^2 \mathbf{u} = \kappa(\bar{\mathbf{u}} - \mathbf{u}),$$

wobei sich $\bar{\mathbf{u}}$ durch Faltung mit einem Glättungsoperator aus \mathbf{u} ergibt: $\bar{\mathbf{u}} = P * \mathbf{u}$. Der Proportionalitätsfaktor κ und der Operator P sind so zu wählen, daß sie den Laplace-Operator geeignet annähern. Die Festlegung von P als gewichteten Mittelwertoperator einer 3×3 Umgebung, der sich daraus ergebende Proportionalitätsfaktor $\kappa = 3$ und der Differenzenoperator zur Berechnung von $\bar{\mathbf{u}} - \mathbf{u}$ sind in Horn und Schunk (1981) explizit angegeben. Der Differenzenoperator ist identisch mit der Differenz zweier Binomialkerne, wobei der Laplace explizit durch $B_{\sigma^2=1/2} - B_{\sigma^2=0}$ approximiert wird. Der Nachweis kann durch einfaches Einsetzen nachvollzogen werden. Das Iterationsschema zur Schätzung des optischen Flusses in allen Bildelementen (i, j) lautet schließlich (Horn und Schunk, 1981):

$$\begin{aligned} u_{i,j}^{(n+1)} &= \bar{u}_{i,j}^{(n)} - \frac{I_x (\nabla I^T \bar{\mathbf{u}}_{i,j}^{(n)} + I_t)}{\alpha^2 + |\nabla I|^2} \\ v_{i,j}^{(n+1)} &= \bar{v}_{i,j}^{(n)} - \frac{I_y (\nabla I^T \bar{\mathbf{u}}_{i,j}^{(n)} + I_t)}{\alpha^2 + |\nabla I|^2} \end{aligned} \quad (3-27)$$

Durch diesen Relaxationsalgorithmus werden die Bildgeschwindigkeiten an allen Bildpunkten iterativ unter Einbeziehung der Information aus der Nachbarschaft ermittelt.

3.2.4 Weitere Ansätze von Hildreth und von Nagel

Mit den Ansätzen von Hildreth (1984) und Nagel (1985, 1987), Nagel und Enkelmann (1986) sollen zwei weitere richtungswisende Arbeiten zur Schätzung des optischen Flusses aufgezeigt werden. Da Intensitätskanten die signifikanteste Information im Bild beinhalten und zudem die Bildgeschwindigkeiten in den Bildbereichen mit starker Textur zuverlässiger bestimmt werden können, hat Hildreth einen konturbasierten Vorschlag zur Bestimmung des optischen Flusses unterbreitet. Zur Überwindung des Aperturproblems bei der Bewegung von Kurven, wie den zero-crossing Konturen eines mit dem *LoG*-(Laplacian of Gaussian)-Operator vorgefilterten Bildes, hat Hildreth Glattheit der Geschwindigkeiten entlang der Kontur gefordert. Zusammen mit dem Beobachtungsmodell (3-21) lautet dann das zu minimierende Funktional

$$E(\mathbf{u}) = \oint_C (\lambda(u_\circ - l_\circ)^2 + u_s^2 + v_s^2) ds. \quad (3-28)$$

An die Stelle des Flächenintegrals tritt das Linienintegral; s bezeichnet die Bogenlänge entlang der Konturen C , über die innerhalb des Bildes integriert wird. Die Regularisierung sorgt dafür, daß jeweils die Geschwindigkeiten (u_s, v_s) in Nachbarerelementen der Kurve nur entlang des Konturverlaufes geglättet werden. u_\circ ist die unbekannte Komponente der Geschwindigkeit \mathbf{u} in Richtung \mathbf{n}_C orthogonal zur Kante, die gemäß $u_\circ = \mathbf{n}_C^T \mathbf{u}$ als Funktion des unbekanntes Geschwindigkeitsvektors \mathbf{u} im Bildpunkt dargestellt werden kann und für die eine entsprechende Beobachtung durch l_\circ vorliegt. Falls die Kontur durch eine Punktfolge mit einheitlichem Intensitätswert beschrieben werden kann, ist die Normalenrichtung dieses Konturverlaufs identisch mit der Gradientenrichtung ($\mathbf{n} = \nabla I / |\nabla I|$). Wird l_\circ aus (3-20) bestimmt, so ist der beobachtungsbezogene Term dem der Formulierung von Horn und Schunk ähnlich, nicht aber die Glättungsrestriktion. Die spezielle Wahl der Kontur und des Meßwertes l_\circ machen auch deutlich, daß der Ansatz (3-28) eine allgemeinere Klasse von Verfahren zur Schätzung des optischen Flusses entlang der Konturen im Bild ermöglicht. Weitere konturorientierte Verfahren zur Bestimmung von l_\circ , bei denen die Normalenrichtung nicht intensitätsbasiert über den Bildgradienten

festgelegt wird, haben wir unter den Beobachtungsmodellen des optischen Flusses zuvor aufgeführt (siehe auch Wohn et al., 1991).

Nagel's Arbeit stellt eine weitere Verallgemeinerung der bisherigen Ansätze dar. Das Ziel ist ein dichtes Bewegungsvektorfeld, wie es auch bei der algorithmischen Lösung von Horn und Schunk erzielt wird. Im Bereich möglicher Bewegungsdiskontinuitäten soll die Glattheitsrestriktion allerdings eingeschränkt bzw. aufgehoben werden, was Hildreth durch die Forderung nach Glattheit ausschließlich entlang der Konturen erwirkt. Die von Nagel propagierte Glattheitsrestriktion wird als gerichtete Glattheit bezeichnet, da für jedes Bildelement eine richtungsabhängige Gewichtung stattfindet. Der physikalische Hintergrund ist auch bei dieser Argumentation, daß durch eine sehr geringe Gewichtung der Glattheitsrestriktion orthogonal zur Bildkante Diskontinuität des optischen Flusses erlaubt wird.

Mit dem Beobachtungsmodell (3-19) lautet dann die Minimierungsaufgabe in dem von Nagel gewählten Ansatz

$$E(\mathbf{u}) = \lambda \int_B (\nabla I^T \mathbf{u} + I_t)^2 dx dy + \int_B (\nabla u^T W \nabla u + \nabla v^T W \nabla v) dx dy \rightarrow \min, \quad (3-29)$$

in dem die symmetrische, positiv definite Gewichtsmatrix W gemäß

$$W = \frac{F}{\text{trace } F} \quad \text{oder} \quad W = \frac{F}{\det F} \quad \text{oder} \quad W = F$$

mit

$$\begin{aligned} F &= \mathbf{t} \mathbf{t}^T + \alpha^2 \bar{H} \bar{H}^T \\ \mathbf{t} &= (I_y, -I_x)^T \\ \bar{H} &= \begin{pmatrix} I_{yy} & -I_{xy} \\ -I_{xy} & I_{xx} \end{pmatrix} \end{aligned}$$

definiert ist. Der Vektor \mathbf{t} ist ein Tangentenvektor der Intensitätskontur, der durch Rotation um 90° aus dem Gradientenvektor ∇I hervorgeht. Die Matrix \bar{H} enthält die zweiten Ableitungen der Intensitätsfunktion. Sie ist durch dieselbe 90° Rotation des Koordinatensystems in die Hesse-Matrix H (Abschnitt 4.2.3) überführbar. Durch α^2 wird die Gewichtung der beiden die ersten und zweiten Ableitungen von I enthaltenden Matrizen zueinander festgelegt.

Wird anstelle der Gewichtsmatrix W in (3-29) die Einheitsmatrix gewählt, so ergibt sich (3-24), d. h. der von Horn und Schunk gewählte Ansatz. Um die Wirkung der Gewichtsmatrix besser erkennen zu können, folgen wir der Argumentation von Nagel und Enkelmann (1986). Die Variation des optischen Flusses eines Punktes in Richtung der Intensitätskontur ergibt den Zeilenvektor

$$\mathbf{t}^T (\nabla u, \nabla v),$$

dessen Norm in der Form

$$|(\nabla u, \nabla v)^T \mathbf{t}| = \text{trace} \{ (\nabla u, \nabla v)^T \mathbf{t} \mathbf{t}^T (\nabla u, \nabla v) \} = \nabla u^T \mathbf{t} \mathbf{t}^T \nabla u + \nabla v^T \mathbf{t} \mathbf{t}^T \nabla v$$

dargestellt werden kann. Der Glattheitsterm bezieht diese Norm ein, so daß die Minimierung starke Glättung in Richtung der Tangente fordert. Die Variation von \mathbf{u} in Richtung der Hauptkrümmungsachse mit der schwächeren Krümmung kann durch die Matrix

$$\bar{H}^T (\nabla u, \nabla v)$$

dargestellt werden. Die Norm, für jede der beiden Zeilen dieser Matrix berechnet, führt aufsummiert zu

$$\text{trace} \{ (\nabla u, \nabla v)^T \bar{H} \bar{H}^T (\nabla u, \nabla v) \} = \nabla u^T \bar{H} \bar{H}^T \nabla u + \nabla v^T \bar{H} \bar{H}^T \nabla v.$$

Diese Norm trägt den zweiten Teil der Glattheitsrestriktion bei, so daß hier die Minimierung eine starke Glättung in Richtung der Hauptkrümmungsachse mit der kleineren Krümmung erzwingt. Durch die Einführung der gerichteten Glattheit wird erreicht, daß an Punkten entlang linearer Bildstrukturen und an Positionen mit starker

Krümmung der Intensitätskonturen in Richtung der Kontur stark, orthogonal zum Konturverlauf schwach geglättet wird.

Da die Euler-Lagrange Gleichungen für die Lösung von (3-29) Ableitungen 3. Ordnung in der Bildfunktion beinhalten, außerdem numerische Probleme für eine Vereinfachung des formulierten Ansatzes sprechen, hat Nagel (1987) die Funktion der ersten Ableitungen von \mathbf{u} und zweiten Ableitungen von I aus differentialgeometrischer Sicht in ihrer Bedeutung für die Minimierungsaufgabe (3-29) diskutiert. Eine der Schlußfolgerungen ist die vereinfachte Modellierung der Gewichtsmatrix W gemäß

$$W = \frac{F}{\text{trace } F} \quad \text{mit} \quad F = \mathbf{t}\mathbf{t}^T + \alpha^2 E, \quad (3-30)$$

in der E eine 2×2 Einheitsmatrix repräsentiert. Durch $\alpha^2 \neq 0$ wird garantiert, daß wenn die Intensitätsstruktur des Bildes lokal konstant ist, der Bereich isotrop geglättet wird. Stark strukturierte Bereiche entlang von Intensitätskonturen üben über die ersten Ableitungen der Bildfunktion einen entsprechend großen Einfluß in Form von orientierter Glattheit aus.

Der Ansatz von Nagel zeigt einige Aspekte, auf die wir bei der Merkmalsextraktion zurückkommen werden. Tatsächlich hat Nagel (1983) zunächst auch Methoden zur Verfolgung von Grauwertecken vorgeschlagen, die dann schrittweise zum Konzept der gerichteten Glattheit erweitert wurden.

3.2.5 Diskussion

Die isotrope Glattheitsforderung, die von Horn und Schunk präsentiert wurde, gilt nicht entlang von Bewegungsdiskontinuitäten. Sowohl die Formulierung des konturorientierten Prozesses, in der von Hildreth ausgearbeiteten Form, als auch das von Nagel präsentierte Konzept orientierter Glattheit sind beides Vorschläge, die Bewegungsdiskontinuitäten gestatten, solange diese mit Intensitätsdiskontinuitäten korrespondieren. Die Korrespondenz ist aus physikalischer Sicht begründet, denn an den Rändern der Objekte oder Oberflächen treten 3D-Tiefen- oder Bewegungsdiskontinuitäten auf, die in der Regel sowohl Diskontinuitäten im optischen Fluß als auch im Intensitätsbild zur Folge haben. Da aber Konturen im Bild oft auch Texturgrenzen sind, die nicht mit Objekträndern einhergehen, wäre an solchen Stellen eine eher isotrope Glättung erwünscht.

Ein Nachteil bei den vorgestellten Schätzprozeduren zur Bestimmung des optischen Flusses ist, daß keine explizite Repräsentation der Bewegungsdiskontinuitäten in den Ansätzen vorgenommen wird. Einen Schritt in diese Richtung stellt die konturbasierte optische Flußbestimmung dar, wenn das zu minimierende Funktional (3-28) für jede geschlossene Kontur eine regionenorientierte Glattheit fordert. Einen ähnlichen Vorschlag haben Cornelius und Kanade (1983) zur Verbesserung des flächenbasierten Ansatzes (3-23) unterbreitet, indem sie Glattheit nur innerhalb der von Konturen begrenzten Regionen fordern. Aus einer zweiten Schätzung des optischen Flusses mit isotroper Glattheit in der gesamten Bildebene nach Horn und Schunk sollen dann durch Überlagerung der Resultate die Bewegungsdiskontinuitäten detektiert werden. Schnörr (1992) formuliert den Ansatz von Horn und Schunk zur Ermittlung eines diskontinuierlichen optischen Flusses um. Er untergliedert die Bildebene in Bereiche, die durch Bewegungsdiskontinuitätsgrenzen voneinander getrennt sind, wobei weder implizit noch explizit Intensitätskonturen zur Formulierung herangezogen werden. Die Minimierungsaufgabe setzt sich dann aus der Summe der für jede einzelne Region formulierten Funktionale (3-24) zusammen. Schnörr leitet ein Verfahren ab, das bei der Bestimmung des optischen Flusses die Diskontinuitätsgrenzen iterativ solange verändert, bis eine Lösung des diskontinuierlichen optischen Flußfeldes gefunden ist. Bislang ist dieses Verfahren auf die Vorgabe guter Näherungswerte für die Bewegungssegmentierung angewiesen.

Ein weiteres Problem dieser gradientenbasierten Ansätze liegt in der kontinuierlichen Formulierung der Optimierungsaufgabe. Die räumlichen und zeitlichen Gradienten der Bildintensitäten sowie die Geschwindigkeiten der Intensitätsstrukturen sind örtlich und zeitlich punktbezogen. Eine Konsequenz daraus ist, daß die auf diskrete Approximationen der kontinuierlichen Formulierung angewiesenen algorithmischen Lösungen nur bei geringen Verschiebungen korrekt arbeiten. Da bei realen Bildfolgen nur selten von generell geringen Verschiebungen innerhalb der gesamten Bildebene ausgegangen werden kann, wurden Mehrebenenprozeduren zu diesen Ansätzen entwickelt. Das zweite Handicap dieser meist als Relaxationsalgorithmen entwickelten Lösungswege ist ein sehr langsames Konvergenzverhalten, das aber ebenfalls durch Mehrebenenverfahren erheblich verbessert

werden kann. Eine hierarchische Prozedur zum Ansatz von Horn und Schunk wurde von Glazer (1984), zum Nagel'schen Ansatz von Enkelmann (1986) vorgestellt. Beide nutzen Gauß'sche Bildpyramiden, starten auf einer groben Auflösungsstufe, schätzen ein Verschiebungsvektorfeld und projizieren dieses auf die nächst feinere Auflösungsstufe, wo es dann als Vorinformation zur Fortsetzung des Schätzprozesses dient. Diese grob skizzierte Vorgehensweise gilt nahezu bei allen Mehrebenenprozeduren; prozedurale Feinheiten (Enkelmann, 1986), lokal variable Gitterweiten (Battiti et al., 1991) oder auch sehr effektive Mehrgitterverfahren sind Themen innerhalb dieses groben Schemas, die heute intensiv erforscht werden. Eine Übersicht gibt Szeliski (1990).

Zur diskreten Berechnung von Ableitungen, z. B. der 2. Ableitungen der Bildfunktion nach dem Ort, wie sie zur Aufstellung der Matrix \bar{H} benötigt werden, muß ein lokales Fenster um den betrachteten Bildpunkt in die Berechnung einbezogen werden. Dabei wird in der Regel innerhalb des Bereichs zusätzlich geglättet, um den Einfluß des Rauschens zu unterdrücken. Diese lokale, fensterbezogene Betrachtung hat Anandan (1987b) auf die Bewegungsparameter selbst ausgedehnt, indem er neben einer globalen isotropen Glattheit annimmt, daß das Bewegungsfeld lokal, z. B. innerhalb eines 5×5 Fensters, durch eine Translation approximiert werden kann. Die lokale Bewegung wird dann durch Korrelation bestimmt. Anstelle der punktbezogenen Beobachtungsgleichungen, die wir oben betrachtet haben, tritt dann das Resultat einer lokalen, fensterbezogenen Messung der Verschiebungsvektoren in allen Bildelementen (siehe Abschnitt 3.3). In Kombination mit der isotropen Glattheit (3-5) erfolgt daraus die Bestimmung eines dichten Flußfeldes.

Die Einbeziehung geometrischer Approximationen der Objektform bei lokaler Betrachtung von Szenenausschnitten ist eine weitere Möglichkeit, um eine geglättete Bewegungsschätzung in örtlich begrenzten Bildausschnitten zu erzielen.

3.3 Bewegungsschätzung in örtlich begrenzten Bildausschnitten

Durch die Einführung von Glattheitsrestriktionen in Form von Stabilisierungsfunktionalen (Abschnitt 3.1) wurde eine Regularisierungsvorschrift gewählt, mit der sich kinematisch oder dynamisch bedingte Änderungen in der 3D-Welt in ihrer Wirkung auf die projizierte Bewegung im Bild in einer sehr allgemeinen Art und Weise berücksichtigen lassen. Selbst wenn die abgebildeten Objekte deformieren, sind die im vorhergehenden Abschnitt vorgestellten Konzepte zur Schätzung des optischen Flusses noch sehr nützlich (Horn, 1986), zumal die Heuristik, die hinter der sehr allgemeinen Modellierung von Glattheit steht, lediglich Kohärenz der Bewegung in der Szene, bzw. Kontinuität in der Bildebene zum Ausdruck bringt.

Wir wollen in diesem Abschnitt von dem an der Kontinuität orientierten Modell der Glattheit (Abschnitt 3.2) über die Starrheit hin zu spezifischeren Approximationen der Modellform vordringen. Spezifischere Formapproximationen haben in der Regel nur lokale Gültigkeit, so daß die Bewegungsschätzung auf kleinere Bildausschnitte begrenzt ist. Für manche Aufgaben, wie z. B. bei der autonomen Fahrzeugführung, kann es nicht zuletzt hinsichtlich der Effizienz vorteilhaft sein, die Bewegungsmessung nicht im gesamten Bild, sondern nur in einigen wenigen ausgewählten Bildausschnitten vorzunehmen. Mit einer weiteren Verringerung der Größe eines Bildausschnittes, z. B. über die 5×5 Fenster, wie sie Anandan (1987b) zur Korrelation verwendet hat, bis hin zum einzelnen Bildelement, wird deutlich werden, daß Stabilisierungsfunktionale und spezifischere Formmodelle sowohl in Konkurrenz zueinander stehen, als auch als Komplement zu betrachten sind. Eine typische Konkurrenzsituation ist gegeben, wenn das Bild in kleine Bildausschnitte unterteilt wird und innerhalb jedes Ausschnittes eine Glattheitsrestriktion eingebracht wird, die sich aus einer entsprechenden Formapproximation des abgebildeten Oberflächenausschnittes ergibt. Eine sich gegenseitig stützende Situation besteht, wenn durch die lokale Formapproximation die lokale Bewegungsmessung verbessert und durch die Stabilisierungsfunktionale (Glattheit in der gesamten Bildebene) auch sehr große Bildbereiche ohne hinreichende Textur überbrückt werden können.

3.3.1 Starre Szene

Die noch sehr allgemeine, unspezifische Modellierung der Szene zur Formulierung von Bewegung durch die Annahme von Starrheit wurde im Kapitel 2 sehr ausführlich diskutiert. Bei punktbezogener Betrachtung kann

der Bewegungsfluß (2-39) entsprechend

$$d_t \mathbf{x}(x, y) = L_V(x, y) \omega + L_Z(x, y) \frac{1}{Z} d_t \mathbf{T}$$

dargestellt werden, was im Abschnitt 2.3.3 abgeleitet wurde. Die Matrizen $L_V = LV$ und $L_Z = -LZ$ sind gemäß (2-38, 2-35)

$$L_V = \frac{1}{c} \begin{pmatrix} xy & -c^2 - x^2 & cy \\ -c^2 - x^2 & xy & -cx \end{pmatrix} \quad L_Z = \begin{pmatrix} -c & 0 & x \\ 0 & -c & y \end{pmatrix}$$

als Funktion der Bildkoordinaten eingeführt worden. Unter der Voraussetzung, daß der optische Fluß und die geometrisch definierte projizierte Geschwindigkeit identisch sind, d. h.

$$\mathbf{u} \stackrel{!}{=} d_t \mathbf{x},$$

können prinzipiell alle möglichen bereits diskutierten Beobachtungsgleichungen (3-18, 3-19, 3-20, 3-21, 3-22, 3-23) des optischen Flusses mit der Interpretation als Starrkörperbewegung zusammengeführt werden. Mit dem Beobachtungsmodell $\nabla I^T \mathbf{u} + I_t = 0$ (3-19) ergibt sich dann

$$I_t + \nabla I^T L_V \omega + \nabla I^T L_Z \frac{1}{Z} d_t \mathbf{T} = 0 \quad (3-31)$$

die *brightness change constraint equation* für die Starrkörperbewegung. Bei punkt- bzw. bildelementbezogener Betrachtung beinhaltet (3-31) neben den 6 Bewegungsparametern ($\omega, d_t \mathbf{T}$) zusätzlich N Strukturparameter (Z), die zu ermitteln sind. Dabei entsteht bei N Bildpunkten aus (3-31) ein (nichtlineares) Gleichungssystem von N Gleichungen und $N + 6$ Unbekannten, d. h. das System ist unterbestimmt. Die Arbeiten in diesem Bereich, die von Horn, Weldon, Negadaripour, u. a. (siehe Abschnitt 2.4.1) vorgestellt wurden, haben noch keine praxistauglichen Experimente vorweisen können. Ungelöst sind noch die Fragen, ob und ggf. welche Art der Regularisierung der Struktur- und Bewegungsparameter sich zur Lösung der schlecht gestellten Aufgabe am besten eignet und alternativ dazu, ob und wie die Erweiterung von zwei auf mehrere diskrete Bezugszeitpunkte erfolgen sollte. Die zweite Möglichkeit wirkt sich allerdings nur dann positiv auf die Redundanz aus, wenn die Strukturparameter nicht zu jedem Zeitpunkt unabhängig angesetzt werden. Die Bewegungs- und Strukturparameter haben in (3-31) offensichtlich auch die Funktion von Ersatzgrößen zur Repräsentation der Bewegung im Bild. Durch den Transfer zu einem ortsfesten Koordinatensystem, wie er in Abschnitt 2.4.3 vorgeschlagen wird, ändern sich die Strukturparameter aller stationären Objekte der Szene nicht mit der Zeit. Damit ist ein möglicher Weg der über die Zeit integrierten Bewegungsschätzung (und Repräsentation) angedeutet. Die Rückprojektion der 3D-Position in die Bildebene gibt die Bewegungsbahn des Objektpunktes im Bild an. Die bei der pixelbezogenen Betrachtung sehr große Zahl der Strukturparameter ist bei einer 3D-Rekonstruktion, wie z. B. einer 3D-Kalman-Filterung, wegen der daraus folgenden beträchtlichen Dimension der Systemgleichungen unüblich und nur durch stark vereinfachende Maßnahmen zu bewältigen. Matthies (1989) hat z. B. jegliche Korrelation unter den Strukturparametern innerhalb der Kalman-Filterung vernachlässigt.

Das gegenteilige Extrem ergibt sich aus den Beziehungen (3-31), wenn die Tiefe eines jeden Bildpunktes vorliegt, z. B. durch einen Distanzmesser¹¹ ermittelt wurde. Es müssen dann lediglich 6 Bewegungsparameter (pro bewegtem Objekt) bestimmt werden, so daß bei N Bildpunkten dies zu einer entsprechend hohen Redundanz führt. Bei mehr als einem Objekt stellt das Hauptproblem die Segmentierung dar, d. h. die Erfassung und Zuordnung der Bildregionen zum entsprechenden Objekt, aber auch zu Objektteilen, die sich durch Diskontinuitäten in der Tiefe unterscheiden. Diese grundlegende Problematik haben wir bereits im vorhergehenden Abschnitt kennengelernt.

3.3.2 Lokal schwach gekrümmte Flächen

Zwischen den beiden gegensätzlichen Voraussetzungen - alle Strukturparameter unbekannt vs. alle Strukturparameter erfaßt - gibt es viele Möglichkeiten, Vorwissen über die Objektform einzubringen. Häufig verwendet

¹¹Die Entwicklung eines solchen "integrierten Distanzmessers", der zu jedem Bildelement die Bildintensität und die entsprechende Tiefe mißt, wurde in Wehr (1991) vorgestellt. Bislang arbeiten flächenhaft scannende Distanzmesser allerdings sequentiell und erreichen noch nicht die hohe zeitliche Abtastrate einer CCD-Kamera, so daß sie zur Aufnahme dynamischer Szenen nur in geringem Maße geeignet sind.

wird die Approximation lokaler Bereiche einer Szene durch Ebenen oder quadratische Flächen (Waxman und Wohn, 1985). Geometrisch komplexere Oberflächen müssen unterteilt werden, bis die Approximation hinreichend genau zutrifft. Für gekrümmte Oberflächen, die durch eine quadratische Oberflächenform beschrieben werden können, ist die projizierte Bewegung gemäß (2-45) durch ein quadratisches Polynom mit 12 Parametern approximierbar:

$$\begin{aligned} u(x, y) &= a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 \\ v(x, y) &= a_6 + a_7x + a_8y + a_9x^2 + a_{10}xy + a_{11}y^2. \end{aligned} \quad (3-32)$$

Die formale Generalisierung auf bivariate Polynome von Grad $m - 1$ gemäß

$$\begin{aligned} u(x, y) &= \mathbf{p}_y^T A_u \mathbf{p}_x \\ v(x, y) &= \mathbf{p}_y^T A_v \mathbf{p}_x \end{aligned}$$

mit den $(m \times m)$ -dimensionalen Parametermatrizen A_u, A_v und den $(1 \times m)$ -dimensionalen Vektoren

$$\begin{aligned} \mathbf{p}_x^T &= (x^0 \ x^1 \ x^2 \ \dots \ x^{m-1}) \\ \mathbf{p}_y^T &= (y^0 \ y^1 \ y^2 \ \dots \ y^{m-1}) \end{aligned}$$

zeigt einen prinzipiell möglichen Weg auf, um komplexere Flächenabschnitte oder größere Bereiche zu modellieren, hat aber vor allem bei höheren Polynomgraden viele negative Begleiterscheinungen (z. B. lokales Aus- oder Überspringen der Funktion).

3.3.3 Lokal ebene Objekte

Kann ein Objekt lokal durch eine 3D-Ebene beschrieben werden, so führt die projizierte Starrkörperbewegung gemäß (2-41) zu

$$\begin{aligned} u(x, y) &= a_0 + a_1x + a_2y + (a_6x + a_7y)x \\ v(x, y) &= a_3 + a_4x + a_5y + (a_6x + a_7y)y. \end{aligned} \quad (3-33)$$

Das optische Flußfeld (u, v) ist ein Polynom zweiter Ordnung in den Bildkoordinaten, das durch 8 Parameter beschrieben wird. Diese 8 Parameter lassen sich im Prinzip schon aus sehr kleinen Bildausschnitten, z. B. bei 10×10 , 15×15 Fenstern, hochredundant ($N - 8$) bestimmen. Die achtparametrische Beschreibung des optischen Flusses gilt streng in der gesamten Bildebene, wenn das abgebildete ebene Objekt den ganzen Bildinhalt abdeckt. Darüber hinaus ist (3-33) eine Approximation mit lokaler Gültigkeit für schwach gekrümmte Oberflächen (Waxman et al., 1987).

3.3.4 Differentialgeometrische Interpretation

Betrachtet man (3-32) aus differentialgeometrischer Sicht, so stellt jede der beiden Gleichungen eine Taylor-Reihe der Funktionen $u(x, y)$ bzw. $v(x, y)$ dar, wobei Glieder 3. und höherer Ordnung abgeschnitten wurden. Bezugspunkt der Reihenentwicklung ist der Ursprung des Koordinatensystems ($\mathbf{x}_0 = 0$). Das bedeutet, das (3-32) auch durch

$$\begin{aligned} \mathbf{u}(\mathbf{x}) &= \sum_{\substack{i=0 \\ (i+j \leq 2)}}^2 \sum_{j=0}^2 \frac{\partial^{i+j} \mathbf{u}}{\partial x^i \partial y^j} \Big|_{\mathbf{x}_0} \frac{x^i y^j}{i! j!} \\ &= \mathbf{u}_0 + \mathbf{u}_x x + \mathbf{u}_y y + \frac{1}{2} \mathbf{u}_{xx} x^2 + \mathbf{u}_{xy} xy + \frac{1}{2} \mathbf{u}_{yy} y^2 \end{aligned} \quad (3-34)$$

dargestellt werden kann, wobei $\mathbf{u}_0 = (a_0, a_6)^T$ und $\mathbf{u}_x = \frac{\partial}{\partial x} \mathbf{u}$, etc. die partiellen Ableitungen von (u, v) nach den Bildkoordinaten darstellen, die alle auf den Punkt $\mathbf{x}_0 = 0$ bezogen sind.¹² Die unbekannt Parameter

¹²Die spezielle Wahl des Bezugspunktes hat den großen Vorteil, daß die Beziehungen zwischen (3-32) und (3-34) direkt ablesbar sind. Für Bildausschnitte, die nicht im Ursprung des Bildkoordinatensystems zentriert sind, kann ein neues, entsprechend zentriertes Koordinatensystem in der durch den Bezugspunkt des Bildausschnittes definierten Richtung festgelegt werden, wobei die Transformation zwischen beiden Koordinatensystemen in einer einfachen Rotation um das festgehaltene Projektionszentrum besteht (Kanatani, 1990, S. 10). Der lokale Bildfluß kann über diese Transformation in ein zentriertes Koordinatensystem überführt werden, so daß die Diskussion auf die zentrierte Problemstellung beschränkt bleiben kann.

$(a_0, a_1, \dots, a_{11})$, deren Beziehung zu den Bewegungs- und Strukturparametern einer lokalen quadratischen Oberfläche im Abschnitt 2.3.5 diskutiert wurde, haben durch (3-34) eine weitere Interpretation als Verschiebung (u_0, v_0) und partielle Ableitungen (u_x, \dots, v_{yy}) des optischen Flusses im Bezugspunkt \mathbf{x}_0 erhalten. Diese sind konkret durch

$$\begin{aligned} u_x &= a_1, & u_y &= a_2, & \frac{1}{2}u_{xx} &= a_3, & u_{xy} &= a_4, & \frac{1}{2}u_{yy} &= a_5, \\ v_x &= a_7, & v_y &= a_8, & \frac{1}{2}v_{xx} &= a_9, & v_{xy} &= a_{10}, & \frac{1}{2}v_{yy} &= a_{11} \end{aligned}$$

gegeben. Diese differentialgeometrische Betrachtung zeigt die Beziehung zu den Regularisierungsansätzen auf, die auf globaler Glattheit beruhen. Ganz offensichtlich ist allen Ansätzen gemeinsam, daß die ersten und zweiten Ableitungen des optischen Flusses bei der Bestimmung der Bewegung im Bild, wenn auch in unterschiedlicher Art und Weise, involviert sind. Beim Ansatz von Horn und Schunk (3-24) werden die Gradientenquadrate des optischen Flusses zusammen mit den Residuenquadraten der lokalen Messungen minimiert. Mit demselben Beobachtungsmodell (3-19) und der polynomischen Beschreibung der projizierten Bewegung der gekrümmten Oberfläche (3-32) lautet die zu minimierenden Energiefunktion

$$E(a_i) = \int_R (\nabla I^T \left(\begin{pmatrix} 1 & x & y & x^2 & xy & y^2 \end{pmatrix} \begin{pmatrix} a_0 & a_6 \\ \vdots & \vdots \\ a_5 & a_{11} \end{pmatrix} \right) + I_t)^2 dx dy \quad (3-35)$$

in der die unbekannt Parameter a_i und damit der optische Fluß und dessen partielle Ableitungen bis zur zweiten Ordnung bezogen auf den Punkt \mathbf{x}_0 zur Repräsentation der Bewegung innerhalb der Region R geschätzt werden. Diese Repräsentationsform des optischen Flusses über die Parameter a_i ist kompakter als die dichten Verschiebungsvektorfelder der Variationsansätze.

Das optische Flußfeld eines ebenen Objektes ist als Sonderfall in (3-34) enthalten. Mit den Restriktionen

$$\frac{1}{2}u_{xx} = v_{xy}, \quad u_{yy} = v_{xx} = 0, \quad \frac{1}{2}u_{xy} = v_{yy} \quad (3-36)$$

für die zweiten Ableitungen des optischen Flusses geht (3-32) über in (3-33). Das bedeutet aber auch, daß die Annahme lokal ebener Oberflächenstücke bei der Regularisierung durch Einführen von (3-36) in den Stabilitätsterm der Variationsansätze berücksichtigt werden kann.

3.3.5 Konturorientierte Beobachtungen

Alternativ zum Beobachtungsmodell (3-19) benützen Waxman et al. (1987) Messungen des *normal flows* (3-21), wie sie auch von Hildreth (1984) zur Schätzung des konturorientierten optischen Flusses verwendet wurden. Liegen Meßwerte l_{\odot} des normal flows entlang des Konturverlaufs zusammen mit den (gemessenen) Normalenrichtungen \mathbf{n}_C vor (siehe Abschnitt 3.2), liefert jeder Konturpunkt eine Beobachtungsgleichung

$$l_{\odot} = \mathbf{n}_C^T \mathbf{u}(x, y)$$

zur Berechnung der 12 bzw. 8 Parameter (3-32, 3-33). Anstelle der Integration über die Region R ist ein Linienintegral entlang der Konturen C innerhalb der Region R auszuwerten. Diskret bedeutet das, daß wenn Meßwerte in mehr als 12 bzw. 8 Konturpunkten vorliegen, eine Kleinste-Quadrate-Schätzung der Flußparameter (a_i) berechnet wird. Waxman et al. (1987), Waxman und Wohn (1985) berichten, daß in der Praxis viele, möglichst mehrere hundert Konturpunkte einbezogen werden sollen. Die geschätzten Flußparameter repräsentieren den optischen Fluß innerhalb der betrachteten Region R , d. h. sie können auch als Merkmale dieser Region angesehen werden. Aufgrund der Kleinste-Quadrate-Minimierung sind die Merkmale einer (starken) Glättung unterzogen worden, so daß über die Beziehungen (3-32, 3-33) ein geglättetes optisches Flußfeld innerhalb des Definitionsbereiches, d. h. der Region R , berechnet wird.

Während bei Hildreth (1984) der optische Fluß nur entlang des Konturverlaufs bestimmt wird, leistet der von Waxman vorgestellte Lösungsweg eine flächenhafte Schätzung (die äquivalent durch die Flußparameter repräsentiert wird) aus konturorientierten Meßdaten. Daß insbesondere bei symmetrischen, wenig strukturierten Konturformen möglicherweise nicht alle Flußparameter schätzbar sind, wird als "Aperturproblem im Großen" bezeichnet.

3.3.6 Affinapproximation

Während die in Waxman angesprochenen “bis zu mehreren Hundert” Konturpunkte auf Bildausschnitte von vielleicht 50×50 Bildpunkte hindeuten¹³, sind für den Affinansatz (2-48)

$$\begin{aligned} u(x, y) &= a_0 + a_1x + a_2y \\ v(x, y) &= a_3 + a_4x + a_5y \end{aligned} \quad (3-37)$$

deutlich kleinere Fenstergrößen von z. B. 15×15 oder 20×20 Bildpunkten typisch. (3-37) kann als Approximation von (3-33) betrachtet werden, wobei die Perspektivprojektion lokal durch eine Parallelprojektion (Abschnitt 2.3.4) angenähert wurde. D. h. auch wenn das Objektmodell einer Ebene die Szene korrekt beschreibt, gilt (3-37) im Gegensatz zu (3-33) nur im lokalen Bereich. Der Affinansatz kann auch als Sonderfall der Flußgleichungen (3-32) bei gekrümmten Oberflächen betrachtet werden, in dem alle normalisierten Krümmungen identisch 0 sein müssen. Durch (3-37) wird die gekrümmte Fläche durch ihre Tangentialebene approximiert.

Während (3-35) noch teilweise in Konkurrenz zu Variationsansätzen steht, ist der affine Ansatz (3-37) eher typisch für Aufgabenstellungen wie die Verfolgung kleiner Objekte oder markanter Details dieser Objekte. Für die Schätzung des optischen Flusses innerhalb dieser kleinen Bildausschnitte werden fast ausschließlich die bildpunktbezogenen Beobachtungsmodelle verwendet. Mit der intensity-constancy assumption (3-19) und der Approximation der projizierten Bewegung durch (3-37) lautet die Minimierungsaufgabe ähnlich (3-35)

$$E(a_i) = \int_R (I_x(a_0 + a_1x + a_2y) + I_y(a_3 + a_4x + a_5y) + I_t)^2 dx dy \rightarrow \min . \quad (3-38)$$

Die geschätzten Parameter \hat{a}_i repräsentieren den optischen Fluß innerhalb des Bildausschnittes R ; sie stehen auch in relativ einfachen Beziehungen zu den Bewegungs- und Strukturparametern (siehe Abschnitt 2.3.4). Es liegt deshalb nahe, diese Informationen zur Segmentierung der Szene heranzuziehen (Adiv, 1985), wofür das Bild zunächst in viele kleine Auschnitte unterteilt wird. Die geschätzten optischen Flußparameter oder die daraus abgeleiteten Strukturparameter dieser lokalen Flächen können zur Analyse von Diskontinuität in der Szene herangezogen werden.

Negahdaripour und Lee (1992) greifen diese Idee auf, unterteilen das Bild in quadratische Regionen von 19×19 Bildelementen (wobei benachbarte Regionen mit 2 Pixel Breite geringfügig überlappen) und fügen Nachbarregionen zusammen, wenn die geschätzten Flußkoeffizienten \hat{a}_i einander ähnlich sind. In einem zweiten Schritt werden aus zwei möglichst großen Regionen mit verschiedenen Tangentialebenen die 12 geschätzten Flußparameter herangezogen, um über die Beziehungen (2-47) die 5 Bewegungs- und 2×3 Strukturparameter zu bestimmen. Mehr Details dieser Rekonstruktion sind in Abschnitt 2.3.4 diskutiert.

Wird anstelle der intensity-constancy assumption (3-19) das weniger restriktive Modell (3-23) verwendet, können in gewissem Umfang Intensitätsänderungen, z. B. aufgrund von Reflexionseigenschaften, Beleuchtungsänderungen, etc. kompensiert werden. Die radiometrische Verschiebungs- und Maßstabskorrektur hat sich vor allem im Bereich der Stereobildzuordnung bewährt. Die radiometrischen Kompensationsparameter (b_0, b_1) werden in der Kleinste-Quadrate-Minimierung gemeinsam mit den Flußparametern (a_i) mitgeschätzt, oder alternativ dazu vorab bestimmt und in der Berechnung des temporalen Gradienten I_t berücksichtigt, so daß im wesentlichen die Minimierungsaufgabe (3-38) zu lösen ist.

3.3.7 Lokale Fenster

Findet eine noch stärkere Einengung des lokal betrachteten Bereiches statt, bedeutet dies Fenstergrößen, wie sie auch bei lokalen Operatoren zur Extraktion von Punkten oder Kantenelementen üblich sind. Tatsächlich kann z. B. die Korrelation mit einer Fenstergröße von 5×5 Bildelementen in erster Linie als Meßoperator betrachtet werden, der bildelementbezogen eingesetzt hauptsächlich in Konkurrenz zu anderen pixelorientierten Beobachtungsmodellen tritt, wie z. B. (3-19, 3-21, 3-23).

¹³Die Größenangabe soll hier lediglich eine grobe Vorstellung über die Ausdehnung im Bild geben. Die Zahl der Konturpunkte ist selbstverständlich von Form und Texturierung des Objekts und vom Verfahren zur Lokalisierung der Konturpunkte abhängig. Der Ausschnitt muß in der Praxis selbstverständlich auch nicht quadratisch geformt sein.

Geometrisch betrachtet wird der optische Fluß innerhalb des lokalen Fensters als konstant angenommen

$$\begin{aligned} u(x, y) &= u_0 \\ v(x, y) &= v_0 . \end{aligned} \quad (3-39)$$

Bedingt durch die kleine Fenstergröße sind die durch Minimierung von

$$E(u_0, v_0) = \int_R (\nabla I^T \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + I_t)^2 dx dy \rightarrow \min$$

geschätzten Verschiebungen besonders texturabhängig und rauschanfällig. Die Ursprünge der Verschiebungsschätzung mit dem Modell (3-39) in Verbindung mit einer Minimierung nach dem Kleinste-Quadrate-Kriterium finden sich in Cafforio und Rocca (1976, 1979) und in Limb und Murphy (1975). Während Limb und Murphy den Quotienten aus zeitlichen und räumlichen Gradienten direkt nützen und zeilenorientiert die Verschiebungskomponente bestimmen, wird in Cafforio und Rocca exakt der Weg der Kleinste-Quadrate-Bewegungsschätzung aufgezeigt. Werden die polynomischen geometrischen Modelle als Generalisierung von (3-39) betrachtet, wird deutlich, daß diese frühen Arbeiten als Wurzel aller Kleinste-Quadrate-Bildzuordnungsverfahren auf ikonischer Datenebene zu betrachten sind.

Lokale Bereiche des Bildes, die eine präzise Schätzung beider Verschiebungskomponenten gestatten, setzen eine starke isotrope Textur voraus, wie sie an Grauwertecken im Bild vorkommt. Sind nicht beide Verschiebungskomponenten bestimmbar, z. B. entlang einer Kante (Aperturproblem), so kann anstelle von (3-39) auch die Messung l_\ominus orthogonal zum Konturverlauf treten. Aus der bildpunktbezogenen Formulierung (3-20, 3-21)

$$l_\ominus = -I_t / |\nabla I|$$

ergibt sich unter der Voraussetzung von lokal konstantem optischem Fluß (3-39) durch die Integration über den lokalen Ausschnitt und die Minimierung gemäß

$$\int_R (|\nabla I| l_\ominus + I_t)^2 dx dy \rightarrow \min$$

eine fensterbasierte Schätzung von l_\ominus .

Die verschiedenen Formen der lokalen Bewegungsschätzung lassen sich wiederum einbringen in die flächen- oder konturorientierten Glattheitsforderungen für das gesamte Bild, so daß sich hieraus eine Verallgemeinerung der im Abschnitt 3.2 vorgestellten Ansätze von Horn und Schunk (3-24) bzw. Hildreth (3-28) ergibt.

Wird die fensterbasierte Bewegungsschätzung in jedem Bildpunkt lokal ermittelt, impliziert das, daß durch die überlappenden Fenster Bildinformation mehrfach genützt wird. Die Mehrfachnutzung hat Anandan (1987a,b) populär gemacht, wobei er durch Korrelation bildpunktbezogen Verschiebungsvektoren ermittelt, mit globaler Glattheit integriert und dabei sowohl global als auch lokal eine sehr gute Schätzung des optischen Flusses erzielt.

3.3.8 Differentialgeometrische Modellierung der Bildfunktion

Mit der differentialgeometrischen Modellierung der Bildfunktion hat Nagel (1983) eine weitere Modellierungskomponente zur Schätzung von Verschiebungsvektoren in lokalen Bildausschnitten eingebracht. Die Grundidee dieses Ansatzes besteht darin, die Intensitätsstruktur des Bildes zu analysieren, um solche Bildausschnitte zu lokalisieren, die die Schätzung beider Komponenten (3-39) gestatten. Durch die lokale Approximation der Bildfunktion werden lokale Deskriptoren der Fenster punktbezogen bestimmt und zur Bewegungsschätzung herangezogen.

Nagel modelliert die Bildfunktion durch ein bivariates Polynom 2. Ordnung in den Bildkoordinaten (was in entsprechender Weise für den optischen Fluß in (3-34) dargestellt wurde) und leitet daraus einen Operator für die Lokalisierung von Grauwertecken ab. Wir werden darauf ausführlich in Kapitel 4 eingehen.

Die lokale Approximation der Bildfunktion $I(\mathbf{x})$ durch ein Polynom 2. Ordnung in $\mathbf{x} = (x, y)^T$ kann gemäß

$$f(x, y) = f_0 + f_x x + f_y y + \frac{1}{2} f_{xx} x^2 + f_{xy} xy + \frac{1}{2} f_{yy} y^2$$

bzw.

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f^T(\mathbf{x}_0) \mathbf{x} + \frac{1}{2} \mathbf{x}^T H_f(\mathbf{x}_0) \mathbf{x}$$

dargestellt werden (vgl. (3-34)). Bezugspunkt der Reihenentwicklung ist $\mathbf{x}_0 = 0$, d. h. es wurde ein lokales Koordinatensystem verwendet, das in der Regel in die Mitte eines symmetrisch gewählten Bildausschnittes gelegt wird. ∇f^T gibt die partiellen ersten Ableitungen, die Hesse-Matrix H_f (siehe auch Abschnitt 4.2.3) die symmetrische Matrix der zweiten Ableitungen

$$H_f = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}$$

an. Zur Schätzung der innerhalb des lokalen Fensters um den Bezugspunkt \mathbf{x}_0 konstant angenommenen Bewegung (3-39) argumentiert Nagel mit zwei diskreten Bezugszeitpunkten. Die Erläuterung des Verfahrens mit zwei Bildern ist tatsächlich anschaulicher, weshalb wir ebenfalls mit $I(t_1)$ und $I(t_2 = t_1 + \delta t)$ zwei aufeinanderfolgende Bilder einer Sequenz ansprechen wollen. Nagel entwirft ein Iterationsschema zur Schätzung des Verschiebungsvektors \mathbf{u}_0 ¹⁴. Zunächst werden die Polynomparameter $\mathbf{p} = (f_0, f_x, \dots, f_{yy})^T$ innerhalb des gewählten Fensters durch eine Kleinste-Quadrate-Schätzung aus

$$\int_R (I(\mathbf{x}, t_1) - f(\mathbf{x}, \mathbf{p}))^2 dx dy \rightarrow \min$$

bestimmt, wobei das Rauschen in den abgetasteten Intensitätswerten von $I(\mathbf{x}, t_1)$ als additives weißes Gauß'sches Rauschen angenommen wird. Über die Parameter $\hat{\mathbf{p}}$ können die geschätzten Funktionswerte $\hat{f}(x, y)$ ¹⁵ der polynomisch modellierten Bildfunktion direkt berechnet werden. Die zugehörigen Kovarianzmatrizen $C_{\hat{\mathbf{p}}}$ und $C_{\hat{f}}$ sind durch Fehlerfortpflanzung einfach bestimmbar. Unter Beachtung von Symmetrie und regelmäßiger Rasterung hat Nagel (1983) die Beziehungen zur direkten elementweisen Berechnung der Kovarianzmatrizen abgeleitet.

Zur iterativen Bestimmung der Verschiebung \mathbf{u}_0 mit dem "Suchbild" $\hat{f}(\mathbf{x} - \mathbf{u}_0)$ und dem "Referenzbild" $I(\mathbf{x}, t_2)$ ¹⁶ kann die Minimierungsaufgabe mit den Intensitätsdifferenzen

$$I_t = I(\mathbf{x}, t_2) - \hat{f}(\mathbf{x} - \mathbf{u}_0)$$

und der durch Fehlerfortpflanzung ermittelten Kovarianzmatrix C_{I_t}

$$C_{I_t} = \sigma^2 E + C_{\hat{f}}$$

gemäß

$$\mathbf{I}_t^T C_{I_t} \mathbf{I}_t \rightarrow \min$$

formuliert werden. Das Rauschen im abgetasteten Bild $I(\mathbf{x}, t_2)$ wurde wie schon bei $I(\mathbf{x}, t_1)$ als Gauß'scher weißer Rauschprozeß angenommen. \mathbf{I}_t ist ein Spaltenvektor, in dem die Differenzen I_t in den einzelnen Bildpunkten innerhalb des Fensters vektororientiert abgelegt sind. Die Linearisierung nach \mathbf{u}_0 führt zur iterativen Schätzprozedur

$$(\mathbf{u}_0)_{\text{Iteration } i} = (\mathbf{u}_0)_{\text{Iteration } i-1} + d\hat{\mathbf{u}}_0.$$

Durch Fehlerfortpflanzung kann die Kovarianzmatrix für den geschätzten Verschiebungsvektor $\hat{\mathbf{u}}_0$ berechnet werden. Nagel (1983) gibt auch hierfür eine explizite Lösung an, die im wesentlichen zur Matrix W führt, die zur Formulierung der orientierten Glattheit eingesetzt wurde. Das bedeutet, daß die 2×2 Kovarianzmatrix der

¹⁴Mit $\delta t \rightarrow 0$ kann formal das Verfahren in die Betrachtungsweise eines momentanen Bezugszeitpunktes überführt werden. Da mit der zeitdiskreten Betrachtung aber ein algorithmischer Lösungsweg aufgezeigt wird, wie von der differentiellen Argumentation auf diskrete Bezugszeitpunkte übergegangen werden kann, lohnt es den Nagel'schen Lösungsweg zu skizzieren. Deutlich wird dabei, daß durch die Diskretisierung der differentiellen Ansätze in der Regel die direkte Lösung in ein Iterationsverfahren übergeht.

¹⁵ \hat{f} ist ein Vektor, der die Schätzung der Funktionswerte in allen Bildelementen der betrachteten Region enthält.

¹⁶Der Suchradius ist durch die üblicherweise kleine Fenstergröße von \hat{f} beschränkt, wenn die Bestimmung bei größeren Verschiebung \mathbf{u}_0 nicht erneut durchgeführt werden soll. Vertauschen der Rolle von Suchbild und Referenzbild impliziert, daß $I(\mathbf{x} - \mathbf{u}_0, t_2)$ bei jeder Iteration resampled werden muß.

lokalen Schätzung direkt in die Gewichtung innerhalb des Regularisierungsterm eingeht. Grob gilt also, daß an Grauwertecken eine genau Schätzung der Verschiebung möglich ist und deshalb eine geringe Gewichtung auf die regularisierenden 1. Ableitungen des optischen Flusses ausgeübt wird. Umgekehrt haben ungenaue Schätzungen eine entsprechend stärkere, zudem richtungsabhängige Regularisierungswirkung. Eine ausführliche Diskussion erfolgte in Abschnitt 3.2. Damit ist auch die Verbindung zwischen der lokalen fensterorientierten Schätzung und dem globalen, für das ganze Bild gültigen Ansatz mit orientierter Glattheit (3-29) aufgezeigt.

3.3.9 Diskussion

Wie schon bei den Variationsansätzen im vorhergehenden Abschnitt wurde die Frage der Näherungswertgewinnung für die Bewegungsschätzung in örtlich begrenzten Bildausschnitten nicht direkt bei den Ansätzen mit besprochen. Der Standardweg besteht darin, eine Mehrebenenprozedur zu entwickeln. Üblicherweise findet die Grobschätzung der Bewegungsparameter auf den groben Auflösungsstufen der Bildpyramide statt und liefert die nötige Näherungsinformation für die Schätzung auf der nächst feineren Ebene der Pyramide. Der Prozeß wird bis zur höchsten Auslösungsstufe, d. h. dem erfaßten oder auch geringfügig geglätteten Bild fortgesetzt und ermöglicht in vielen Fällen eine zuverlässige Bildzuordnung selbst bei sehr kleinen Bildausschnitten. Weitere Aspekte pyramidaler Zuordnungsverfahren, insbesondere im Hinblick auf eine Alternative, den sog. Scale-Space, werden wir im Kapitel 5 vertiefen.

Für die differentiellen Bildzuordnungsverfahren, insbesondere wenn sie wie in (3-31) Bewegungs- und Strukturparameter direkt als zu schätzende Parameter einbeziehen, wird oft betont, daß Sie ohne Korrespondenz zurechtkommen. Dies ist deshalb so bedeutend, da allgemein das Korrespondenzproblem, d. h. die Aufgabenstellung Merkmale in beiden Bildern zu identifizieren, die Projektionen ein und derselben Objektstruktur in der 3D-Welt sind, als sehr schwierig gilt. Betrachten wir den optischen Fluß diskret in der Zeit, wie er bei der Realisierung auch geschätzt werden kann, gibt $\mathbf{u}(\mathbf{x})$ einen Verschiebungsvektor an (was bei Normierung durch das Zeitintervall auch als Verschiebungsvektorrates bezeichnet wird). Jeder Verschiebungsvektor definiert die Lage korrespondierender Bildpunkte (\mathbf{x}) und ($\mathbf{x} + \mathbf{u}(\mathbf{x})$) im Stereopartner, d. h. approximiert die Korrespondenz der elementarsten Merkmale des Bildes, nämlich der Bildelemente selbst (siehe auch Weng et al., 1992).

Eines der schwierigsten Probleme bei der Schätzung von Bewegung im Bild ist die Lokalisierung von Bewegungsdiskontinuitäten. Wird die Objektform lokal durch schwach gekrümmte oder ebene Flächen approximiert, sind die Objektränder und die Unstetigkeiten innerhalb der Randbeschreibung die natürlichen Grenzen der lokalen Ausdehnung, die in der Projektion die entsprechende lokale Begrenzung des Bildausschnittes festlegen. Schwierigkeiten bei der Bildzuordnung aufgrund der Okklusion von Objekten oder Objektteilen gehen mit der Diskontinuitätsproblematik einher, insbesondere bei größerer 3D-Bewegung (insbesondere Translation) zwischen zwei Bezugszeitpunkten.

Wie schon bei den globalen Ansätzen (Abschnitt 3.2) werden die konturorientierten Schätzverfahren als eine Möglichkeit betrachtet, um die Bewegungsdiskontinuitäten an den Rändern von meist geschlossenen Konturen in die Bewegungsschätzung einzubeziehen. Solange die Konturen mit Texturkanten auf einem glatten Oberflächenausschnitt korrespondieren, sind die Voraussetzungen für die Formapproximation gemäß (3-32, 3-33, 3-37) in der Regel gegeben. Gleiches gilt prinzipiell, wenn die Kontur mit einer Oberflächendiskontinuität korrespondiert, solange der von der Kontur eingeschlossene Ausschnitt der entsprechenden Flächenbeschreibung genügt. Ein Problem der konturorientierten Verfahren wird dabei offensichtlich: die starke Abhängigkeit von der Qualität der vorausgehenden Segmentierung des Bildes. Besitzt umgekehrt die Prozedur zur Bewegungsschätzung eine Möglichkeit, um aus den extrahierten Konturen diejenigen zu detektieren, die mit Bewegungsdiskontinuitäten korrespondieren, wird dadurch das Segmentierungsergebnis um eine weitere Interpretationskomponente ergänzt.

Die Verfahren zur Lokalisierung von Bewegungsdiskontinuitäten haben bislang ein Experimentierstadium kaum überschritten. Es gibt zwar klare Vorstellungen über die grundsätzliche gegenseitige Abhängigkeit von Bewegungssegmentierung im Bild und Szenensegmentierung (Diskontinuitäten des optischen Flusses und dessen erster Ableitung korrespondieren mit C^0 - und C^1 -Diskontinuitäten der 3D-Oberfläche, Okklusionskonturen und -regionen, die eine Untermenge in den Diskontinuitäten des optischen Flusses bilden, korrespondieren mit entsprechenden 3D-Konturen und verdeckten Flächen (Waxman et al., 1987 - siehe auch Abschnitt 3.1), aber wenig Konsens, dieses "Huhn-oder-Ei-zuerst"-Problem (Negahdaripour und Lee, 1992) anzupacken. Wäre eine

der Aufgaben (z. B. die optische Flußmessung) präzise gelöst, würde dies die jeweils andere Aufgabe (z. B. die Szenensegmentierung) wesentlich erleichtern. Wir werden die Komplexität des Problems in Abschnitt 3.5 noch näher betrachten.

Anstatt die Vorschläge zur Segmentierung (z. B. Waxman und Wohn, 1985, Adiv, 1985, Hoff und Ahuja, 1986, Waxman et al., 1987, Meygret und Thonnat, 1990, Negahdaripour und Lee, 1992) einzeln zu betrachten, sollen mehr grundsätzlich die wesentlichen Gesichtspunkte herausgearbeitet werden:

1. Meist werden die Residuen (Quadratsumme) der Meßdaten nach einer Kleinste-Quadrate-Schätzung getestet, wobei Inkonsistenz mit dem Modell, z. B. einer schwach gekrümmten Oberfläche, die Messung und das Modell in Frage stellen.
2. Um möglichst große zusammenhängende Regionen zu finden, werden kleine, lokal benachbarte Ausschnitte zusammengruppiert, wenn die geschätzten Flußparameter einander ähnlich sind. Ebenso können die Flußparameter in Strukturparameter (lokale Oberflächenneigung, etc.) transformiert und diese auf Ähnlichkeit geprüft werden.
3. Die Vergrößerung der Fenstergröße für ein gewähltes geometrisches Modell über bildpunktorientierte Regionenwachstumsverfahren oder entsprechend dem zuvor angedeuteten Gruppierungsverfahren geht einher mit der Suche nach der bestmöglichen Modellbeschreibung, wobei vor allem (3-32, 3-33, 3-37) konkurrieren.
4. Die Beziehungen (3-34) zwischen den lokal geschätzten optischen Flußparametern und den Ableitungen des optischen Flusses lassen theoretisch Vorinformation für die globalen Regularisierungsansätze aus lokalen Schätzwerten gewinnen. Diese Vorinformation könnte entsprechend dem von Terzopoulos präsentierten Vorschlag (3-4) direkt einbezogen werden.
5. Daß die Segmentierung mit der 3D-Interpretation einhergeht, zeigt sich beispielsweise bei den konturorientierten Verfahren, mit der Interpretation
 - Texturkante, wenn die Kontur innerhalb einer glatten Region liegt,
 - Diskontinuität, wenn der optische Fluß bzw. dessen erste Ableitung zwischen zwei Regionen an der gemeinsamen Konturlinie entsprechend stark differiert.

In der Regel werden auch Regionen bleiben, die zu rauh sind, um durch eine glatte Fläche beschrieben werden zu können. Ebenso Regionen, die zu wenig Textur aufweisen, um eine aussagekräftige Beurteilung zu ermöglichen. Auch Okklusionsgebiete zählen zu den schwer detektierbaren Regionen, zumal der entsprechende Ausschnitt der Szene nur in einem Bild abgebildet ist.

Ein letzter Aspekt, den wir diskutieren wollen, hat vor allem für die Punktverfolgung große Bedeutung. Verfahren, wie die Bewegungsschätzung aus langen Sequenzen (Teil III) profitieren algorithmisch von wenigen ausgewählten Punkten, an denen die Verschiebung sehr präzise in beiden Komponenten bestimmbar ist. Nagel (1983) hat nachgewiesen, daß an Grauwertecken die lokale punktbezogene Information (1. und 2. Ableitung der Bildfunktion) hierfür ausreicht. Mit wachsender Fenstergröße (Nagel extrahiert die Eckpunkte aus einer diskreten 3×3 Umgebung) und der Einbeziehung von Objektformapproximation wird eine mögliche Unbestimmbarkeit als "Aperturproblem im Großen" bezeichnet. Bildausschnitte, die eine sehr gute Bestimmbarkeit ermöglichen (allgemeinere Zusammenhänge werden im Kapitel 4 diskutiert), sollten zur lokalen Schätzung des optischen Flusses entsprechend ausgewählt werden. Die lokale Bestimmung der optischen Flußparameter kann durch ein Verschiebungsvektorfeld der Region dargestellt werden, d. h. im Prinzip ist es möglich, jeden beliebigen Punkt der Region weiterzuverfolgen. Üblicherweise wird die Fenstermitte oder der mit den Gradientenquadrate gewichtete Schwerpunkt als Bezugspunkt definiert. Dieser Ablauf steht in umgekehrter Reihenfolge zu den merkmalsbasierten Zuordnungsverfahren. Dort werden in jedem der Bilder unabhängig Punktmerkmale detektiert und lokalisiert. Nach erfolgreicher Zuordnung gibt die Positionsdifferenz den Verschiebungsvektor zwischen zwei Bezugszeitpunkten an. In den Kapiteln 4 und 5 werden wir an Experimenten die Vor- und Nachteile beider Verfahren aufzeigen.

3.4 Ein Vergleich von Variations- und Kleinste-Quadrate-Formulierungen

In den beiden vorausgehenden Abschnitten lag der Schwerpunkt darin, die Vielfalt der Verfahren zur Bestimmung der Bewegung im Bild unter Einbeziehung verschiedenster Modelle aufzuzeigen. Die Formulierungen der Optimierungs- oder Minimierungsaufgaben können abhängig von der zu bestimmenden Größe in zwei Gruppen unterteilt werden: Die eine Gruppe zielt auf die Bestimmung einer unbekanntes *Funktion*, z. B. des optischen Flusses $\mathbf{u}(x, y, t)$ oder eines Höhenmodells $Z(X, Y)$, die zweite Gruppe auf die Schätzung von *Parametern*. Zur ersten Gruppe haben wir die Grundlagen und Lösungswege in den Abschnitten 3.1 und 3.2 besprochen. Der Übergang von der kontinuierlichen Formulierung der Optimierungsaufgabe (Zielfunktion, Euler'sche Gleichungen) zur diskreten Approximation der Euler'schen Gleichungen ist ein für die algorithmische Lösung der Aufgaben entscheidender Schritt, den wir am Beispiel des von Horn und Schunk (1981) präsentierten Relaxationsalgorithmus sehr ausführlich dargelegt haben. Im Gegensatz dazu beinhalten die in Abschnitt 3.3 vorgestellten Ansätze die Schätzung von Parametern und werden vorwiegend direkt in die Standardform Kleinsten-Quadrate-Ansätze gebracht. Wird die Integration über die betrachtete Region durch eine Summation über alle Meßdatenpositionen (z. B. alle Bildelemente oder alle Bildelemente entlang einer Kontur) ersetzt, ist der Übergang zur datendiskretisierten Darstellung der Zielfunktion vollzogen. Im Gegensatz zu den Variationsansätzen wird als zu bestimmende Größe nicht die Funktion \mathbf{u} , sondern werden die optischen Flußparameter eingeführt, die aufgrund von (3-34) auch durch den Verschiebungsvektor \mathbf{u}_0 und die partiellen Ableitungen $(\mathbf{u}_x, \dots, \mathbf{u}_{yy})$ interpretiert werden können.

3.4.1 Bildzuordnung in der Photogrammetrie

Die wesentlichen Arbeiten zur Bildzuordnung in der Photogrammetrie lassen sich alle der Gruppe der Parameterschätzverfahren zuordnen. Obwohl sie vorwiegend zeitdiskrete Stereo- oder Multibildkonfigurationen¹⁷ eines Bildverbandes zugrunde legen, lassen sie sich im Kontext der vorgestellten Bildfolgenansätze diskutieren und einordnen. Dies wollen wir im folgenden in knapper Form tun. Einer Gruppe dieser Arbeiten liegt ein Konzept zugrunde, das in enger Verbindung zu den Variationsansätzen zur Schätzung des optischen Flusses steht. Die Diskussion der Gemeinsamkeiten und Unterschiede gibt ein tieferes Verständnis für die Querbeziehungen zwischen den Ansätzen und Lösungswegen der sich in der mathematischen Formulierung grundlegend unterscheidenden Arbeiten.

Sieht man von einigen wenigen merkmalsbasierten Arbeiten ab (z. B. Förstner, 1986, Schenk et al., 1990, Müller und Hahn, 1992), nützen die Ansätze zur Bildzuordnung in der Photogrammetrie hauptsächlich die ikonische Bilddatenrepräsentation. Im wesentlichen gehen die Intensitäten oder Intensitätsdifferenzen, die ein projizierter Punkt der Szene in allen beteiligten Bildern hervorruft, in die Zielfunktion ein. Das Äquivalent zur Intensitätsdifferenz zweier Bilder bildet der diskretisierte temporale Bildgradient.

Die Betrachtungen von Förstner (1982) zur geometrischen Genauigkeit der Bildzuordnung haben viele Gemeinsamkeiten mit Limb und Murphy (1975), die ebenfalls zwei konstant angenommene Verschiebungsparameter innerhalb eines Bildausschnittes schätzen. Der Affinansatz (3-37) zur Modellierung der geometrischen Transformation zwischen Bildausschnitten wurde von Ackermann (1984) in Kombination mit dem radiometrischen Modell (3-23) präsentiert. Die experimentellen Erfolge (z. B. Pertl, 1984) dieser "Hochgenauen Kleinste-Quadrate-Bildkorrelation" haben der digitalen Verarbeitung der Bilder in der Photogrammetrie starke Impulse verliehen. Grün (1985a) greift den Vorschlag von Ackermann auf und setzt darüber hinaus voraus, daß die relative oder äußere Orientierung bekannt ist. Diese Vorinformation führt er in Form einer Schnittbedingung (Kollinearität) in das Modell ein. Die Objektkoordinaten des Schnittpunktes und dessen projizierte Bildkoordinaten (ein Punkt innerhalb des affin zugeordneten Bildausschnittes) werden simultan bestimmt. Die Erweiterung dieses Konzepts auf mehrere Bilder und eine sehr detaillierte Beschreibung vieler, für die algorithmische Lösung wichtiger Aspekte (z. B. die Gewinnung von Näherungswerten) der als "Multi Photo Geometrically Constraint matching" bezeichneten Prozedur sind in Baltsavias (1991) dargestellt. Der klassische Weg, die bekannte Orientierung eines

¹⁷Mit dem Begriff des Stereobildpaares werden implizit zum Teil epipolare Bildgeometrie zugrundegelegt, vor allem bei der Stereobildfolge. Oft werden auch die nicht epipolar rektifizierten Bilder damit gemeint, deren relative oder äußere Orientierung gegeben ist. Wir wollen hier unter Stereobildern Bildpaare einer Szene verstehen, bei denen lediglich die innere Orientierung explizit bekannt ist, die in der Regel durch Kalibrierung bestimmt wurde. Obwohl das "trinocular stereo" eine Generalisierung des Stereobegriffes über mehr als zwei Bilder impliziert, wird bei mehreren Bildern oft auch explizit von Multibildzuordnung gesprochen.

Bildpaares einzubeziehen besteht darin, die Bilder auf Normalbilder zu transformieren (Kreiling, 1976). Sehr umfangreiche Untersuchungen zur Objektpunktkoordinierung und Erfassung von Geländemodellen mit Normalbildern haben den Nachweis des hohen Qualitätspotentials (1/10 000 der Flughöhe) bei der automatischen Messung mit Bildzuordnungsverfahren erbracht (Hahn und Förstner, 1988, Hahn, 1989). Sie waren eine entscheidende Grundlage für die Entwicklung des Programmsystems MATCH-T (Ackermann und Krzystek, 1991), das heute bislang als eines der wenigen bildzuordnungsbasierten Produkte in der photogrammetrischen Praxis zur automatischen Oberflächenrekonstruktion aus digitalen Luftbildern eingesetzt wird. Zur Bildzuordnung (intensitätsbasiert und merkmalsbasiert) wurde in diesen Untersuchungen das affine Modell verwendet, das um die Epipolarrestriktion erweitert, d. h. mit 3 unabhängigen geometrischen Parametern formuliert wurde. Aus den resultierenden Parallaxen lassen sich mit der bekannten Orientierung 3D-Objektpunktkoordinaten ermitteln, die wiederum die primäre Information für die Interpolation bzw. Approximation der Geländeform durch Oberflächenmodelle darstellen. Auf die Diskussion verschiedener Aspekte der Regularisierung und finite Elemente Modellierung auch bezüglich der Erfahrungswerte aufgrund dieser Untersuchungen zur Geländemodellierung sei hier verwiesen (siehe Abschnitt 3.1).

Die Bewegungsschätzung unter der Annahme von Starrheit (3-31) hat mit den 3D-Bewegungs- und Strukturparametern in direkter Weise zu bestimmende Größen des Objektraumes in den Ansatz einbezogen. Konzepte mit ähnlicher Intention, die zudem hinsichtlich der Mehrbildzuordnung allgemeiner formuliert wurden, sind von Helava (1988) unter dem Begriff "Object-Space Least-Squares Correlation", von Wrobel (1987) mit "Facets Stereo Vision" und von Ebner et al. (1987) präsentiert worden. Gemeinsam ist all diesen Arbeiten, daß sie ein ortsfestes Referenzsystem in der stationären Szene festlegen und das Objekt durch zwei Flächen modellieren, nämlich eine geometrische und eine intensitätsbezogene Oberfläche. Die beiden Flächen werden durch eine finite Elemente Modellierung approximiert, wobei üblicherweise eine regelmäßige, meist quadratische Rasterung zugrundegelegt wird. Die Höhen bzw. Intensitätswerte in den Knotenpunkten sind zu bestimmen. Zwischen den Knotenpunkten wird die Formbeschreibung durch einen entsprechenden Interpolator festgelegt, wobei vorwiegend ein bilineares Interpolationsmodell verwendet wird. Die Rasterung für die Intensitätsoberfläche orientiert sich am Bildelement, d. h. der Rasterabstand wird grob durch die Größe eines Pixels in der Bildebene multipliziert mit dem Maßstabsfaktor der Aufnahme abgeschätzt. Da die geometrische Oberfläche meist viel glatter ist als die Intensitätsoberfläche, wird hier eine entsprechend gröbere Rasterung gewählt. Experimente und Details der algorithmischen Realisierung dieser Ansätze sind in Heipke (1990), Zheng (1991) und Weisensee (1991) dargestellt.

Diese auf die Modellierung im Objektkoordinatensystem konzentrierten Ansätze lassen sich alle der Gruppe der monolithischen Algorithmen zuordnen (siehe Abschnitt 2.4.1). Die beobachteten Daten sind die in der Bildebene registrierten Intensitätswerte. Die Modellbildung beinhaltet die Hypothese einer stationären Szene, deren Geometrie durch eine einfache, parametrische Oberflächenform beschrieben werden kann. Dieser geometrischen Formbeschreibung wird zusätzlich eine intensitätsbezogene Oberfläche überlagert, die auch als Orthophoto interpretiert werden kann. Die geometrische Beziehung zwischen Bild- und Objektraum ist durch die Perspektive gegeben, wobei meist vorausgesetzt wird, daß innere und äußere Orientierung der Bilder bekannt sind. Erfolgreiche Experimente mit diesen Ansätzen sind bislang vor allem mit synthetischem Bildmaterial in Simulationen präsentiert worden. Algorithmische Schwierigkeiten ergeben sich aus der Regularisierung (Wrobel, 1989, Franek und Müller, 1990) und der Rasterung der geometrischen und intensitätsbezogenen Oberflächen (dichte Raster erhöhen beträchtlich die Zahl der zu schätzenden Parameter, wobei die Zahl der Iterationen bis zur Konvergenz der Prozedur mit ansteigt (Heipke, 1990)). Da die monolithischen Ansätze hauptsächlich aufgrund der inversen Perspektive (2D-Intensitäten sind beobachtet, 3D-Formparameter zu rekonstruieren) zu den sehr schlecht gestellten Aufgaben zählen (Ackermann und Zheng, 1990), sind neben Stabilisierungsfunktionalen vor allem auch stabilisierende Prozeduren sehr wichtig. Hierfür werden generell hierarchische Vorgehensweisen auf der Basis von Bildpyramiden und Mehrgitterrepräsentationen (vor allem für die Höhenmodelle) verwendet, wie sie in Abbildung 3.3 dargestellt sind. Eine sehr große Schwierigkeit stellt vor allem die physikalische Modellierung dar. Die Beziehung zwischen der optischen Dichte an der Objekt Oberfläche und den registrierten Intensitäten in allen beteiligten Bildern sind sehr komplex und beinhalten Beleuchtungs- und Reflexionsmodelle mit sehr vielen Unbekannten (Wrobel, 1988, Weisensee, 1991). Hier sind starke Vereinfachungen üblich (Lambert'sche Oberfläche: perfekt diffuse Reflexion, ruhende weit entfernte Lichtquelle, etc.), so daß in der Praxis die Modellierung in der Regel durch eine radiometrische Korrektur zwischen Intensitätsbild und optischer Dichte

des Objekts entsprechend (3-23) lokal approximiert wird. Obwohl auch viele weitere Probleme, wie z. B. die automatische Lokalisierung von Okklusionen und Diskontinuitäten durchaus klar erkannt sind (z. B. Wrobel, 1989), stehen die Entwicklungen diesbezüglich bislang noch am Anfang. Einige Experimente hierzu wurden von Zheng und Hahn (1990) präsentiert, die zur Detektion von Diskontinuitäts- und Okklusionsregionen die robuste Parameterschätzung eingesetzt haben. Die gesamte prozedurale Entwicklung dieser aus theoretischer und praktischer Sicht sehr interessanten monolithischen Ansätze wird sicherlich noch einige Zeit in Anspruch nehmen. Möglicherweise führt der Weg zum Erfolg auch zurück zur Zerlegung des monolithischen Ansatzes in weitgehend unabhängige Komponenten (Bildzuordnung, -segmentierung, 3D-Rekonstruktion, -Segmentierung, etc.), die dann durch Integration sehr viel flexibler und robuster eine - den monolithischen Formulierungen überlegene - Gesamtlösung herbeiführen (Aloimonos, 1989).

3.4.2 Variations- versus Kleinste-Quadrate-Formulierung: Ein Beispiel

In den bislang besprochenen Arbeiten der photogrammetrischen Gruppe haben wir eine Entwicklungsrichtung ausgelassen, die in der Aufgabenstellung in enger Verbindung zu den Variationsansätzen zur Schätzung des optischen Flusses stehen. Rosenholm (1986), Rauhala (1986) und Li (1989) haben ein Konzept vorgestellt, das auch als "Multipoint Matching" Prozedur bezeichnet wird. Die Prozedur setzt ein epipolar entzerrtes Bildpaar voraus. Zu bestimmen ist ein Feld von Parallaxen, d. h. die diskrete Version einer skalarwertigen Parallaxenfunktion. Die Bildebene wird hierzu aufgerastert, wofür z. B. Rosenholm (1986) eine quadratische Rasterung mit einem Knotenabstand von 8 Bildelementen bei seinen Experimenten zugrundegelegt hat. Zur Regularisierung werden vorwiegend die zweiten Ableitungen der Parallaxenfunktion in den Knotenpunkten verwendet. Innerhalb jeder Rastermasche wird ein Bilinearinterpolator als Approximationsmodell für die Flächenbeschreibung verwendet. Radiometrisch werden zwischen den beiden Bildern die zwei Parameter - offset und gain - gemäß (3-23) zur Korrektur verwendet. Eine Erweiterung auf 6 radiometrische Parameter mit positionsabhängigen Koeffizienten hat Rosenholm (1988) vorgeschlagen.

Zum Vergleich haben wir in Tabelle 3.1 den Multipoint-Ansatz von Rosenholm (1986) der Formulierung des optischen Flusses von Horn und Schunk (1981) gegenübergestellt. Der Vergleich eines lokalen intensitätsbasierten Ansatzes mit der globalen Bestimmung des optischen Flusses ist in Hahn und Proß (1981) aufgearbeitet. Signifikant unterschiedlich ist zunächst vor allem die Formulierung der Ansätze in kontinuierlicher bzw. diskreter Form. Die Intensitätsfunktion, der optische Fluß und die Ableitungsfunktionen entsprechen bei diskreter Formulierung den Intensitätswerten in den Bildpunkten, einem Verschiebungsvektorfeld und den diskret formulierten Gradienten.

Die Intensitätsdifferenzen $I_2(x, y) - I_1(x + p, y)$ haben gemäß (3-18, 3-19) ein entsprechendes Gegenüber in der nichtlinearen optischen Flußgleichung. Im Gegensatz zum optischen Fluß $(u, v)^T$ ist die Parallaxenfunktion (p) allerdings nur skalarwertig formuliert. Die Parallaxen in allen Bildpunkten sind beim Multipoint Matching über eine Bilinearfunktion auf relativ grob (z. B. 8×8 pixel) gerasterte Stützpunkte p^\ominus bezogen. Die zu bestimmende Größe bei der Variationsformulierung ist die vektorwertige Funktion $\mathbf{u}(x, y)$, während bei der Kleinste-Quadrate-Formulierung die Parameter p^\ominus , d. h. die Parallaxen in den grob gerasterten Gitterpunkten, ermittelt werden. Die Lösung bei kontinuierlicher Formulierung führt zu den Euler'schen Gleichungen. Für die Parameterschätzung wird mithilfe von Näherungswerten nach den Parametern linearisiert und iterativ über die Lösung der Normalgleichungen die optimale Schätzung erzielt. Da die digitalen Bilder grundsätzlich nur datendiskret vorliegen, muß eine Diskretisierung der Euler'schen Gleichungen erfolgen. Dabei wird bei der Schätzung des optischen Flusses in jedem Bildelement ein Verschiebungsvektor bestimmt, d. h. die diskretisierten Differentialgleichungen führen zu einem nichtredundanten Gleichungssystem. Bei Nb Bildelementen sind dabei $2Nb$ Gleichungen mit $2Nb$ Unbekannten zu lösen. Dem steht bei Multipoint Matching ein hochredundantes Gleichungssystem mit beträchtlich geringerer Anzahl Unbekannter (siehe Tabelle 3.1) gegenüber.

Während beim optischen Fluß durch eine isotrope Glättung mit den ersten Ableitungen überhaupt erst das Aperturproblem überwunden wird, d. h. die Bestimmbarkeit der zweidimensionalen Verschiebung ermöglicht wird, tritt dieses Problem bei Multipoint Matching durch die grobe Rasterung in wesentlich entschärfter Form auf. Dennoch hat Rosenholm seinen Ansatz damit motiviert, daß er Gebiete mit sehr schwacher Textur zu

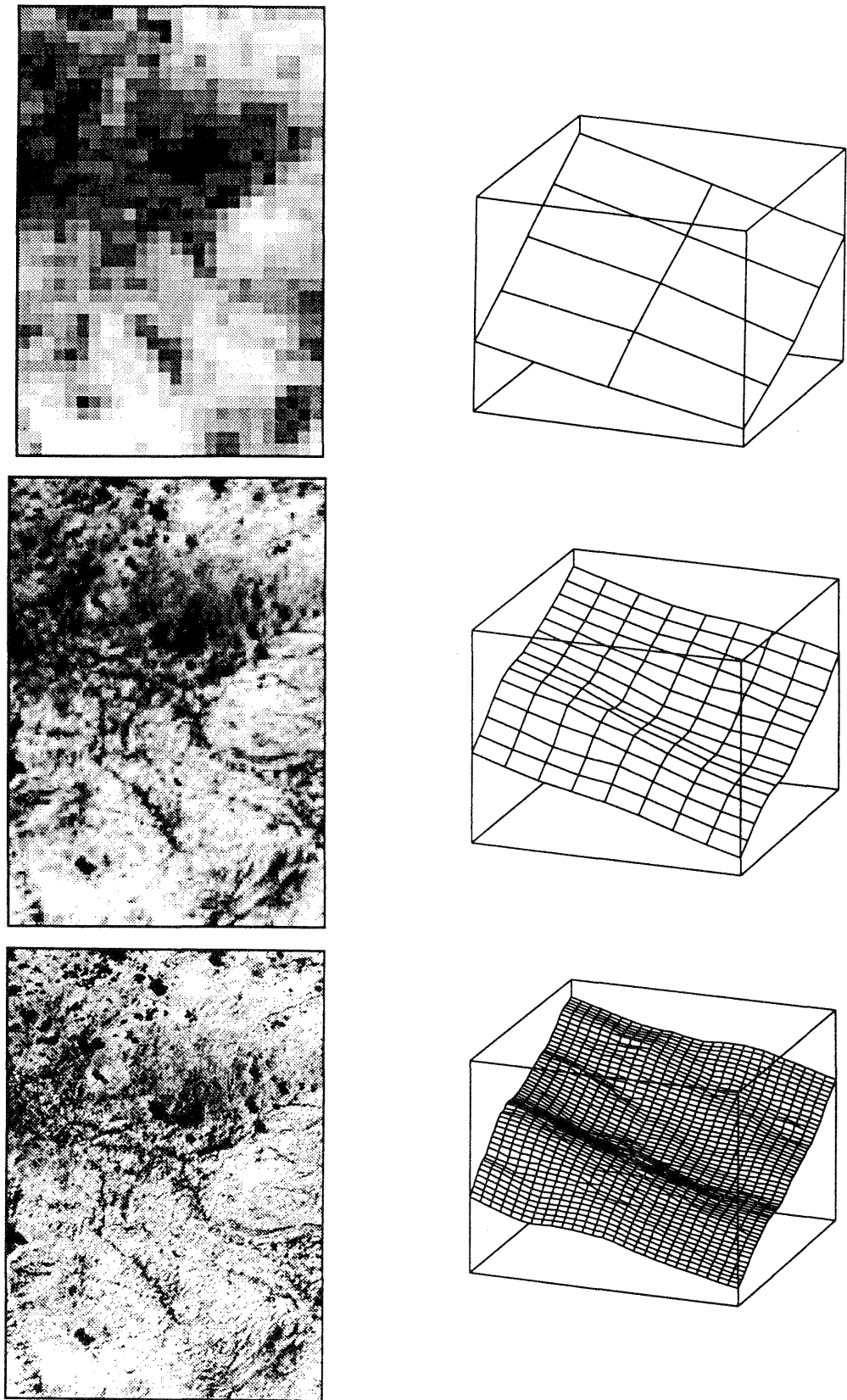


Abbildung 3.3: Bildpyramide und rekonstruiertes DHM (aus Ackermann und Hahn, 1991). Links: 3 Ebenen der Bildpyramide eines Stereobildpartners. Rechts: Rekonstruktionsresultate der entsprechenden Pyramidenebenen

| Variationsformulierung (Horn und Schunk, 1981) | Kleinste-Quadrate-Formulierung (Rosenholm, 1986) |
|---|---|
| Ansatz | |
| kontinuierlich formuliert in Raum u. Zeit | diskret formuliert in Raum u. Zeit |
| Zielfunktion | |
| $E(\mathbf{u}) = \lambda \int_B (\nabla I^T \mathbf{u} + I_t)^2 dx dy + \int_B (\nabla u ^2 + \nabla v ^2) dx dy$ $\rightarrow \min$ | $E(p) = \sum_B (I_2(x, y) - I_1(x + p, y))^2 + \sum_B (p_x^{\odot 2} w_{1x} + p_y^{\odot 2} w_{1y} + p_{xx}^{\odot 2} w_{2x} + p_{yy}^{\odot 2} w_{2y})$ $\rightarrow \min$ |
| ∇I - Intensitätsgradienten $(I_x, I_y)^T$ I_t - Intensitätsgradient nach der Zeit (x, y) - Bildebene \mathbf{u} - optischer Fluß $\mathbf{u} = (u, v)^T$ $\nabla u, \nabla v$ - Gradienten des optischen Flusses (1. Ableitung nach x bzw. y) λ - Regularisierungsfaktor (glob. Gew.-faktor) | I_1, I_2 - Intensitäten eines Bildpaares (x, y) - Position der Bildelemente p - x -Parallaxen in (x, y) = Bilinearfunktion von p^{\odot} (4 Nachbarn) p^{\odot} - x -Parallaxen in Gitterpunkten $p_{x[y]}$ - diskrete 1. Ableitung nach x bzw. y $p_{xx[yy]}$ - diskrete 2. Ableitung nach x bzw. y $w_{1[2]}$ - Gewichte der 1. bzw. 2. Abl. von p^{\odot} |
| Schätzung von | |
| Variationsfunktion $\mathbf{u}(x, y)$ (optischer Fluß) | Parameter p^{\odot} (Parallaxen in Gitterpunkten) |
| Lösung | |
| Euler'sche Gleichungen, Differentialgleichungen: $\lambda \nabla^2 u = I_x^2 u + I_x I_y v + I_x I_t$ $\lambda \nabla^2 v = I_x I_y u + I_y^2 v + I_y I_t$ | Differentiation nach den Parametern, Normalgleichungen: $N \delta \hat{p}^{\odot} = A^T l$ (explizit nicht aufgeführt) |
| ∇^2 - Laplace-Operator | N, A - Normalgleichungs-, Designmatrix $\delta \hat{p}^{\odot}, l$ - Zuschläge zu Näherung p^{\odot} , Beob.-en |
| Diskretisierung | |
| $I_{x[y,t]}$ - diskrete Abl. in allen Bildelem. (x, y) \mathbf{u} - Verschiebungsvektorfeld (2D) | p^{\odot} - Parallaxenfeld (1D) |
| Lösungsweg, Anzahl zu schätzender Größen Nu | |
| Approx. des Laplace durch Differenzenoperator zur Vermeidung einer Inversion → Relaxationsalg.(3-27), lokale Berechnung Iterationen: sehr große Anzahl | Inversion der Normalgleichungsmatrix globale, direkte Berechnung Iterationen: große Anzahl |
| 2 Unbek. aus zwei Restriktionen pro pixel, bei Nb Bildelementen: $Nu = 2Nb$ | bei Ng Gitterpunkten, z.B. $Ng = 32^2$ bei $Nb = 256^2$, $Nu = Ng \ll Nb$ |

Tabelle 3.1: Vergleichende Gegenüberstellung der Variationsformulierung von Horn und Schunk (1981) und der Kleinste-Quadrate-Prozedur von Rosenholm (1986)

überbrücken, d. h. ein “Aperturproblem im Großen” zu überwinden sucht. Durch die Ableitungen der Parallaxen in den Stützstellen ist dies auch garantiert. Die individuelle Gewichtung der quadrierten ersten und zweiten Ableitungen ist potentiell wesentlich flexibler als die isotrope Gewichtung bei der Variationsformulierung von Horn und Schunk. Die eigentliche Schwierigkeit besteht aber darin, individuelle Gewichte automatisch zu gewinnen, wofür keine überzeugenden Vorschläge in diesen Arbeiten gemacht wurden. Eine beträchtliche Regularisierung stellt auch die Bilinearapproximation der Parallaxenfunktion dar, die letztendlich bewirkt, daß bei 8×8 Bildpunkten pro Rastermasche eine entsprechend starke Glättung erzielt wird.

Beide Ansätze berichten über sehr viele notwendige Iterationen bis zur Konvergenz der Prozedur. Verantwortlich hierfür sind zum einen die Linearisierung der nichtlinearen Schätzaufgabe, zum andern die Lösung des sehr großen dünn besetzten Gleichungssystems durch einen Relaxationsansatz, wie er von Horn und Schunk ausgearbeitet wurde. Deutliche Vorteile beim Lösungsweg von Horn und Schunk sind die simple Algorithmik, die nur lokale Berechnungen beinhaltet und deshalb eine Verarbeitung auf Parallelrechnerarchitekturen nahe legt. Die dichte Rasterung gestattet es, auch lokale Effekte sehr gut zu repräsentieren, hat aber den Nachteil, daß die lokale Redundanz und damit insbesondere die Zuverlässigkeit der Schätzung sehr viel geringer ist. Unter der Voraussetzung, daß das Interpolationsmodell für die Parallaxen innerhalb der vordefinierten Stützstellen als hinreichende Approximation der projizierten Objektfläche betrachtet werden kann, sollten deshalb die in den Knotenpunkten geschätzten Parallaxen eine höhere Genauigkeit aufweisen. Die Regularisierung nimmt beim Multipoint Matching, verglichen mit der dichten Rasterung beim optischen Fluß, eine eher untergeordnete Rolle ein. Dafür spricht auch der Vorschlag Rauhala’s (1988), die Gewichte für die Ableitungen von Iteration zu Iteration zu verringern, so daß sie anfangs dazu beitragen, daß eine Optimallösung gefunden wird und in der Lösung dann bei geringem Gewicht eine genauere Schätzung erzielt werden kann. Die Lokalisierung und Berücksichtigung von Diskontinuitäten auf dem gröberen Stützpunktraaster ist von Li (1989) ausgearbeitet worden. Der sehr viel einfachere Weg, Diskontinuität auf dem hochaufgelösten Pixelraaster direkt zu repräsentieren, erscheint allerdings wesentlich eleganter und bezüglich der Datenstruktur effektiver als der Vorschlag von Li.

Sowohl größere Konvergenzradien, als auch einen beträchtlichen Gewinn an Effizienz hat Li durch die Verwendung von Bildpyramiden in einer hierarchischen Prozedur mit Mehrgitterrepräsentation der Parallaxenstützstellen experimentell nachgewiesen. Die Erweiterung auf ein zweidimensionales Vektorfeld bei nichtepipolarer Geometrie wurde von Xiao et al. (1988) vorgeschlagen.

Aus heutiger Sicht erscheinen manche Begriffe für die in der Photogrammetrie ausgearbeiteten Zuordnungsverfahren nicht genügend spezifisch ausgewählt. Das *least squares matching* (Ackermann, 1984) wird, zumal nahezu alle Zuordnungsverfahren nach Kleinste-Quadrate-Kriterien formuliert sind, deshalb heute auch als *area based matching*, oder auch, da Intensitätswerte direkt einbezogen werden, als *intensity based matching* bezeichnet. Völlig verkehrt erscheint die Bezeichnung *single point matching* (Rosenholm, Li), da die geometrischen Transformationsparameter (affin oder quadratisch) gemäß Abschnitt 3.3 Flächenausschnitte der Bilder zuordnen und auch in der 3D-Welt als Strukturparameter entsprechender Oberflächen interpretierbar sind. Wird das flächenbasierte Zuordnungsergebnis auf ein punktbezogenes Ergebnis reduziert (Verschiebungsvektor), so beinhaltet dies eine gewisse Willkür in der Punktauswahl (z. B. der Fenstermitte oder des gradientengewichteten Schwerpunktes) innerhalb der Fläche und auch einen Verlust flächenspezifischer Zuordnungsinformationen. Anstelle der begrifflichen Verallgemeinerung auf *multipoint matching* wäre eine treffendere Bezeichnung die “Schätzung eines Parallaxenfeldes oder -modells” in grob gerasterten Stützstellen. Durch das zugrundegelegte Interpolationsmodell handelt es sich aber beim *multipoint matching* auch um eine flächenbasierte bzw. intensitätsbasierte Zuordnung, für eine in der Regel große Bildregion.

3.5 Bildflußparadigma

Daß der Themenbereich “Bildzuordnung” keineswegs eine abgeschlossene Theorie besitzt, bei dem die Methoden soweit entwickelt sind, daß sie zur Zuordnung z. B. bei Bildsequenzen nur noch eingesetzt werden müssen, ist aus den vorangegangenen Diskussionen sicherlich deutlich geworden. Ein sehr häufig postulierter Anspruch an den Meßprozeß der Bewegung in der Bildsequenz ist, daß er sich am Leistungspotential des menschlichen visuellen Systems orientieren soll. Dabei geht es nicht darum, eine Kopie des menschlichen visuellen Systems zu entwerfen. Hochberg (1987) drückt das so aus: *Machines should not see as people do, but must know how*

people see. Das Wissen um wie bzw. was Menschen sehen, liegt aber leider auch nicht als eine in sich geschlossene Theorie vor. Den bislang bemerkenswertesten Ansatz zu einer um eine *Computational Theory* erweiterte *Representational Theory of Mind* hat Marr (1982) ausgearbeitet. Das tragende Paradigma in seiner Arbeit ist die Transformation sensorischer Daten in symbolische Daten, wobei *Prozesse*, *Repräsentationen* und *Kontrollmechanismen* die Strukturelemente eines Bildinformation verarbeitenden Systems darstellen. In dieser Theorie können viele Prozesse weitgehend unabhängig voneinander betrachtet werden, ergänzen sich aber bei der Zusammenführung auf dem entsprechenden Repräsentationsniveau. Zu diesen Prozessen sind die verschiedenen Verfahren der Bildzuordnung und auch der Segmentierung zu zählen. Aus heutiger Sicht wird die Hypothese weitgehender Unabhängigkeit der Einzelprozesse zunehmend in Frage gestellt.

Daß der Mensch in der Lage ist, aus der stereoskopischen Betrachtung eines Zufallsmusters Form zu erkennen, komplementär aber auch ausschließlich konturhafte Bildinformation verarbeiten kann, um Form, Bewegung oder Deformation wahrzunehmen, haben wir bereits im ersten Kapitel diskutiert. Auch Parallelverarbeitung und Mehrebenenprozesse sind Versuche, der Funktionalität des menschlichen visuellen Systems näher zu kommen. Sowohl merkmals- als auch intensitätsbasiert konnten auf algorithmischer Basis Erfolge unter gewissen Voraussetzungen (z. B. einfacher Objektform) nachgewiesen werden. Eine eher ungelöste Aufgabe, für die sich möglicherweise durch Fusion merkmalsbasierter und intensitätsbasierter Verfahren neue Perspektiven auftun, liegt in Erfassung und Berücksichtigung von Bewegungsdiskontinuitäten des optischen Flusses. Ist eine gute Schätzung des optischen Flusses vorgegeben, können Diskontinuitäten durch Analyse des Bewegungsvektorfeldes potentiell einfacher detektiert werden. Umgekehrt ist die Segmentierung des Bildes bezüglich Diskontinuitäten des optischen Flusses eine sehr wichtige Vorinformation, um eine bessere Schätzung zu erzielen. Die Regularisierung bei bekannten Bewegungsdiskontinuitäten könnte ähnlich wie bei Nagel (1987) durch eine gerichtete Glattheitsforderung eingebracht werden. Diese entspricht dann einer Vektorgeneralisierung des von Terzopoulos (Gleichung (3-4)) vorgeschlagenen Konzepts zur kontrollierten Glattheit. Die Integration verschiedener Meßprozesse verbunden mit der Repräsentation der Meßergebnisse muß, wie schon eingangs von Teil II ausgeführt, daran ausgerichtet werden, in welcher Art und Weise das Meßresultat weiter verwendet werden soll, d. h. auch, welche Anwendung damit verfolgt wird.

Paradigmatische Arbeiten zur Bewegungsanalyse in Bildfolgen sollten deshalb zur konzeptionellen Klarheit hinsichtlich solcher Fragestellungen beitragen, die bei der Entwicklung geeigneter Vorgehensweisen, der Modellierungen und andern Formen des Einbringens problemspezifischen Wissens, sowie der Repräsentation aufgeworfen werden.

Die Fragestellungen um das Ringen nach geeigneten Strategien wollen wir an einem Beispiel verdeutlichen: *Ein autonom navigierendes Fahrzeug soll auf einer belebten Straße fahren*. Ziel des Beispiels ist es, die Komplexität bezüglich der Bildzuordnung in sehr groben Zügen zu verdeutlichen, die mit der Analyse einer solchen Sequenz einhergeht.

1. Die Bild-zu-Bild-Zuordnung wird als der erste Schritt angesehen, der den Meßprozeß für die Erfassung der projizierten Bewegung repräsentiert. Die in der Bildebene ermittelte Bewegung ist dabei eine direkte Interpretation der visuell sichtbaren, sich mit dem Parameter Zeit ändernden Intensitäten. Die Änderung der Bildintensitäten innerhalb eines Zeitintervalls kann selbstverständlich ohne eine entsprechende Modellierung zu keiner Messung der Bewegung in der Bildfolge führen.
2. Nehmen wir an, daß die Bewegung der Kamera (des navigierenden Fahrzeugs) bekannt ist, so liefern die Zuordnungsergebnisse bei einem modularen Aufbau des Analysekonzepts die primäre Information für die Rekonstruktion von Struktur- oder Formparametern über eine beobachtete starre Szene, einschließlich der sich in ihr bewegenden Objekte. Sind die Bewegungsparameter des Fahrzeugs nicht bekannt, so müssen diese, wie auch die Bewegungsparameter einzelner (starrer) Objekte als weitere Unbekannte mitrekonstruiert werden. Die Bewegungs- und Strukturparameter bilden dann eine nächste Interpretationsstufe, denn sie sind die direkte 3D-Interpretation der im Bildraum beobachteten Bewegung.
3. Die Aggregation der Strukturparameter aufgrund der Bewegungsparameter der einzelnen Objekte oder auch von Wissen über die Form der stationären Szene kann zumindest theoretisch zur 3D-Segmentierung der Szene in bewegte Objekte und stationären Hintergrund führen. Auch Tiefensprünge, Diskontinuitäten

der Oberflächenorientierung und Verdeckungen sind potentiell in einer solchen Aggregierungsphase identifizierbar und lokalisierbar. Daß gerade diese 3D-Interpretation aber eine enorm wichtige Wissenskomponente für die Bildzuordnung darstellt, ist anschaulich leicht nachvollziehbar. Die einzelnen, aggregierten und Objekten zugeordneten Strukturen korrespondieren mit entsprechenden Diskontinuitäten der projizierten Geschwindigkeiten. Bildzuordnung, die eine 3D-Bewegungssegmentierung und - was bei einem bewegten Beobachter ein ähnliches Problem aufwirft - eine Tiefensegmentierung (Bruchkanten, Tiefensprünge, Verdeckungen) potentiell gestatten soll, setzt einen Meßprozeß voraus, der auch die Diskontinuitäten des Bildflusses beinhaltet.

4. Die Einbeziehung expliziter, objektspezifischer Modelle kann schließlich zu Interpretationen auf sehr abstrakter Ebene führen, deren Resultat z. B. in einer Aussage wie "Auto überholt Fußgänger mit 50 km/h" bestehen könnte. Die Zuordnung der über die Bewegung segmentierten Objekte, z. B. mittels der rekonstruierten 3D-Strukturen, zu einem entsprechenden Objektmodell ist dabei eine eigenständige komplexe Aufgabe. Das Resultat der Zuordnung kann die Identifikation des Objekts Auto sein; dessen Geschwindigkeit ergibt sich aus der 3D-Bewegungsschätzung, und der Interpretationsterm "überholt" legt nahe, daß Modelle über Bewegungsabläufe in diesem Bewegungsraum notwendig sind, um eine solche Interpretation zu leisten und entsprechend zu artikulieren (Neumann und Novak, 1986). Ist Vorwissen über den Objektraum, z. B. in Form eines geometrischen Straßenmodells, und über die sich in dieser Szene bewegendenden Objekte, z. B. durch Modelle für Autos, Fußgänger, a priori vorhanden, so hat dies verschiedene Auswirkungen auf die Bildzuordnungsverfahren. Die Zuordnung von Modell und Bild beinhaltet zunächst das Problem, die Objekte zu erkennen. Identifizierte Objekte über die Sequenz hinweg zu lokalisieren, ist dann ein Prozeß der Objektverfolgung. Diese Zuordnungsproblematik von Objekten, die geometrisch durch Knoten, Kanten und Flächen spezifiziert sind, zu entsprechenden Größen im Bild hat sehr viele Gemeinsamkeiten zur Bild-zu-Bild-Zuordnung (Gennery, 1992, Lowe, 1991, 1992), vor allem in der langen Bildsequenz im Kontext der 3D-Rekonstruktion. Die Bild-zu-Bild-Zuordnung setzt die Existenz solcher Objektprimitive in 3D, z. B. bei Punkten 3D-Verknüpfungspunkte, voraus, auch wenn diese Primitive nicht explizit durch ein Objektmodell gegeben sind.

Eine Konsequenz aus dieser beispielhaften Skizzierung verschiedener Aspekte der Bildfolgenanalyse ist, daß die Bildzuordnung Interpretationsprozesse verschiedener Abstraktionsstufen anspricht und dabei ohne Approximation wohl nicht auskommen kann. Ein perfekter Meßprozeß der in den Bildern erkennbaren Bewegung beinhaltet entsprechende Interpretationsprozesse, wird dadurch äußerst komplex und wie das Beispiel zeigt, auch in starkem Maße durch Wissen und Modelle in und über die Szene beeinflusst. Dabei wurden in diesem Beispiel zeitabhängige physikalische Aspekte bezüglich Beleuchtung, Reflektion, etc. in die Diskussion noch gar nicht einbezogen.

Das von Waxman (1987) ausgearbeitete Bildflußparadigma (Abbildung 3.4) zielt auf eine Szenerie mit mehreren bewegten Objekten und Starrkörperbewegung. Es betont den konturorientierten Ansatz, ist deshalb bezüglich der originären Meßdaten einseitig. Besondere Beachtung verdienen die Einzelprozesse und ihre Verknüpfung, die Repräsentation und der Kontrollmechanismus des Analysekonzepts. Ein allgemeiner Anspruch an die Kontrolle ist, daß diese als problemunabhängiges prozedurales Wissen behandelt werden sollte, das auf verschiedenen Repräsentationsformen operiert (Sommer, 1992).

Die durch Kantendetektion in den aufeinanderfolgenden Bildern lokalisierten geschlossenen Konturen bilden die primäre Information zur Erfassung der Bewegung. Der lokal entlang der Konturen bestimmte Normalfluß (Abschnitt 3.2) wird herangezogen, um optische Flußparameter (Abschnitt 3.3) für die von den Konturen begrenzten kleinen Regionen zu ermitteln. D. h. die lokale Bewegung dieser Regionen wird durch die entsprechenden geschätzten Parameter repräsentiert. Diese Parameter beinhalten, da sie in analytischer Beziehung zur lokal approximierten Oberflächenform (gekrümmt oder eben) stehen, auch Information über die lokale 3D-Struktur. Die Abbildung auf dichte Flußfelder ermöglicht für die benachbarten Bildregionen einen bildelementbezogenen Vergleich der Verschiebungsvektoren in Überlappungszonen entlang der Konturen. Kompatibilität in dieser Zone dient als Kriterium für ein Zusammenfügen der lokal benachbarten Regionen. Die nachbarschaftlich zusammengewachsenen Regionen werden dann durch Flußparameterfelder repräsentiert. Dieser Wachstumsprozeß findet entlang der Konturen seine Grenzen, die inkompatible Flußinformation aufweisen. Das Resultat dieser

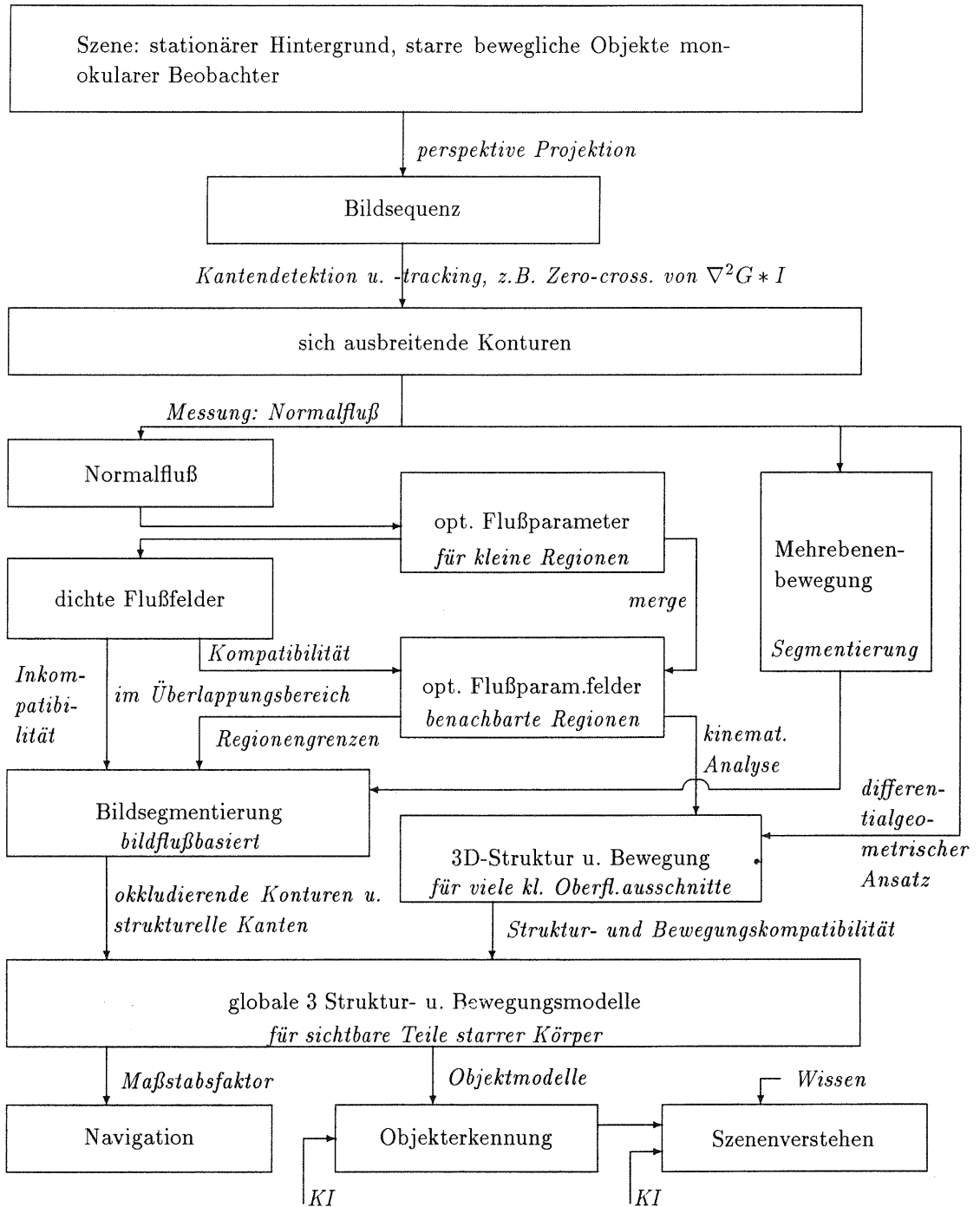


Abbildung 3.4: Elemente des Bildflußparadigmas (vereinfachte Fassung aus Waxman, 1987)

Analyse ist eine bildflußbasierte Segmentierung des Bildes in Regionen mit glatten optischen Flußfeldern. Die Regionengrenzen sind Diskontinuitäten des optischen Flusses innerhalb der gesamten Bildebene.

Eine Form der Kontrolle ist die Mehrebenenprozessierung des Bildflusses. Während die Bewegungsschätzung und Segmentierung über die Flußparameterfelder auf Einzelprozessen beruhen, die alle auf dem Bild direkt operieren, beinhalten die pyramidenbasierten Prozesse üblicherweise eine grob-zu-fein Kontrollstrategie, die in ähnlicher Konzeption für sehr viele verschiedene Aufgabenstellungen des Rechner Sehens Anwendung findet. Die auf grober Auflösungsstufe extrahierten Bewegungsparameter beinhalten eine sehr generalisierte Informa-

tion, die als Vorinformation für die Analyse auf der nächsten Auflösungsstufe direkte Verwendung findet. Aus der Verknüpfung zwischen den Pyramidenebenen kann auch Vorinformation für die Segmentierung gewonnen werden.

Der Transfer der Flußparameterfelder in den Objektraum wird in Abbildung 3.4 als kinematische Analyse bezeichnet. Die Objekte werden durch eine Vielzahl kleiner Oberflächenausschnitte angesprochen, für die die Flußparameter in 3D-Struktur- und Bewegungsparameter überführt werden. Die kinematische Analyse ist demzufolge der Interpretationsprozeß der bildbezogenen, regionenorientierten Parameter. Werden die sich ausbreitenden Konturen direkt verwendet und über eine differentialgeometrische Modellierung in den 3D-Raum transferiert, lassen sich im Prinzip 3D-Raumkurven gewinnen, die auf den Oberflächenausschnitten (Texturgrenzen) liegen oder diese begrenzen sollten (3D-Konturen). Strukturelle Kanten und okkludierende Konturen korrelieren potentiell mit den Diskontinuitäten des optischen Flusses. Die 3D-Analyse mit dem Zusammenfügen kleinerer Oberflächenausschnitte zu einer globaleren Modellbeschreibung der Form und Bewegung der starren Körper beinhaltet auch die Interpretation der Regionengrenzen (Bildflußsegmentierung) in Form von strukturellen Kanten oder okkludierenden Konturen.

Die unterste Zeile in Abbildung 3.4 deutet die Einbettung der erfaßten Struktur- und Bewegungsinformation in weitere Themenbereiche an. Während für Aufgaben der autonomen Navigation die essentiellen Informationen vorliegen, sind zur Lösung von Aufgaben der Objekterkennung zusätzliche Objektmodelle notwendig, die eine eindeutige Zuordnung der rekonstruierten Struktur- und eventuell auch Bewegungsinformation zu diesen Modellen gestatten. Das in der Bildfolgenanalyse lokalisierte und rekonstruierte Objekt "starrer Körper" bekommt damit eine explizitere Interpretation. Als komplexeste Aufgabe ist das Thema Szenenverstehen angedeutet. Einen Beitrag, den die Bildfolgenanalyse dazu leisten kann, haben wir bereits bei der Diskussion des oben angeführten Beispiels (*ein autonom navigierendes Fahrzeug ..*) unter Punkt 4. angeführt.

Diese relativ ausführliche Skizzierung einer paradigmatischen Arbeit muß als ein Plan verstanden werden, der zum Ausdruck bringen soll, welche wesentliche Bedeutung der Gesamtkonzeption, den Repräsentationsformen, den prozeduralen Einzelschritten und den Kontrollmechanismen zukommt. Eine entscheidende Schwäche haben wir bereits genannt: Der konturorientierte Ansatz ist natürlich in starkem Maße von der Qualität der lokalisierten Konturen abhängig, die bei den Zero-crossings eines LoG-gefilterten Bildes nicht sehr hoch ist. Eine Integration anderer lokaler Meßdaten (Abschnitt 3.2) ist aus heutiger Sicht ebenso geboten, wie die Ausweitung der Mehrebenenprozedur auf die Kantendetektion, z. B. entsprechend dem von Bergholm (1989) propagierten *edge focusing*.

4 Markante Punkte und Korrespondenz

Wissenschaftlicher Beitrag der Arbeit in diesem Kapitel:

Als Klassifizierungsschema für die Verfahren zur Extraktion markanter Punkte wird eine am Interpretationspotential des Verfahrens orientierte Unterteilung in statistische, differentialgeometrische und klassifizierende Verfahren vorgeschlagen. Die bekannten Punktfinder werden hierin eingeordnet und ihre gegenseitigen Beziehungen diskutiert. Zur Verfolgung der markanten Punkte werden zwei Verfahren vorgeschlagen, die konzeptionell sehr unterschiedlich sind. Eines der Verfahren nützt die rekursiv rekonstruierten 3D Punkte und zielt auf eine 2D-3D Korrespondenz. Als effizienter hat sich allerdings ein 2D-2D Korrespondenzverfahren herausgestellt, bei dem der Expansionspunkt zur Suchbereichsbeschränkung einbezogen ist. Zwei experimentelle Erkenntnisse lassen sich hervorheben: Obgleich mit dem Förstner'schen Punktfinder ein sehr qualifizierter Operator eingesetzt wurde, hat sich mehr als die Hälfte der markanten Punkte als wenig stabil in der Bildsequenz erwiesen. Dennoch verbleibt eine hinreichend große Zahl von Punkten (mehr als 100 im Mittel), die bei der merkmalsbasierten Punktverfolgung über 4 und mehr Bilder verfolgt wurden. Im Einzelfall konnte ein Punkt trotz beträchtlicher Maßstabsänderung (um einen Faktor 2.5) in 45 Bildern aufgefunden werden.

Neben den intensitätsbasierten Verfahren zur Schätzung des optischen Flusses haben die merkmalsbasierten Verfahren zur Bewegungsbestimmung im Bildraum große Beachtung bei der Bildfolgenanalyse gefunden. Die Betrachtung eines Bildausschnittes durch eine kleine Blende macht deutlich, daß der Intensitätswert eines Bildpunktes bzw. die Änderung der Bildintensität mit der Zeit nicht genügend Information für eine Bewegungsbestimmung enthält. Vergrößert man die Blende, d. h. betrachtet man einen lokalen Bildausschnitt, tritt die Strukturierung, d. h. die geometrische Verteilung und Organisation der Intensitäten in den Vordergrund. Diese Information explizit zu machen führt zur symbolischen Repräsentation des Bildes, deren Elemente als Primitive bezeichnet werden. Eine Möglichkeit zur Klassifizierung der Primitiven der symbolischen Bildrepräsentation ist die nach Punkten, Linien und Regionen. Sowohl spezifischer als auch allgemeiner hat Marr (1982) die Begriffe *Primitiv* oder *Token* verwendet. Er versteht darunter Elemente der symbolischen Bildrepräsentation wie Blobs, Zero-crossings, Diskontinuitäten, Kanten, virtuelle Linien, Gruppierungen u. a., die hierarchisch organisiert, im sog. *primal sketch* verschiedene Abstraktionsstufen beinhalten. Diese Primitive beziehen sich vor allem auf Bildintensitätsänderungen. Der grundlegende Anspruch an Primitive ist, daß sie zweckorientiert sind und vor allem mit realen physikalischen Änderungen der betrachteten Objektflächen korrespondieren (Marr, 1982).

Noch weiter gefaßt als der Begriff des Primitivs ist der Begriff des Merkmals. Zu Merkmalen zählen im einfachsten Fall die Intensitätswerte in den Bildelementen selbst. Aber auch Primitive, wie z. B. Eckpunkte, die durch einen Eckenoperator lokalisiert wurden und schon sehr stark auf eine 3D-Interpretation abzielen, werden teilweise als Merkmale bezeichnet. Merkmalsbasierte Bildzuordnungsverfahren sind dann alle Verfahren, die auf Elemente der symbolischen Beschreibungsebene der Bilder zurückgreifen. Charakteristisch für die merkmalsbasierte Bildzuordnung ist, daß die Primitive einander meist über ihre charakteristischen Größen - Merkmale oder Attribute - zugeordnet werden. Die ganz allgemeine Problemstellung, aus zwei Datensätzen korrespondierende Merkmale ausfindig zu machen, wird als Korrespondenzproblem bezeichnet.

Werden zur Lösung des Korrespondenzproblems zusätzlich Relationen zwischen den Primitiven hinzugezogen, spricht man zur Unterscheidung von der merkmalsbasierten Zuordnung von der relationalen Zuordnung. Relationale Zuordnung findet meist zwischen einer symbolischen Beschreibung eines Bildes und einer adäquaten Beschreibung von Objektmodellen mit dem Ziel der Objekterkennung und Objektlokalisierung statt. Im Rahmen der Bildfolgenanalyse ist die relationale Zuordnung für die Bewegungs- und Strukturrekonstruktion zur 3D-Szenenbeschreibung nur von untergeordneter Bedeutung. Eine mögliche Anwendung besteht darin, daß durch relationale Zuordnung Startwerte für die Orientierung der Kamera zu Beginn der Aufnahme einer längeren Sequenz gefunden werden, wenn es ein Ziel ist, daß die Szenenbeschreibung in einem vorab durch Paßinformation aus der 3D-Welt festgelegten geometrischen Bezugssystem erfolgen soll. Eine Übersicht über die relationale Bildzuordnung gibt Vosselman (1992).

Der Übergang von der ikonischen zur symbolischen Bildrepräsentation wird schon seit langem als ein sehr kritischer und entscheidender Schritt der Bildanalyse betrachtet. Die Extraktion von Primitiven findet oft in aufeinander aufbauenden Schritten statt. Meist werden zunächst Intensitätsänderungen im Bild lokalisiert und

daraus Primitive extrahiert, in dem die lokale Struktur analysiert wird. Weitere Analyseschritte können auf dieses Ergebnis aufbauen und die Primitive zu abstrakteren Primitiven zusammenfassen, woraus schließlich eine hierarchische Organisationsstruktur für die symbolische Bildrepräsentation folgt. Diese Prozesse werden auch als Gruppierung bezeichnet.

Obwohl im Grundsatz gilt, daß der Zuordnungsprozeß mit abstrakteren Primitiven durch die Einschränkung des Suchraumes bei höheren Repräsentationsebenen einfacher wird, beruhen die meisten merkmalsbasierten Bildzuordnungsverfahren auf simplen Primitiven wie Punkten oder Bildkanten. Ein Grund dafür ist, daß die Segmentierungs- und Gruppierungsprozesse zu abstrakteren Primitiven in der Theorie noch immer viel zu wenig ausgearbeitet sind. Für die Suchraumbeschränkung kann meist geometrisches Wissen über den Objektraum oder die Bewegung, das auf die Bildebene projiziert ist, sehr effektiv eingebracht werden. Die aufgabenspezifische, zweckorientierte Betrachtung, auch als *Purposive Vision* bezeichnet, kann heute als paradigmatische Antwort auf die Suche nach einer allgemeinen Theorie der Bildanalyse angesehen werden.

In den Experimenten zur merkmalsbasierten Bewegungsbestimmung aus Bildfolgen werden wir uns auf die Punktzuordnung bzw. -verfolgung und dabei auf den von Förstner entwickelten Operator beschränken. Trotzdem wollen wir eine Auswahl an Operatoren zur Extraktion markanter Punkte kurz charakterisieren, um den verwendeten Operator besser einordnen zu können. Im nächsten Abschnitt werden Kriterien an die Auswahl markanter Punkte formuliert, danach einige weitverbreitete Operatoren zur Lokalisierung markanter Punkte diskutiert. Experimente mit dem von Förstner vorgeschlagenen Interestoperator sollen einen experimentellen Zugang und einen Vergleich von Anspruch und Wirklichkeit bezüglich der Qualität des Operators bei der Punktverfolgung in Bildsequenzen darlegen.

4.1 Kriterien an die Extraktion markanter Punkte

Mathematisch ist der Punkt als Größe infinitesimaler Ausdehnung definiert, der durch seine Koordinaten beschrieben wird. Für den Bildpunkt, der synonym auch als Bildelement verstanden wird, gilt dies wegen der endlichen Ausdehnung des Bildelements nur noch in gewisser Näherung. Eine einfache Lösung wäre deshalb, unter einem Bildpunkt nur die Position eines Bildelementes anzusehen. Ein markanter Punkt im Bild hat mit dem geometrisch definierten Punkt gemeinsam, daß er positionsbezogen durch seine Koordinaten festgelegt wird. Darüber hinaus kommen aber noch eine Reihe von Charakteristika hinzu, für die wir in diesem Abschnitt Kriterien vorstellen wollen.

Als Kriterium an einen markanten Punkt aus Sicht der Bildzuordnung ist zumindest zu stellen, daß er die Bestimmung von Bewegung in der Bildfolge, d. h. die Ermittlung einer Änderung der Position im Bild, in jeder Richtung gestattet. Dazu soll der Punkt in seiner Lage im Bild eindeutig lokalisiert werden können.

Markante Punkte bezeichnen die Stellen im Bild, die auch bei visueller Betrachtung besonders auffällig sind. Zum Beispiel sind das helle oder dunkle Flecken, Endpunkte oder Schnittpunkte von Linien, oder auch Ecken. Die physikalischen Ursachen solcher markanter Punkte in der 3D-Welt liegen in der Texturierung der Oberfläche, in Objektkanten, in der Reflexion (z. B. bei Glanzlichtern), aber auch in virtuellen Schnittpunkten, z. B. bei sich verdeckenden Objektkonturen. Diese 3D-Interpretation kann und soll von Operatoren zur Extraktion markanter Punkte nicht geklärt werden. Das Ziel ist es aber, daß durch die 2D-Interpretation, z. B. als Grauwertecke, und die Lokalisierung eines Punktes die essentielle Information für eine höhere Stufe der Bildinterpretation gewonnen wird.

In der Formulierung der Kriterien für die Auswahl markanter Punkte folgen wir im wesentlichen auch begrifflich Förstner (1991). Die Kriterien lassen sich weiter unterscheiden nach Anforderungen an die markanten Punkte und Anforderungen an den Operator bzw. die Extraktionsprozedur (Bartsch, 1987), was Sinn macht, da zum einen die Repräsentation, zum anderen der Prozeß angesprochen ist, der von der ikonischen in die symbolische Ebene abbildet.

Im folgenden Kriterienkatalog sind die ersten drei Kriterien mehr der Repräsentation, die letzten drei mehr dem Auswerteprozess zuzuordnen.

Kriterienkatalog

- (1) *Deutlichkeit*: Als markant gilt ein Bildpunkt, wenn er sich deutlich von seiner Umgebung abhebt.

Dieses Kriterium kann als das primäre Definitionskriterium für den Begriff des markanten Punktes betrachtet werden. Markante Punkte sind Stellen im Bild, die im lokalen Bildkontext besonders auffällig sind. Sie sollen sich bezüglich bestimmter Eigenschaften signifikant von den benachbarten Bildpunkten unterscheiden. Zu diesen Eigenschaften zählt, daß die Lage des Punktes eindeutig lokalisiert werden kann.

- (2) *Seltenheit*: Ein markanter Punkt soll sich von allen selektierten Punkten des ganzen Bildes signifikant unterscheiden.

Die Seltenheit zielt auf die globale Einzigartigkeit eines markanten Punktes. Damit unterscheidet sich ein global markanter Punkt von anderen Merkmalen, insbesondere aber von seinesgleichen, womit er sich für eine einfache Zuordnung besonders empfiehlt.

Seltenheit legt implizit nahe, daß die Anzahl markanter Punkte klein, zumindest klein gegenüber der Anzahl der Bildpunkte sein wird. Für eine geringe Dichte der Punkte spricht eine geringere Komplexität der Bildzuordnung zu den Punkten des nächsten Bildes. Fordert die Anwendung allerdings eine reichhaltige Beschreibung, führt dies natürlich zum Konflikt mit diesem Kriterium.

- (3) *Interpretierbarkeit*: Ein markanter Punkt soll die Struktur im lokalen Bildausschnitt mit einer geometrischen Erklärung belegen können, d. h. er soll zu Bildstrukturen korrespondieren, denen eine Bedeutung im Sinne der Interpretation des Bildes beigemessen werden kann.

Die Interpretation der lokalen Bildstruktur führt zu einer begrifflich spezifischeren Ansprache eines markanten Punktes, z. B. als Eckpunkt oder als Kreismittelpunkt. Grauwertecken oder -kreise sind geometrische 2D-Primitive der Intensitätsfunktion. Diese können mit entsprechenden Textur- oder Formprimitiven der 3D-Welt korrespondieren, was sie für die Bildzuordnung und die Objektidentifizierung sehr interessant macht.

- (4) *Invarianz*: Die Auswahl und die Position des markanten Punktes soll invariant gegen die zu erwartenden geometrischen und radiometrischen Änderungen der Bildfunktion sein.

Änderungen der relativen Lage zwischen Aufnahmekamera und Objekt und Änderungen der Beleuchtung und Reflexion schlagen sich im allgemeinen in Verzerrungen des Bildes nieder. Der Auswahlprozeß und die Position, die als Bezugspunkt der lokalen Struktur festgelegt oder geschätzt wird, soll invariant gegen geometrische und radiometrische Verzerrung des Bildes sein. In enger Beziehung zur Invarianz steht die

- (5) *Stabilität*: Die Auswahl der markanten Punkte soll robust gegen Störungen sein.

Als Störungen sollen die "unerwarteten" geometrischen und radiometrischen Änderungen des Bildes und vor allem auch Rauschen im Aufnahmesystem betrachtet werden. Markante Punkte müssen vom Operator gut reproduziert werden können, was vor allem für das Wiederfinden dieser Punkte in einer Bildsequenz sehr wichtig ist.

- (6) *Qualität*: Die Positionierung markanter Punkte soll möglichst genau und biasfrei erfolgen.

Die ersten fünf Kriterien zielen auf eine zuverlässige Lokalisierung. Für Meßprozesse ist darüber hinaus die Genauigkeit ein wichtiger Aspekt. Die Position eines markanten Punktes soll präzise bestimmt und soll möglichst mit der projizierten Position eines korrespondierenden Objektpunktes übereinstimmen. Für die Bestimmung der Tiefeninformation eines Objekts ist die Lagedifferenz der markanten Punkte in einem Stereobildpaar entscheidend, so daß an deren Bestimmung hohe Genauigkeitserwartungen geknüpft sind. Die Forderung nach einer genauen, biasfreien Lokalisierung ist bei einer längeren Sequenz von Bildern notwendig, da sonst durch die Fortpflanzung systematischer Positionsfehler z. B. eine Deformation im stationären Bereich der Szene vorgetäuscht wird.

Leider stehen die verschiedenen Kriterien in Konkurrenz zueinander. So wird z. B. die Stabilität bei der Auswahl markanter Punkte in stark verrauschten Bildern erhöht, wenn durch eine entsprechende Filterung das Rauschen

unterdrückt wird. Durch eine starke Filterung werden in der Regel auch die Strukturen im Bild verwaschen, was sich negativ auf die Deutlichkeit und die Qualität auswirkt.

Die formulierten Kriterien an die Repräsentation und Extraktion markanter Punkte sind Heuristiken, was bedeutet, daß sie normalerweise zufriedenstellende Resultate erwarten lassen und negative Aspekte aus experimentellen oder theoretischen Untersuchungen nicht bekannt sind, oder nur sehr selten beobachtet werden. Sie können anwendungsbezogen weiter ergänzt werden, z. B. wird gelegentlich ein schneller Algorithmus als zusätzliches Kriterium gefordert. Andererseits können Kriterien abgeschwächt werden, z. B. die Seltenheit eines markanten Punktes, wenn ein entsprechender Zuordnungsprozeß den Suchbereich für einen korrespondierenden Punkt sehr stark eingrenzen kann, mit der Konsequenz, daß der Aspekt der Seltenheit im lokalen Ausschnitt eine völlig untergeordnete Bedeutung einnimmt.

In diesen Kriterien finden sich Heuristiken der merkmalsbasierten Bildzuordnung wieder, insbesondere die Kompatibilitätsbedingung (Abschnitt 1.4), die garantieren soll, daß ähnliche Merkmale zuordnungsfähig sind und Merkmale gegen veränderbare Größen bei der Aufnahme invariant sind.

4.2 Operatoren zur Extraktion markanter Punkte

Eine grundlegende Klassifizierung der Operatoren zur Extraktion markanter Punkte kann danach erfolgen, ob sie *direkt* auf den Bildintensitäten operieren, oder auf einer vorausgehenden Segmentierung basieren und dann als *indirekte* Verfahren bezeichnet werden können (Kitchen und Rosenfeld, 1982, Zuniga und Haralick, 1983). Die weitaus größte Gruppe dieser Operatoren - wir wollen sie auch als Punktfinder (Dreschler, 1981) bezeichnen, da der Begriff des Punktoperators sich eng an die Punktoperation anlehnt, die bereits für die bildelementbezogene Manipulation der Intensitäten in der Bildverarbeitung verwendet wird - zielt direkt auf *Ecken*. Ecken im Bild werden dadurch definiert, daß zwei oder mehr gerade Kanten(stücke) im Eckpunkt einen Knoten bilden, d. h. sich darin schneiden oder berühren. Zum Teil wird in der Literatur der Begriff der Ecke, vor allem bei Grauwerteckenfindern, auf eine einfache Keilform (einen L -Knoten) beschränkt.

Die Ecke ist also eine Diskontinuität innerhalb des Kantenverlaufs. Zwei Aspekte folgen direkt aus der Definition der Ecke: (1.) Eckpunkte sind eine Untermenge der Kantenpunkte. (2.) Eckpunkte zeichnen sich durch (starke) Richtungsänderungen in den Kanten aus (Mehrotra und Nichani, 1990). Entsprechend der geometrischen Struktur kann eine Klassifizierung der Ecke als L -, T -, Y -, X -, K -Knoten, etc. erfolgen. Der Eckpunkt bekommt durch diese Klassifizierung ein zusätzliches topologisches Attribut.

Die Forderung, daß ein markanter Punkte sich deutlich von seiner Umgebung unterscheidet, schließt Stellen entlang linear ausgerichteter Strukturen im Bild aus, wie die an geraden Grauwertkanten. Aus Sicht der Bewegungsbestimmung im Bild steht hinter dieser Forderung der Zuordnungsaspekt (bzw. das Aperturproblem, Kapitel 3), daß die zwei Bewegungsfreiheitsgrade jedes markanten Punktes auch bestimmbar sein sollen. Werden diese Freiheitsgrade durch eine Epipolarrestriktion beschränkt, können als markante Punkte auch Punkte auf Grauwertkantenstücken betrachtet werden, wenn die Kantenrichtung sich von der Richtung der Epipolarlinie deutlich unterscheidet. Die eindimensionalen Punktfinder unterscheiden sich nicht signifikant von Operatoren zur Detektion und Lokalisierung von Kanten, so daß wir sie nachfolgend nicht besprechen wollen. Daß die Kantendetektion für die Extraktion markanter Punkte im allgemeineren 2D-Fall aber dennoch eine wesentlich Rolle spielt, wird bei den segmentierungsbasierten Verfahren zur Lokalisierung markanter Punkte besonders deutlich.

4.2.1 Klassifizierung in indirekte und direkte Verfahren

(1) Indirekte Verfahren: Segmentierungsbasierte Punktfinder

Die sich auf eine vorausgehende Segmentierung beziehenden indirekten Verfahren werden spätestens seit Kitchen and Rosenfeld (1982) der weniger attraktiven Gruppe von Punktfindern zugeordnet. Die typische Vorgehensweise besteht darin, daß eine kantenorientierte oder regionenorientierte Segmentierung vorausgeht, die Grenzen der einzelnen Regionen als Linienzüge in Kettencode erfaßt und schließlich dort Ecken lokalisiert werden, wo starke Richtungsänderungen auftreten. Eine ausführliche Übersicht geben Rutkowski und Rosenfeld (1978).

Aus theoretischen und experimentellen Untersuchungen ist bekannt, daß viele Kantendetektoren an Ecken versagen. Für die Lokalisierung von Kantenelementen gibt es zwei elementare Verfahren, die obwohl sie in formal enger Beziehung zueinander stehen, doch sehr unterschiedliche Resultate hervorbringen können. Gradientenbasierte Kantendetektoren, die die Extrema der ersten Ableitung in Gradientenrichtung ermitteln (Canny, 1986a) und damit eine anerkannt gute Beschreibung des Bildes produzieren, hinterlassen bei der diskreten Berechnung in der Umgebung einer Ecke oft Lücken. D. h. die auf den Extrema der ersten Ableitung basierenden Kantenoperatoren geben im allgemeinen direkt keine Auskunft über eine Ecke. In den theoretischen Arbeiten, die alle das Bild durch eine Modellfunktion einer Ecke - in der Regel einen Gaußgefilterten Stufenkeil - beschreiben, sind systematische Abweichungen in der berechneten Position der Ecke (z. B. Bergholm, 1987, De Micheli et al., 1989) nachgewiesen worden. Bei kleinem Öffnungswinkel einer keilförmigen Struktur (vgl. Abbildung 4.2 in Abschnitt 4.2.4) und bei starker Glättung kann der systematische Fehler mehrere Pixel betragen (Rohr, 1992). Die alternative Möglichkeit zur Kantenlokalisierung besteht in der Identifikation der Kantenposition als Nulldurchgang der zweiten Ableitung der Bildfunktion. Durch Filterung mit einem *LoG*-Operator (Marr und Hildreth, 1980) oder einem Operator, der die zweite Ableitung eines gefilterten Bildes in Richtung des Gradienten berechnet, ergeben sich im Gegensatz zu den gradientenbasierten Extrema-Operatoren bei den Nulldurchgängen geschlossene Konturen. Berzins (1984) hat gezeigt, daß die Kontur durch den Eckpunkt eines *L*-Knotens verläuft, unabhängig von der Größe des Glättungsfilters. Für einen *Y*-Knoten mit verschiedenen (rauschfreien) Intensitätsniveaus in den drei benachbarten Regionen gilt gleiches; allerdings wird eine zusätzliche unechte Kante (engl. *spurious edge*) detektiert (De Micheli et al., 1989). Generell sind die gradientenbasierten Kantendetektoren weniger rauschempfindlich als die Nulldurchgangsoperatoren, was auch zu erwarten ist, da bei jeder Differentiation Rauschen verstärkt wird.

Verfahren zur Detektion von geometrischen Primitiven wie Ecken oder auch kreisförmigen Strukturen nützen die Segmentierung oft in Zusammenhang mit entsprechendem Vorwissen. Ist z. B. a priori bekannt, daß es sich bei dem segmentierten Bildausschnitt um einen Kreis oder eine Ellipse handelt, kann die Position der Kantenpunkte zur Schätzung der entsprechenden Parameter (Mittelpunkt, Radius oder Halbachsen) verwendet werden. Für die Berechnung von Ecken als Schnittpunkt von geraden Kantenstücken gilt gleiches. Durch die scheinbar restriktive Voraussetzung, daß entsprechendes Vorwissen vorliegen muß oder eventuell auch nur als Heuristik postuliert werden kann, scheint sich dieses Vorgehen nur bedingt zur Detektion markanter Punkte zu empfehlen. Dennoch sind viele solche Verfahren zur Nachbereinigung von Segmentierungen in der Praxis mit Erfolg im Einsatz. Beispielsweise hat Vosselman (1992, S. 118-121) in Experimenten mit Blocksweltszenen Eckpunkte, die von einem Punktfinder nicht oder fehlerhaft lokalisiert wurden, durch entsprechende Verfahren ergänzt, eliminiert bzw. auch korrigiert. Wir werden später noch sehen, daß entsprechende Modellierungen, z. B. einer Ecke als Schnittpunkt von Geraden in Konturrichtung, auch bei den direkten intensitätsbasierten Verfahren formuliert werden. Ein beträchtlicher Informationsgewinn bei entsprechenden Punktfindern folgt gerade aus dieser Modellierung: Ein markanter Punkt kann dadurch z. B. als Grauwertecke interpretiert werden.

Die starke Popularisierung des *LoG*-Operators durch Marr und Hildreth (1980) hat in der Folge viele Forschungsarbeiten motiviert. Knickpunkte in diesen Konturen lassen sich durch eindimensionale Analyse der Richtungsänderungen, abgetragen über der Bogenlänge, detektieren. Daß darin auch potentiell Ecken enthalten sind, haben wir oben schon angesprochen. Durch Feinheiten der prozeduralen Realisierung, z. B. mit Filtern, Prädiktion und Approximation der Konturen, auch mit verschiedenen Krümmungsmodellen, entstanden viele Varianten zur Detektion der Knickpunkte, die treffend auch als Krümmungsprimitiva bezeichnet werden. Experimente zur Stereozuordnung mit solchen markanten Punkten finden sich in Schenk et al. (1990), zur lageinvarianten Formbeschreibung in Korn (1986) und Asada und Brady (1986). Letztgenannte haben die Krümmung der Kontur über verschiedenen Auflösungen repräsentiert, um damit auch eine skalierungsinvariante Formbeschreibung zu finden. Die lokalen Extrema der verschiedenen Auflösungsebenen werden miteinander verbunden und durch einen Baum repräsentiert. Die Position der Knotenpunkte der Kontur können über die Bäume verfolgt werden. Diese Eigenschaft macht sie als markante Punkte auch für die Zuordnung sehr interessant.

(2) Direkte Verfahren: Intensitätsbasierte Punktfinder

Die attraktiveren direkten Verfahren zur Lokalisierung markanter Punkte sollen in den nächsten Abschnitten detaillierter vorgestellt werden. Sie beruhen nicht auf Segmentierungsergebnissen, sondern auf den Intensitätswerten oder Intensitätsgradienten direkt. Die durch direkte Verfahren extrahierten markanten Punkte können

eine wichtige Komponente innerhalb der Segmentierung darstellen, was ein Vorteil dieser Punktefinder ist. Für diese Operatoren ist typisch, daß sie lokale fensterbezogene Information auswerten und dabei in der Mehrzahl alle auf den ersten und zweiten Ableitungen der Bildintensitäten basieren, womit eine gewisse Abhängigkeit von der Qualität der verwendeten Ableitungsoperatoren besteht.

Die direkten Verfahren wollen wir weiter unterteilen nach

- *statistischen Verfahren,*
- *differentialgeometrischen Verfahren* und nach
- *klassifizierenden Verfahren,*

je nachdem ob elementare statistische Eigenschaften¹ oder differentialgeometrische Größen der Bildfunktion ausgewertet werden, um einen markanten Punkt zu detektieren, oder ob durch eine Klassifizierung im Zuge einer Parameterschätzung detektierte markante Punkte nach Klassen unterschieden werden können.² Wenn Operatoren mehrere Eigenschaften in sich vereinen, werden sie der Gruppe zugeordnet, die sie am meisten auszeichnet. Tendenziell kann davon ausgegangen werden, daß der Zeitaufwand für die Berechnung von den statistischen über die differentialgeometrischen hin zu den klassifizierenden Verfahren ständig zunimmt.

4.2.2 Punktefinder: Statistische Verfahren

Die Beschreibung von Eigenschaften der Bildfunktion nach statistischen Kriterien erfolgt unter dem Gesichtspunkt, Maße für die Texturierung des Bildes zu finden, die es gestatten, einen inhomogen texturierten Bereich um einen Punkt von seiner eher homogenen Umgebung zu unterscheiden. Entsprechende Punkte hat Moravec (1977) als *Distinctive Areas*, den Operator zur Lokalisierung dieser Punkte als Interestoperator bezeichnet. Ein Hintergrund dieser Vorgehensweise ist die Beschreibung eines Bildes durch ein stochastisches Bildmodell (Jähne, 1989). Die auf der Basis dieses Bildmodells abgeleiteten ikonischen (lokalen) Merkmale, wie Mittelwert, Varianz und Kovarianzen liegen bildelementbezogen berechnet vor und können auch als Merkmalskarte bezeichnet werden (Tsotsos, 1987). Die Strukturierung der Bildintensitäten, mit der der Texturbegriff im allgemeinen identifiziert wird, ist im Regelfall allerdings sehr komplex. Es ist deshalb nicht leicht, Textur im allgemeinen durch einfache Merkmale hinreichend zu charakterisieren (Jähne, 1989).

Hannah 1974

Die Auswertung von Stereobildsequenzen zur Navigation auf der Basis markanter Punkte hat Hannah (1974) vorgestellt. Die markanten Punkte des Stereopaars werden zugeordnet und dienen als Verknüpfungspunkte für die Bewegungsbestimmung des sich bewegenden Fahrzeuges. In dieser frühen Arbeit hat Hannah als Kriterium für einen Punktefinder die Varianz des Signals im Fenster

$$O_{\text{Varianz}} = \frac{1}{N-1} \sum_W (I(i, k) - \bar{I}(x, y))^2$$

gewählt, wobei $\bar{I}(x, y)$ den mittleren Intensitätswert der N Bildpunkte im Fenster um den Punkt (x, y) angibt. Ein lokal hohes Operatorsignal legt die Lage der markanten Punkte fest. Der Varianzoperator ist rotationsinvariant, was auch noch gilt, wenn anstatt der Gleichgewichtung eine andere isotrope Gewichtung verwendet wird, z. B. die Faltung mit einem Binomialkern anstelle einer Rechteckmaske erfolgt. Das Filterergebnis ist dadurch unabhängig von der Orientierung der Textur. Der Operator spricht auf alle Strukturen im Bild an, insbesondere auch Kanten, weshalb er für die Bestimmung markanter Punkte im Grunde ungeeignet ist.

¹Statistik wird hier im Sinne der Elementarstatistik verstanden. Bei den Verfahren gilt es vorwiegend Momente zu berechnen.

²Die klassifizierenden Verfahren nützen die Residuenquadratsummen, die bei der Parameterschätzung mit entsprechenden Modellen für markante Punkte berechnet werden können. Entsprechend der Struktur kann eine Klassifikation der markanten Punkte in Ecken und kreisförmige Texturen erfolgen, oder verfeinert nach topologischen Gesichtspunkten in Ecktypen als $L-$, $T-$, etc. $-$ Knoten.

Interestoperator von Moravec 1977

Der Interestoperator von Moravec (1977, 1979, 1981) wird in der Literatur sehr häufig zitiert, insbesondere wenn über Experimente berichtet wird, so daß dieser Operator der in der Praxis wohl am häufigsten verwendete Punktfinder ist. Der Operator beruht auf der Berechnung von vier gerichteten Varianzen in jedem Bildpunkt (x, y) gemäß

$$\begin{aligned} O_{0^\circ} &= \frac{1}{N} \sum_W (I(i, k) - I(i, k+1))^2 \\ O_{45^\circ} &= \frac{1}{N} \sum_W (I(i, k) - I(i+1, k+1))^2 \\ O_{90^\circ} &= \frac{1}{N} \sum_W (I(i, k) - I(i+1, k))^2 \\ O_{135^\circ} &= \frac{1}{N} \sum_W (I(i, k) - I(i+1, k-1))^2 . \end{aligned} \quad (4-1)$$

Um in allen vier Richtungen gleiche Anzahl an Summanden (N) zu haben, nimmt Moravec Bildelemente außerhalb eines quadratisch oder rechteckig gewählten Fensters W um den Mittelpunkt (x, y) hinzu. Ansonsten werden bei Summation über alle Bildelemente, die nur innerhalb eines entsprechenden Bildausschnittes liegen, in den beiden Diagonalrichtungen weniger Summanden aufaddiert. Bei den Argumenten, über die in (4-1) summiert wird, handelt es sich um die Quadrate der in Zeilen- und Spaltenrichtung sowie in den beiden Diagonalrichtungen berechneten Gradienten. Um statistisch begründet von Varianzen sprechen zu können, muß dabei vorausgesetzt werden, daß die Erwartungswerte der Gradienten in den entsprechenden Richtungen identisch Null sind. Aus theoretischer Sicht wäre für die Diagonalrichtungen auch eine Normierung der Gradienten mit $\sqrt{2}$ (quadratische Rasterung der Länge 1 vorausgesetzt) angebracht.

Als Operatorsignal des Moravec-Operators wird das Minimum der vier gerichteten Varianzen bestimmt und das Resultat dem zentralen Bildelement zugewiesen:

$$O_{Moravec}(x, y) = \min(O_{0^\circ}, O_{45^\circ}, O_{90^\circ}, O_{135^\circ}) \quad (4-2)$$

Eine hohe minimale gerichtete Varianz ist ein Indikator für Punkte mit hoher Varianz der Bildfunktion in allen Richtungen. Die Selektion der markanten Punkte erfolgt durch *Schwellwertbildung* und anschließende *lokale Maxima*-Bestimmung in $O_{Moravec}(x, y)$. Schwellwertbildung und Unterdrückung aller Nicht-Maxima finden sich auch bei vielen anderen Punktfindern wieder. Der Schwellwert dient dazu, zufällige und wenig deutliche, auch durch das Rauschen im Bild verursachte lokale Maxima auszuschließen. Die Chance diese Punkte in anderen Bildern wiederzufinden wird gering eingeschätzt. Punkte, die keine hohe Stabilität erwarten lassen, wurden aber im Kriterienkatalog ausgeschlossen. Durch die lokale Maxima-Suche wird in einer stärker texturierten lokalen Umgebung schließlich die auffälligste Stelle detektiert.

Für die Ermittlung von Punktkorrespondenzen hat Moravec (1981) markante Punkte in einem Bild lokalisiert und deren Auftreten an dieser Position über mehrere Auflösungsstufen überprüft. Die Fenstergröße wurde in Experimenten üblicherweise mit 6×6 Bildelementen festgelegt. Die reduzierten Auflösungsstufen sind durch Mittelung und Reduktion um den Faktor 2 (linear) als Bildpyramide vorab bestimmt wurden. Die Zuordnung erfolgt durch Korrelation mit dem Stereobild, wobei die Position der besten Zuordnung einer Auflösungsstufe als Näherung eines über die Pyramide kontrollierten Suchbereichs von grob-zu-fein weitergegeben wurde. Auf dem Originalbild führt dann die Korrelation zu Punktkorrespondenzen der markanten Punkte. Auf dem Hintergrund der Rechenzeiteinsparung wurde als weitere Besonderheit der Operator ursprünglich nicht an jeder Stelle im Bild berechnet, sondern jeweils um eine halbe Operatorbreite versetzt. Da dies auf Kosten der Präzision der lokalisierten Punkte geht und die Rechnerleistung im letzten Jahrzehnt beträchtlich gesteigert wurde, ist diese Maßnahme heute nicht mehr sinnvoll.

Weitere Arbeiten

Eine Schwäche des Moravec-Operators besteht darin, daß die gerichteten Varianzen in vier Richtungen explizit berechnet werden. Moravec (1981) weist darauf hin, daß bei Kanten, die genau zwischen diesen vier Richtungen

verlaufen, eine relative hohe minimale gerichtete Varianz zu erwarten ist, was zur Lokalisierung markanter Punkte an solchen Stellen führen kann. Diese konzeptionelle Schwäche läßt sich zwar verringern, in dem die Winkelabstufung feiner unterteilt wird, aber nicht grundlegend beseitigen. Hannah (1980) hat den Operator abgewandelt, indem sie nicht das Minimum der gerichteten Varianz verwendet, sondern

$$O_{Hannah}(x, y) = O_{Varianz} \cdot \min \left(\frac{O_{0^\circ}}{O_{90^\circ}}, \frac{O_{90^\circ}}{O_{0^\circ}}, \frac{O_{45^\circ}}{O_{135^\circ}}, \frac{O_{135^\circ}}{O_{45^\circ}} \right). \quad (4-3)$$

berechnet. Die multiplikativ eingehende Varianz des Signals trägt dazu bei, die stark texturierten Stellen im Bild zu finden, die Quotienten der gerichteten Varianzen in zueinander orthogonalen Richtungen sorgen für eine gleich hohe gerichtete Varianz in allen Richtungen.

Weitere Versuche den Moravec-Operator zu verbessern hat Thorpe (1984) vorgeschlagen. Er versucht zunächst über einen Kantenoperator eine Vorzugsrichtung der Textur zu ermitteln, um dann in dieser Richtung eine gerichtete Varianz zu bestimmen und aus den lokalen Maxima markante Punkte zu lokalisieren. Der Vorschlag löst im Prinzip das Problem der Verfeinerung diskreter Winkelabstufungen, ist allerdings abhängig von der Qualität der über den Kantenoperator bestimmten Richtung.

Markante Punkte, die über ihre Merkmale ausschließlich statistische Eigenschaften einer Bildfunktion bemessen, sind nicht unbedingt leicht interpretierbare Punkte, vor allem nicht mit Blick auf die 3D-Szene. Um diese Schwierigkeit zu vermeiden, sind Punktefinder entwickelt worden, die in der Mehrzahl zusätzlich ein geometrisches Modell einer Grauwertecke verwenden und solche Stellen im Bild zu lokalisieren versuchen.

4.2.3 Punktefinder: Differentialgeometrische Verfahren

Die differentialgeometrische Beschreibung von Flächen durch lokale Extrema, Sattelpunkte, und auch verschiedene Krümmungsmaße kann entsprechend auf eine kontinuierliche Bildfunktion angewandt werden, wenn vorausgesetzt wird, daß die Bildfunktion mindestens zweimal stetig differenzierbar ist. Diese differentialgeometrisch begründeten Merkmale der Bildfunktion lassen sich verwenden um daraus Punktefinder abzuleiten. Einige wenige elementare Grundlagen der Differentialgeometrie wollen wir der Diskussion der Punktefinder voranstellen.

Ganz allgemein kann eine Bildoberfläche \mathbf{S} durch die Gleichung

$$\mathbf{S} = \begin{pmatrix} x \\ y \\ I(x, y) \end{pmatrix}$$

in Parameterform definiert werden, wobei $I(x, y)$ die Intensität über der Bildebene (x, y) angibt und \mathbf{S} die Bildoberfläche im Raum $\{x, y, I\}$ repräsentiert. Differenzierbarkeit von \mathbf{S} vorausgesetzt, kann der auf Länge 1 normierte Normalenvektor der Fläche in jedem Punkt gemäß

$$\mathbf{n}_S = \frac{1}{|\mathbf{S}_x \times \mathbf{S}_y|} (\mathbf{S}_x \times \mathbf{S}_y) = \frac{1}{\sqrt{I_x^2 + I_y^2 + 1}} \begin{pmatrix} -I_x \\ -I_y \\ 1 \end{pmatrix}$$

berechnet werden. Für die Betrachtung von Eigenschaften der Oberflächen werden üblicherweise die erste und zweite quadratische Fundamentalform von \mathbf{S}

$$\Phi_1 = d\mathbf{S}^T d\mathbf{S} = \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

und

$$\Phi_2 = -d\mathbf{S}^T d\mathbf{n}_S = \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} L & M \\ M & N \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}$$

herangezogen. Die Metrik auf der Fläche wird vollständig durch die erste quadratische Fundamentalform definiert, deren Koeffizienten für die Bildoberfläche sich zu $E = \mathbf{S}_x^T \mathbf{S}_x = I_x^2 + 1$, $F = \mathbf{S}_x^T \mathbf{S}_y = I_x I_y$, $G = \mathbf{S}_y^T \mathbf{S}_y =$

$I_y^2 + 1$ ergeben. Ebenso einfach ist die Berechnung der Koeffizienten der zweiten quadratischen Fundamentalform, die mit dem Einheitsnormalenvektor der Flächennormalen \mathbf{n}_S gemäß $L = \mathbf{S}_x^T \mathbf{n}_S$, $M = \mathbf{S}_x^T \mathbf{S}_y \mathbf{n}_S$, $N = \mathbf{S}_y^T \mathbf{n}_S$ ermittelt werden können.

Für die Klassifizierung der Flächenpunkte entscheidend sind die Koeffizienten L, M, N , bzw. die Hesse-Matrix H der Intensitätsfunktion

$$H = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix} = \frac{1}{\sqrt{I_x^2 + I_y^2 + 1}} \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

die bis auf die Normierung mit $|\mathbf{S}_x \times \mathbf{S}_y|$ der Matrix der zweiten Fundamentalform mit den Koeffizienten L, M, N entspricht. Die Determinante der Hesse-Matrix

$$\det H = I_{xx}I_{yy} - I_{xy}^2,$$

die von der Orientierung des Bezugssystems unabhängig ist, hat dabei eine besondere Funktion. Mit der Determinante $\det H$, oder äquivalent dazu mit den noch zu behandelnden Hauptkrümmungen, lassen sich Flächenpunkte klassifizieren. Dabei handelt es sich bei einem solchen Flächen- oder Bildpunkt um einen *elliptischen*, einen *hyperbolischen* oder einen *parabolischen* Punkt, wenn gilt:

$$\begin{array}{ll} \text{elliptischer Punkt} & : \det H > 0 \\ \text{hyperbolischer Punkt} & : \det H < 0 \\ \text{parabolischer Punkt} & : \det H = 0 \end{array} .$$

Bei elliptischen Punkten liegen alle Punkte der Umgebung des Punktes auf einer Seite der Tangentialebene. Ist die Umgebung sattelähnlich, d. h. zeigen die konkaven Seiten der Hauptnormalschnitte nach entgegengesetzten Richtungen, handelt es sich um einen hyperbolischen Punkt. Das analytische Kennzeichen eines parabolischen Punktes ist $\det H = 0$, was einer Hauptkrümmung von Null gleichkommt.

Neben dieser Flächenpunktklassifizierung bauen die differentialgeometrisch argumentierenden Punktefinder insbesondere auf die Krümmung der Fläche. Zur Charakterisierung der Krümmungen einer Fläche werden hauptsächlich die mittlere und die Gauß'sche Krümmung herangezogen. Mit den *Hauptkrümmungen* κ_{max} und κ_{min} , die als Wurzeln der quadratischen Gleichung (z. B. Bronstein-Semendjajew, 1977, S. 223)

$$(EG - F^2)\kappa^2 - (EN - 2FM + GL)\kappa + (LN - M^2) = 0$$

erhalten werden, sind die *mittlere Krümmung* κ_M und die *Gauß'sche Krümmung* κ_G gemäß

$$\begin{aligned} \kappa_M &= \frac{1}{2}(\kappa_{max} + \kappa_{min}) = \frac{EN - 2FM + GL}{2(EG - F^2)}, \\ \kappa_G &= \kappa_{max}\kappa_{min} = \frac{LN - M^2}{EG - F^2} = \frac{I_{xx}I_{yy} - I_{xy}^2}{(I_x^2 + I_y^2 + 1)^2} \end{aligned}$$

definiert. Die Gauß'sche Krümmung ist einfach dadurch zu berechnen, daß die Determinante der Hesse-Matrix $\det H$ mit der Länge der Flächennormalen in diesem Punkt normiert wird.

Die im folgenden zu betrachtenden Kurven sind solche, die sich als Schnittlinien der Intensitätsoberfläche mit Ebenen ergeben, die orthogonal zur Bildebene stehen. Für diese "Vertikalschnitte" werden die Begriffe Normalkrümmung, Hauptkrümmung, Gauß'sche Krümmung, etc. z. B. von Nagel (1983) und vielen anderen ebenfalls verwendet. Sie sind damit nicht konsistent mit der oben dargestellten generell üblichen Verwendung dieser Begriffe. Mit der Richtung der Spur dieser Vertikalebene in der Bildebene \mathbf{s} , $|\mathbf{s}| = 1$ und der Hesse-Matrix ist die *Normalkrümmung* der Schnittlinie dieser Ebene mit der Intensitätsfunktion $I(x, y)$ durch

$$\kappa(\mathbf{s}) = \mathbf{s}^T H \mathbf{s}$$

festgelegt. Die Hauptkrümmungen κ_{max} und κ_{min} der Oberfläche

$$\kappa_{max} = \max \kappa(\mathbf{s}), \quad \kappa_{min} = \min \kappa(\mathbf{s}),$$

können als Eigenwerte der Hesse-Matrix berechnet werden, d. h. sie ergeben sich aus der quadratischen Gleichung $\kappa^2 - (\text{trace } H)\kappa + \det H = 0$. Entsprechend stehen die Spur von H und die Determinante von H

$$\text{trace } H = \kappa_{max} + \kappa_{min}, \quad \det H = \kappa_{max}\kappa_{min}$$

in Verbindung mit der mittleren und der Gauß'schen Krümmung.

Weitere Krümmungen, auf die wir zurückgreifen werden, sind die *Konturkrümmung* κ_C mit Richtung $\mathbf{s}^T = (-I_y, I_x)/|\nabla I|$ und die *Krümmung in Gradientenrichtung* κ_N mit $\mathbf{s}^T = (I_x, I_y)/|\nabla I|$ (Bartsch, 1987)

$$\kappa_C = \frac{I_y^2 I_{xx} - 2I_x I_y I_{xy} + I_x^2 I_{yy}}{I_x^2 + I_y^2}$$

$$\kappa_N = \frac{I_x^2 I_{xx} + 2I_x I_y I_{xy} + I_y^2 I_{yy}}{I_x^2 + I_y^2}$$

Zur Bestimmung der Extrema bzw. der Krümmungen sind in den bisherigen Betrachtungen kontinuierliche Flächen vorausgesetzt. Ein zweiter Aspekt ist daher der Übergang vom diskret vorliegenden Bild zur kontinuierlichen Bildfunktion, was in der Theorie bei entsprechend bandbegrenzten diskreten Bildern aufgrund des Abtasttheorems von Shannon exakt möglich ist. Die Begrenzung der Bandbreite hat zur Folge, daß die Bildfunktion keine Unstetigkeiten aufweist. Um eine kontinuierliche Funktion lokal zu beschreiben, bietet sich die Möglichkeit diese in eine Taylor-Reihe zu entwickeln. Vorausgesetzt werden muß dabei, daß die Funktion hinreichend oft differenzierbar ist. Die Rekonstruktion der Bildfunktion in einem lokalen Ausschnitt erfolgt in der Regel durch lokale Approximation. Die Flächenapproximation durch bivariate Polynome ist der fast ausschließlich beschrittene Lösungsweg, den wir bei der Bewegungsschätzung in örtlich begrenzten Bildausschnitten als eine Modellierungskomponente bereits ausführlich dargelegt haben (siehe Abschnitt 3.3).

Die lokale Anpassung der Bildfunktion an ein bivariates Polynom zweiter Ordnung

$$f(x, y) = f_0 + f_x x + f_y y + \frac{1}{2} f_{xx} x^2 + f_{xy} xy + \frac{1}{2} f_{yy} y^2$$

bzw.

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f^T(\mathbf{x}_0) \mathbf{x} + \frac{1}{2} \mathbf{x}^T H_f(\mathbf{x}_0) \mathbf{x}$$

durch Schätzung der Koeffizienten des bivariaten Polynoms nach Kleinsten Quadraten

$$\sum_R (I(\mathbf{x}) - f(\mathbf{x}, \mathbf{p}))^2 \rightarrow \min$$

erhält man eine Schätzung des Parametervektors $\mathbf{p}^T = (f_0, f_x, f_y, f_{xx}, f_{xy}, f_{yy})$, dessen Elemente die Ableitungen der lokal approximierten bivariaten Bildfunktion in der in einem Bildpunkt (x, y) örtlich festgelegten Entwicklungsstelle ($\mathbf{x}_0 = 0$) angeben. Diese Schätzung kann soweit ausgearbeitet werden, daß daraus Filtermasken für die Bestimmung der Ableitungen explizit vorliegen; ein Lösungsweg, den vor allem Beaudet (1978) vorangebracht hat (siehe auch Nagel, 1983).

In der Kleinsten-Quadrate-Zielfunktion zur Schätzung der Ableitungen werden alle Bildelemente innerhalb des Fensters gleich gewichtet. Anstelle dieser Faltung mit einer Rechteckmaske kann auch eine andere Gewichtung, zum Beispiel ein Binomialkern, treten.

Einige Operatoren zur Lokalisierung markanter Punkte wollen wir nun im einzelnen darstellen. Auf eine strenge Unterscheidung von f, f_x, \dots, f_{yy} und I, I_x, \dots, I_{yy} kann dabei verzichtet werden. Die Ableitungen I_x, \dots sind im weiteren als die aus der Kleinsten-Quadrate-Anpassung geschätzten Größen \hat{f}_x, \dots zu verstehen.

Beaudet 1978

Beaudet schätzt die Ableitungen der lokal approximierten Bildfunktion durch die bereits angesprochenen von ihm entwickelte Operatormasken, die für die Größen von 3×3 bis 8×8 Pixel und für bivariate Polynome bis zum 4. Grad explizit ausgearbeitet vorliegen. Der vorgeschlagene Operator

$$O_{DET} = I_{xx} I_{yy} - I_{xy}^2 \quad (4-4)$$

hat als Operatorsignal die Determinante der Hesse-Matrix $\det H$. Als markante Punkte werden solche Stellen im Bild detektiert, die lokale Extrema sind und in denen der Absolutwert des Operatorsignals eine gewisse Schwelle überschreitet. Ein hohes Operatorsignal wird an Ecken und Sattelpunkten erzielt, wobei aber in Experimenten mit scharfen Konturen der Operator weniger gut abgeschnitten hat (Kitchen und Rosenfeld, 1982).

Kitchen und Rosenfeld 1982

Die Autoren schlagen mehrere Verfahren zur Detektion von Grauwertecken vor, wobei sie als Basis ein Maß für die "Eckigkeit" (engl. cornerity oder auch cornerness) definieren, das sich aus dem Produkt von Krümmung und Kantenstärke errechnet. Besondere Aufmerksamkeit hat der Operator erlangt, der eine lokal geschätzte Intensitätsoberfläche voraussetzt. Abgeleitet wurde der Operator über die Änderung der Gradientenrichtung entlang des Konturverlaufs durch den betrachteten Punkt und multipliziert mit der lokalen Gradientenstärke.

$$O_{K\&R} = \frac{I_y^2 I_{xx} - 2I_x I_y I_{xy} + I_x^2 I_{yy}}{I_x^2 + I_y^2} \quad (4-5)$$

Das Operatorsignal ist offensichtlich identisch mit der Konturkrümmung κ_C . Eine Alternative zur Ableitung des Operators besteht darin, die Kontur als Isolinie mit konstantem Intensitätswert $I(x, y) = konst$ durch den betrachteten Punkt zu definieren. Die Krümmung dieser ebenen Kurve, deren Gleichung in Parameterform mit $(x = x(s), y = y(s))$ definiert sein soll, kann mit den ersten und zweiten Ableitungen nach s : x', y', x'', y'' gemäß

$$\kappa = (x'y'' - x''y') / (x'^2 + y'^2)^{3/2} \quad (4-6)$$

berechnet werden. Die Multiplikation von κ mit der Gradientenstärke führt zu (4-5). Der Berechnungsweg über die Differentiation der durch $I(x(s), y(s)) = I(x_0, y_0)$ implizit gegebenen Isolinie ist in Kitchen und Rosenfeld (1982) und Nagel (1983) angedeutet. Setzen wir $I_y \neq 0$ und wegen (4-5) $|\nabla I| \neq 0$ voraus (präzise wäre die Existenz und Eindeutigkeit der durch $I(s) = I(0)$ definierten Kurve zu klären), können y', y'' als Funktionen von x', x'' in der lokalen Umgebung des betrachteten Punktes dargestellt werden

$$\begin{aligned} y' &= -\frac{I_x}{I_y} x' \\ y'' &= \frac{1}{I_y^2} ((I_{xy} I_x - I_{xx} I_y) x'^2 + (I_{yy} I_x - I_{xy} I_y) x' y' - I_x I_y x'') \end{aligned}$$

Eingeführt in (4-6) ergibt sich schließlich (4-5). Die lokalen Maxima des Operatorsignals werden, sofern sie eine gewisse Schwelle überschreiten, als Grauwertecken ausgewählt. Als weitere Heuristik wird von den Autoren empfohlen, die Kantenstärke $|\nabla I|$ vor der Multiplikation mit der Krümmung durch Nicht-Maxima Unterdrückung entlang der Gradientenrichtung zu bearbeiten. Nagel (1983) zeigt, daß durch diese Heuristik insbesondere die Kurve der lokalen Wendepunkte selektiert wird (Kurve durch Q_y , T, Q_x in Figur 4.1). Maximaler Operatorrespons ist dann für den Eckpunkt T zu erwarten.

Dreschler und Nagel 1982

Das in Abbildung 4.1 dargestellte Modell einer Grauwertecke ist der Ausgangspunkt für die Ableitung des von Dreschler-Nagel vorgestellten Punktefinders. Geometrisch liegt dem Modell eine keilförmige Struktur zugrunde, wobei der Übergang zwischen den beiden Intensitätsniveaus von Keil und Umgebung glatt sein soll. Der Operator beruht auf der Gauß'schen Krümmung bzw. der Determinante der Hesse-Matrix. Ein wesentliches Kriterium für den Entwurf des Operators war die Beobachtung, daß in der Umgebung von Grauwertecken Minima und Maxima der Gauß'schen Krümmung paarweise und mit entgegengesetztem Vorzeichen auftreten (in Punkt E positiv, in Punkt H negativ). Der Operator beinhaltet dann folgende Operationen:

1. Bestimme die Gauß'sche Krümmung $\kappa_G = \det H$ und lokalisierere Punkte mit extremaler - positiver und negativer - Gauß'scher Krümmung. Punkt E gibt die Position des elliptischen Punktes mit maximaler positiver Krümmung κ_G , Punkt H die Position des hyperbolischen Punktes mit maximaler negativer Krümmung κ_G an.

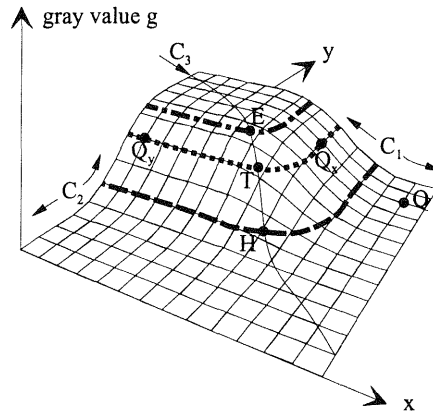


Abbildung 4.1: Modell einer Grauwertecke (aus Nagel, 1983, S. 89)

2. Die Punkte E und H werden einander zugeordnet, wenn die Hauptkrümmungen mit entgegengesetztem Vorzeichen ungefähr die gleiche Ausrichtung wie die Verbindungslinie E-H haben. Für die Zuordnung von E und H muß die Größe des Suchfensters, z. B. von 3×3 Pixel, festgelegt werden.
3. Selektiere den Punkt T auf der Verbindungslinie E-H, in dem die entsprechende Hauptkrümmung das Vorzeichen wechselt. Die Hauptkrümmung von 0 entspricht einer maximalen Neigung der Kurve E-H und hat eine Gauß'sche Krümmung $\kappa_G = 0$ zur Folge, d. h. es handelt sich um einen parabolischen Punkt.

Kommen bei der Paarbildung der Extremalpunkte in der lokalen Umgebung mehrere Kandidaten zustande, wird das Paar ausgewählt, dessen Ausrichtung der Hauptkrümmungsachsen am exaktesten übereinstimmt. Daß der Operator neben Ecken auch helle oder dunkle Flecken detektieren kann, ergibt sich aus folgender Betrachtung: Ist das Maximum der Gauß'schen Krümmung von einem Ring von Minima umgeben, so liegt ein isolierter Punkt vor, was anschaulich in Abbildung 4.1 dem entspricht, daß die Verbindungslinie E-H eine Mantellinie einer mehr oder weniger rotationssymmetrischen Fläche darstellt. Die Position des Grauwertextremums ist durch $|\nabla I| = 0$ gekennzeichnet (siehe auch Nagel, 1985).

Zur Robustifizierung wurde in experimentellen Untersuchungen mit dem Operator mit weiteren Schwellwerten gearbeitet, z. B. für die lokalen Extrema der Gauß'schen Krümmung, deren Absolutwerte ein Minimum nicht unterschreiten sollen (Dreschler, 1981), oder mit Schwellwerten für die Gradienten (Zuniga und Haralick, 1983). Weitere Verfahrensvariationen, wie die zweistufige Berechnung der Gauß'schen Krümmung mit Beaudet-Operatoren der Größe 5×5 und, zur präziseren Lokalisierung, mit 3×3 -Operatoren, haben experimentell nicht den erwünschten Erfolg gebracht (Dreschler, 1981, Bartsch, 1987).

Zuniga und Haralick 1983

In engem Zusammenhang zu Kitchen und Rosenfeld (1982) steht die Arbeit von Zuniga und Haralick. Als parametrische Form zur lokalen Flächenapproximation wählen sie ein bikubisches Polynom in beiden Koordinaten (x,y) . Die Approximation der Intensitätsfunktion durch eine kontinuierliche Fläche, die aus der lokalen Nachbarschaft berechnet wird, ist essentiell für das von Haralick (1980) ausgearbeitete Facettenmodell. Die Eckendetektion setzt auf zwei Grundbausteine. Zum einen wird eine Kante vorausgesetzt, zum andern werden signifikante Änderungen der Kantenrichtung ermittelt. Unter drei vorgeschlagenen Prozeduren war folgendes Verfahren am erfolgreichsten, nach dem ein Punkt als Eckpunkt identifiziert wird, wenn

1. der Punkt ein Kantenpunkt ist. Um als Kantenpunkt detektiert zu werden, muß die Gradientenstärke einen Schwellwert überschreiten und ein Nulldurchgang in der zweiten gerichteten Ableitung vorliegen. Die zweite gerichtete Ableitung wird dabei in Richtung des Gradienten berechnet. Der Nulldurchgang muß bei einem Radius von 1 Pixel innerhalb von $\pm 14,9^\circ$ auftreten, wobei für die Berechnung der lokalen Intensitätsfunktion eine Umgebung von 7×7 Pixel herangezogen wurde.

2. der Betrag der Krümmung der Kontur im betrachteten Kantenpunkt

$$|\kappa| = |(I_y^2 I_{xx} - 2I_x I_y I_{xy} + I_x^2 I_{yy}) / (I_x^2 + I_y^2)^{3/2}| > \kappa_{Schwell}$$

einen gewissen Schwellwert überschreitet.

Durch den zur Kantendetektion verwendeten Nulldurchgangsoperator werden ähnlich den segmentierungsbasierten Verfahren zunächst Konturpunkte lokalisiert. Eine Verkettung dieser Konturpunkte entfällt aber, da durch die bei der lokalen Flächenapproximation ermittelten zweiten Ableitungen direkt die Krümmung dieser Kurve in jedem Kantenpunkt berechnet werden kann. Abgesehen von der Multiplikation mit der Gradientenstärke entspricht der Operator in diesem zweiten Schritt dem Operatorsignal von Kitchen und Rosenfeld.

Ergänzungen

Bei der Darstellung der Operatoren haben wir uns immer wieder auf Abbildung 4.1 bezogen, um damit die große Ähnlichkeit der differentialgeometrischen Verfahren direkt aufzuzeigen. Den Nachweis und die Randbedingungen für die Äquivalenz der Punktfinder von Dreschler-Nagel und Kitchen-Rosenfeld hat Nagel (1983) erbracht.

Im Gegensatz zu den statistisch motivierten Verfahren haben die differentialgeometrischen Verfahren vor allem eine Bedeutung als *Eckenfinder* erlangt. Prinzipiell könnten entsprechende Verfahren auch zur Detektion anderer Primitive entwickelt werden. Beim Dreschler-Nagel Operator wurde dies für Flecken angedeutet. Der Operator muß dabei relativ wenig modifiziert werden.

Das Operatorsignal der differentialgeometrischen Punktfinder besteht hauptsächlich aus Krümmungsmaßen, die sich alle auf Funktionen der ersten und zweiten Ableitungen der kontinuierlich approximierten Bildfunktion beziehen. Um als Ecke markiert zu werden, muß das Operatorsignal einen gewissen Schwellwert überschreiten. Bislang wenig befriedigend gelöst ist die Wahl der Schwellwerte, die meist individuell szenenabhängig festgelegt werden müssen. Guiducci (1988) sieht den Nachteil darin, daß die Verbindung der Schwellwerte zu signifikanten physikalischen Charakteristika der abgebildeten Szene unklar ist.

4.2.4 Punktfinder: Klassifizierende Verfahren

Die Klassifizierung markanter Punkte nach verschiedenen Punkttypen ist die wesentliche Eigenschaft der dritten Gruppe von Punktfindern. Hierzu werden Modelle für verschiedene Klassen von solchen Punkten (z. B. Ecken und Kreise) formuliert. Zu einem betrachteten lokalen Bildausschnitt werden durch Parameterschätzung die Modelle in Beziehung gebracht, wobei die Bildstruktur über die Grauwerte bzw. Grauwertgradienten in die Schätzprozesse eingeht. Die Übereinstimmung zwischen einem Bildausschnitt und einem Modell drückt sich in der Residuenquadratsumme der entsprechenden Schätzung aus und wird zur Klassifizierung herangezogen.

Unter den klassifizierenden Verfahren wollen wir zwei sehr verschiedene Arbeiten diskutieren. Der Operator von Förstner ist der viel flexiblere Punktfinder und gestattet die Klassifizierung der markanten Punkte in Ecken, kreisförmige Texturen (Kreise) und sonstige isotrope Strukturen. Innerhalb der Parameterschätzung wird die Position eines Punktes möglichst exakt bestimmt. Das von Rohr (1992) vorgestellte Verfahren konzentriert sich auf L -, T -, Y -, ...-Knoten. Ein parametrisches Modell der Intensitätsoberfläche solcher Ecktypen wird durch Superposition von elementaren Modellfunktionen formuliert und gestattet die Lokalisierung der exakten Eckposition und das Erkennen des entsprechenden Knotentyps. Dabei ist das Verfahren auf gute Vorinformationen angewiesen.

Paderes et al. 1984, Förstner und Gülch 1987

Die Entwicklung des Förstner'schen Operators erfolgte in zwei Stufen, die sich in jeweils zwei Verfahrensschritten des Punktfinders äußern. Der erste Schritt sieht die Berechnung einer 2×2 Matrix aus Varianzen³ und

³Wie schon bei der Darstellung des Operators von Moravec diskutiert, gilt auch hier, daß um von Varianzen und Kovarianzen sprechen zu können, die Erwartungswerte an die Gradienten Null sein müssen. Auf eine formale Unterscheidung zwischen geschätzten ($\hat{\sigma}^2$ aus (4-8)) und nicht geschätzten (σ^2 in (4-7)) Varianzen bzw. Kovarianzen wird bei der Darstellung verzichtet.

Kovarianzen der Bildgradienten in jedem Bildpunkt (x, y) vor

$$C(x, y) = \begin{pmatrix} \sigma_{I_x}^2 & \sigma_{I_x I_y} \\ \sigma_{I_x I_y} & \sigma_{I_y}^2 \end{pmatrix}, \quad (4-7)$$

die mit den Gradienten ∇I und mit einer symmetrischen, normalisierten Fensterfunktion ψ durch Faltung gemäß

$$C(x, y) = (\nabla I(x, y) \nabla I^T(x, y)) * \psi(x, y) \quad (4-8)$$

ermittelt werden können. In der Argumentation und bei Experimenten benützten Förstner und Gülch (1987) gleichgewichtete Fensterfunktionen, d. h. 2D-Rechteckfilter, so daß die Elemente des Operators (4-8) diskret durch lokale Summation der Gradientenquadrate bzw. -produkte

$$\frac{1}{N} \sum_W I_x^2, \quad \frac{1}{N} \sum_W I_x I_y, \quad \frac{1}{N} \sum_W I_y^2$$

berechnet werden. Summiert wird dabei über die N Bildpunkte innerhalb des Fensters W . Die Matrix (4-7, 4-8) wird auch als Matrix der 2. Momente bezeichnet und als Texturdeskriptor vielseitig verwendet (siehe z. B. die Referenzen in Lindeberg und Gårding, 1993).

Die Elemente von C sind translations- aber nicht rotationsinvariant, so daß der Übergang auf ein lokales Hauptachsensystem naheliegt. Die Eigenwerte werden aus der quadratischen Gleichung $\lambda^2 - (\sigma_{I_x}^2 + \sigma_{I_y}^2)\lambda + \sigma_{I_x}^2 \sigma_{I_y}^2 - (\sigma_{I_x I_y})^2 = 0$ erhalten und geben die maximale bzw. minimale Varianz an: $\sigma_{\nabla I, max}^2 = \lambda_{max}$, $\sigma_{\nabla I, min}^2 = \lambda_{min}$. Der Vergleich mit dem Interestoperator Moravec's macht deutlich, daß (4-7) als Generalisierung von (4-1, 4-2) aufgefaßt werden kann. Minimale und maximale gerichtete Varianz werden über die Eigenwerte direkt erhalten. Förstner bezeichnet seinen Operator ebenfalls als Interestoperator, was aufgrund der engen Verbindung zur Arbeit von Moravec, aber auch der statistischen Betrachtungsweise von Förstner (1991) naheliegend ist.

Als weitere Koordinatensystem invariante Merkmale können die Determinante und die Spur von C , aber auch deren Inversen C^{-1}

$$\begin{aligned} \det C &= \sigma_{I_x}^2 \sigma_{I_y}^2 - (\sigma_{I_x I_y})^2, & \text{trace } C &= \sigma_{I_x}^2 + \sigma_{I_y}^2, \\ \det C^{-1} &= 1/\det C, & \text{trace } C^{-1} &= \text{trace } C/\det C \end{aligned}$$

angegeben werden. Die Richtung zum größeren Eigenwert der Texturellipse C

$$\phi = \frac{1}{2} \arctan \frac{2\sigma_{I_x I_y}}{\sigma_{I_x}^2 - \sigma_{I_y}^2},$$

gibt eine aus dem lokalen Fensterausschnitt bestimmte mittlere Gradientenrichtung an, womit angedeutet ist, daß der Operator auch als Kantenoperator genutzt werden kann. Die Isotropie bzw. Anisotropie der lokalen Texturierung kann über das Verhältnis der Eigenwerte $\nu \in [1, \infty)$ bzw. das davon monoton abhängige Rundheitsmaß $q \in [0, 1]$

$$\nu = \frac{\lambda_1}{\lambda_2}, \quad q = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 = 4 \frac{\det C}{(\text{trace } C)^2}$$

berechnet werden. All diese Texturmerkmale liegen bildelementbezogen in Form von Merkmalskarten vor.

Die erste Stufe des von Förstner vorgeschlagenen Punktfinders beinhaltet folgende Schritte:

1. Bestimme das Operatorsignal $O_{F\ddot{o}}$ mit den beiden Merkmalen (w, q)

$$O_{F\ddot{o}}(x, y) = \begin{pmatrix} w(x, y) \\ q(x, y) \end{pmatrix} = \begin{pmatrix} \det C/\text{trace } C \\ 4\det C/(\text{trace } C)^2 \end{pmatrix}$$

2. **Detektion:** Selektiere mit den Schwellenwerten w_{min} und q_{min} alle Punkte (x, y) , für die das Operatorsignal das "Gewicht" $w > w_{min}$ und die Rundheit $q > q_{min}$ überschreitet. Die lokale Maxima-Suche unter allen verbleibenden Punkten mit dem Merkmal w führt schließlich zur Lokalisierung aller markanter Fenster mit isotroper Textur.

Das Merkmal $w = 1/\text{trace } C^{-1}$ steht in Beziehung zur inversen Texturellipse C^{-1} , die, wie wir noch sehen werden, ein Maß für die Genauigkeit der Lokalisierung angibt. Durch die Spur $\text{trace } C^{-1}$ wird ein mittleres Genauigkeitsmaß, folglich durch den Quotienten $1/\text{trace } C^{-1}$ ein Gewicht festgelegt. Das Rundheitsmaß q kann aufgrund der Beziehung zum Verhältnis der Eigenwerte ν nur Werte zwischen 0 und 1 annehmen, wobei isotrope Texturen, z. B. ein rotationssymmetrisches Muster, $\nu = 1 \mapsto q = 1$ und anisotrope Texturen, z. B. ein translationssymmetrisches Muster, $\nu \rightarrow \infty \mapsto q = 0$ hervorrufen.

Eine Besonderheit des Operators besteht darin, daß diese 1. Stufe nur zur Lokalisierung *markanter Fenster* führt. Die exakte Position des markanten Punktes im Fenster wird dann in der zweiten Stufe bestimmt. Diese erste Stufe des Förstner-Operators bezeichnen Harries (1987) und Noble (1988) offensichtlich in Unkenntnis der Förstner'schen Arbeit als Plessey-Eckendetektor. Sie nützen nicht die Rundheit q , sondern die beiden Eigenwerte von C und die Eckigkeit $\text{trace } C / \det C$, worin aber kein wirklicher Unterschied zum Operatorsignal $O_{F\delta}$ besteht.

Die zweite Stufe des Punktfinders beinhaltet eine Klassifikation der lokalen Bildstruktur zusammen mit der Lokalisierung des Schnittpunktes von Kantenelementen (bei Grauwertecken) bzw. Gefälleelementen (bei rotationssymmetrischen Figuren, Flecken) auf der Basis entsprechender Modelle für diese beiden Texturtypen.

Ausgangspunkt ist das Modell, in dem eine Konturlinie (Kante) in einem Bildpunkt als Isolinie mit konstantem Intensitätswert definiert ist:

$$I(x(s), y(s)) = \textit{konst} \quad \text{bzw.} \quad I_x dx/ds + I_y dy/ds = 0;$$

eine Hypothese, die schon bei der Ableitung des Operators von Kitchen und Rosenfeld (4-5, 4-6) zugrundegelegt wurde. Die Tangente an die Isolinie mit dem im Punkt (x, y) berechneten Normalenvektor $\nabla I / |\nabla I|$ kann als Geradengleichung in Hesse-Normalform

$$\frac{1}{|\nabla I|} \nabla I^T \begin{pmatrix} x \\ y \end{pmatrix} - l = 0$$

dargestellt werden⁴. $l = l(x, y)$ gibt den Abstand der Geraden durch den Punkt (x, y) vom (lokal gewählten) Koordinatenursprung an. Ein beliebiger Punkt (x_0, y_0) hat von dieser Geraden den Abstand v

$$v = \frac{1}{|\nabla I|} \nabla I^T \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - l \quad (4-9)$$

Die simple Idee besteht nun darin, Geradengleichungen für alle Bildpunkte innerhalb des Fensters gemäß (4-9) aufzustellen, wodurch ein parametrisches Modell für die Struktur "Grauwertecke" als lineares Modell mit den Unbekannten (x_0, y_0) formuliert ist. Bestimmt wird die Position des Punktes (x_0, y_0) , der von allen Geraden die kleinste gewichtete Abstandsquadratsumme $\mathbf{v}^T P \mathbf{v} \rightarrow \min$ aufweist. Die Eckposition ist damit als Knotenpunkt lokalisiert, in dem sich die Kantenstücke von zwei oder mehreren linear ausgerichteten Konturen schneiden. Die Gewichtung von l leitet Förstner aus $|\nabla I| = dI/dl$ unter der Annahme eines additiven weißen Rauschprozesses für die Bildintensitäten mit Varianz σ_{noise}^2 ab. Aus der Fehlerfortpflanzung resultiert⁵ $\sigma_l^2 = (1/|\nabla I|^2) \sigma_{noise}^2$, d. h. P ist eine Diagonalmatrix der Dimension $N \times N$ mit den Elementen $|\nabla I|^2$. Die Konturinformation im betrachteten Bildausschnitt, repräsentiert durch starke Gradienten, übt demzufolge einen großen Einfluß auf die geschätzte Punktposition (\hat{x}_0, \hat{y}_0) aus; rauschbedingte Strukturen hingegen beeinflussen das Ergebnis sehr viel weniger.

Alternativ zum Modell "Grauwertecke" (4-9) kann ein zweites Modell "rotationssymmetrische Struktur" formuliert werden, in dem der Schnittpunkt von Geraden in Gradientenrichtung (Gefälleelementen) geschätzt wird. Die Normale steht senkrecht auf der Gradientenrichtung, d. h. sie ist in Richtung $(-I_y, I_x)^T$ orientiert. Analog zu (4-9) läßt sich ein entsprechendes Modell aufstellen. Die beiden Ansätze sind nachfolgend summarisch angegeben:

⁴Um nicht weitere Symbole einführen zu müssen, hat hier (x, y) eine Doppelfunktion als beliebiger Punkt auf der Geraden und als betrachteter fester Bildpunkt.

⁵Streng wird die Beobachtung l aber aus $l = (1/|\nabla I|)(I_x x + I_y y)$ bestimmt. Die stochastischen Größen in dieser Gleichung sind die Bildgradienten, in denen sich das Rauschen der Bildfunktion bekanntlich zudem verstärkt. In $\sigma_l = (1/|\nabla I|) \sigma_{noise}^2$ wird nur die Positionsunsicherheit in Gradientenrichtung berücksichtigt. Damit wird der eigentliche Unsicherheitsterm, die Richtungsungenauigkeit, hier nicht stochastisch angenommen ist (siehe auch Förstner, 1991, S. 62).

1. Modell "Grauwertecke": Schnittpunkt von Geraden in Konturrichtung

funktionales Modell: formal $E(\mathbf{l}) = A\mathbf{p}$

$$l(x, y) + v(x, y) = \frac{1}{|\nabla I|}(I_x \hat{x}_0 + I_y \hat{y}_0) \quad \text{mit} \quad l(x, y) = \frac{1}{|\nabla I|}(I_x x + I_y y)$$

stochastisches Modell: formal $D(\mathbf{l}) = \sigma_{noise}^2 P^{-1}$

$$p_l(x, y) = |\nabla I|^2, \quad P = \text{diag}(p_l)$$

Normalgleichungen: formal $C\hat{\mathbf{p}} = \mathbf{h}$

$$\begin{pmatrix} \sum_W I_x^2 & \sum_W I_x I_y \\ \sum_W I_x I_y & \sum_W I_y^2 \end{pmatrix} \begin{pmatrix} \hat{x}_0 \\ \hat{y}_0 \end{pmatrix} = \begin{pmatrix} \sum_W I_x^2 x + \sum_W I_x I_y y \\ \sum_W I_x I_y x + \sum_W I_y^2 y \end{pmatrix}$$

Geschätzte Residuenquadratsumme: $\mathbf{v}^T P \mathbf{v} = \mathbf{l}^T P \mathbf{l} - \hat{\mathbf{p}}^T \mathbf{h}$

$$\Omega_E = \mathbf{v}^T P \mathbf{v} = \sum_W (I_x x + I_y y)^2 - \hat{\mathbf{p}}^T \mathbf{h} \quad (4-10)$$

2. Modell "rotationssymmetrische Struktur": Schnittpunkt von Geraden in Gradientenrichtung

funktionales Modell:

$$l(x, y) + v(x, y) = \frac{1}{|\nabla I|}(-I_y \hat{x}_0 + I_x \hat{y}_0) \quad \text{mit} \quad l(x, y) = \frac{1}{|\nabla I|}(-I_y x + I_x y)$$

stochastisches Modell:

$$p_l(x, y) = |\nabla I|^2, \quad P = \text{diag}(p_l)$$

Normalgleichungen:

$$\begin{pmatrix} \sum_W I_y^2 & -\sum_W I_x I_y \\ -\sum_W I_x I_y & \sum_W I_x^2 \end{pmatrix} \begin{pmatrix} \hat{x}_0 \\ \hat{y}_0 \end{pmatrix} = \begin{pmatrix} \sum_W I_y^2 x - \sum_W I_x I_y y \\ -\sum_W I_x I_y x + \sum_W I_x^2 y \end{pmatrix}$$

Residuenquadratsumme:

$$\Omega_K = \mathbf{v}^T P \mathbf{v} = \sum_W (I_x y - I_y x)^2 - \hat{\mathbf{p}}^T \mathbf{h} \quad (4-11)$$

Die Normalgleichungsmatrix C von (4-10) ist identisch mit der Matrix der 2. Momente (4-7, 4-8), so daß die enge Beziehung von statistischer Texturbeschreibung und geometrisch-parametrischer Schätzung der Position des markanten Punktes aufgezeigt ist.

Bei der Darstellung in (4-10) und (4-11) wurde die eingängigere Form der Summation über das Fenster bevorzugt. Die Verallgemeinerung gemäß (4-8) auf eine beliebige symmetrische normalisierte Fensterfunktion soll aber noch kurz angegeben werden. Sie folgt direkt daraus, daß die Summation über das entsprechende Argument durch die Faltung mit ψ entsprechend (4-8) ersetzt wird: $\sum_W(\cdot) \rightarrow \psi * (\cdot)$. Für die Normalgleichungen in (4-10) führt das zu

$$(\psi * (\nabla I \nabla I^T)) \begin{pmatrix} \hat{x}_0 \\ \hat{y}_0 \end{pmatrix} = \psi * \left((\nabla I \nabla I^T) \begin{pmatrix} x \\ y \end{pmatrix} \right)$$

Dieser Übergang auf ψ , für den insbesondere die Gaußfunktion von Interesse ist, kann direkt als Überlagerung der Gewichtung P im Fenster mit ψ interpretiert werden. Entsprechend den diskreten Werten des Gaußkerns werden die zentralen Pixel höher, die Randpixel geringer gewichtet.

Die Inversen der Normalgleichungsmatrizen (4-10, 4-11) geben die Genauigkeit der geschätzten Position an, so daß hierin die Begründung für das Merkmal w des Operators zu sehen ist. Durch hohe Gewichte w sind demzufolge markante Fenster gekennzeichnet, die wiederum eine sehr präzise Bestimmung der Position erlauben. In Förstner (1991) werden beide Modelle zusätzlich als "gewogene Schwerpunktsformulierungen" dargestellt.

| Fenster- größe $N =$ $f =$ | Fraktile $F_{f,f,1-\alpha}$ | | | | | | | | |
|----------------------------------|-----------------------------|--------------|--------------|--------------|--------------|--------------|----------------|----------------|----------------|
| | 4×4 | 5×5 | 6×6 | 7×7 | 8×8 | 9×9 | 10×10 | 11×11 | 13×13 |
| | 14 | 23 | 34 | 47 | 62 | 79 | 98 | 119 | 167 |
| $1 - \alpha = 99\%$ | 3.70 | 2.72 | 2.26 | 1.99 | 1.82 | 1.70 | 1.61 | 1.54 | 1.44 |
| $1 - \alpha = 99.9\%$ | 5.93 | 3.85 | 2.98 | 2.51 | 2.22 | 2.02 | 1.88 | 1.77 | 1.62 |

Tabelle 4.1: Schwellwert für die Klassifikation: Die Fraktile der Fisher-Verteilung $F_{f,f,1-\alpha}$ bei $f = N - 2$ Freiheitsgraden

Dies beruht auf der Dualität der Geradendarstellung mit stochastisch modellierter Beobachtung, die z. B. in Hesse-Normalform gemäß

$$l = (1/|\nabla I|)(I_x x + I_y y) \quad \text{mit Gewicht} \quad |\nabla I|^2$$

oder äquivalent als Punkt mit singulärer 2×2 Gewichtsmatrix gemäß

$$\begin{pmatrix} x \\ y \end{pmatrix} \quad \text{mit Gewicht} \quad \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \quad (4-12)$$

angegeben werden kann. Was hier exemplarisch für die Geraden in Konturrichtung aufgezeigt wurde, gilt entsprechend auch für Geraden in Gradientenrichtung.

Die Residuenquadratsummen der beiden Modelle können zur Klassifikation herangezogen werden. Nach Förstner folgt die Testgröße

$$T = \frac{\Omega_E}{\Omega_K} \sim F_{f,f}$$

einer zentralen Fisher-Verteilung, d. h. die Zufallsvariablen Ω_E und Ω_K sind stochastisch unabhängig und besitzen jeweils f Freiheitsgrade. $f = N - 2$ gibt die Redundanz im linearen System bei N Beobachtungen und 2 Unbekannten an. Die Klassifikation erfolgt nun gemäß

$$\begin{array}{ll} T < 1/F_{f,f,1-\alpha} & \text{Hypothese "Ecke" bestätigt} \\ T > F_{f,f,1-\alpha} & \text{Hypothese "Kreis" bestätigt} \\ \text{sonst} & \text{Klassifiziert als allgemeine isotrope Textur} \end{array} \quad (4-13)$$

Die Fraktile der Fisher-Verteilung zu Signifikanzniveau $1 - \alpha$ kann aus Tafelwerken herausgegriffen, oder sehr effizient mit der in Hahn et al. (1989) vorgestellten Prozedur berechnet werden. Für quadratische Fenster der Größe $N = 4 \times 4$ bis 13×13 sind die Fraktile $F_{f,f,1-\alpha}$ der Fisher-Verteilung bei $f = N - 2$ für die Signifikanzniveaus $1 - \alpha = 99\%$ und 99.9% in Tabelle 4.1 beispielhaft aufgelistet.

Damit kann die zweite Stufe des Punktefinders mit den Schritten *Klassifikation* und *Punktlokalisierung* folgendermaßen angegeben werden:

- Klassifikation:** Klassifiziere mit der Testgröße $T = \Omega_E/\Omega_K$ gemäß (4-13) die Textur in (A) Modell: Ecke, (B) Modell: Kreis, oder (C) sonstige isotrope Textur
- Punktlokalisierung:** Schätze abhängig vom Klassifikationsresultat die Position des markanten Punktes: für das Modell Ecke gemäß (4-10), für das Modell Kreis gemäß (4-11).

Für allgemeine isotrope Texturen, die nicht als Ecke oder Kreis klassifiziert werden können, liegt kein Modell zur Positionsschätzung vor, so daß als Heuristik die Fenstermitte oder der in beiden Koordinatenrichtungen unabhängig durch Gewichtung mit den Gradientenquadraten bestimmte Schwerpunkt berechnet werden kann (Ackermann, 1984). Eine weitere Möglichkeit besteht darin, die Fraktile auf 1 zu setzen, womit die kleinere

Residuenquadratsumme das als wahrscheinlicher erachtete Modell angibt und die Parameterschätzung nach diesem Modell erfolgt. Daß dabei die Interpretation sehr unzuverlässig erfolgt ist offensichtlich. Auf eine generelle Diskussion des Zuverlässigkeitskonzepts, wie es von Baarda (1968) begründet wurde, wollen wir in dieser Arbeit verzichten. Die Zuverlässigkeit behandelt die Fragestellung, wie gut die Beobachtungen innerhalb der Parameterschätzung kontrolliert sind, ab welcher Größe einzelne grobe Fehler mit einer gewissen Mindestsicherheit (Güte des statistischen Tests) aufgedeckt werden können (innere Zuverlässigkeit) und in welchem Maße ein solcher unentdeckter grober Fehler die zu schätzenden Parameter und daraus ableitbare Größen verfälscht (äußere Zuverlässigkeit). Die Formulierung der Zuverlässigkeit für Parametergruppen (z. B. nur die zu koordinierenden Punkte, nicht aber mitzuschätzende Orientierungen oder andere Parameter) ist in Hahn und Jäger (1987) ausgearbeitet.

Die Erweiterung der Baarda'schen Theorie zur Abstimmung von Hypothesentests zwischen Nullhypothese (z. B. in (4-13) die Hypothese: "allgemeine isotrope Textur") und Alternativhypothesen (in (4-13) die Hypothesen $H_{a,1}$: "Ecke" und $H_{a,2}$: "Kreis") ist für allgemeine $F_{m,n}$ -verteilte Teststatistiken in Hahn et al. (1989, S. 234-248) sehr detailliert präsentiert worden. Die Abstimmungstheorie beruht auf der Grundüberlegung, daß Hypothesentests so formuliert werden müssen, daß sie zu einer widerspruchsfreien Klassifizierung führen. Einer der klassischen Widersprüche (Baarda, 1968) ist der, daß die Nullhypothese "alle Beobachtungen sind mit der Modellierung konsistent" bestätigt wird, die Einzelprüfung der Beobachtungen aber einzelne Fehler anzeigt. In Hahn et al. (1991, S. 15-26) sind Anwendungen zur Zuverlässigkeits- bzw. Abstimmungstheorie aufgezeigt, wobei auch ein Beispiel zur Klassifikation mit dem Förstner'schen Punktoperator diskutiert wurde.

Rohr 1990, 1992

Mit dem Ziel Ecken in Bildern auch hinsichtlich des Knotentyps zu erkennen, formuliert Rohr parametrische Modelle einer Klasse von charakteristischen Intensitätsoberflächen, die sich durch Superposition aus elementaren Modellen zusammensetzen. Die Bestimmung der Modellparameter zielt auf eine detaillierte Charakterisierung der Grauwertstruktur im lokalen Bildausschnitt. Das elementare Modell einer Ecke besteht aus einer keilförmigen Struktur (siehe Abbildung 4.2), auch als ideale Ecke oder Stufenkeil bezeichnet, die mit einem Gaußkern gefiltert

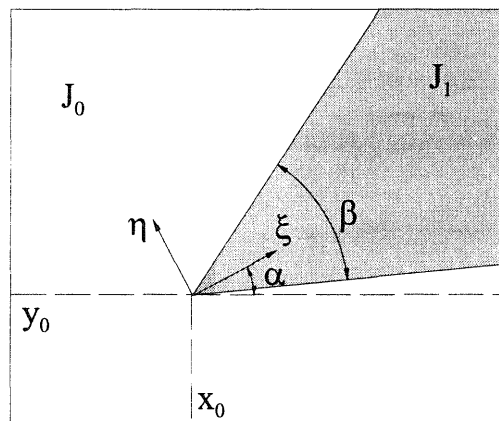


Abbildung 4.2: Keilförmige Struktur (ideale Ecke)

wird. Die ideale Grauwertecke soll im lokalen Koordinatensystem (ξ, η) durch den Öffnungswinkel β und die Amplitude $I_a = I_1 - I_0$ gemäß

$$E(\xi, \eta) = \begin{cases} I_a & \text{falls } \xi \geq 0 \wedge |\eta| \leq \xi \tan(\beta/2) \\ 0 & \text{sonst} \end{cases}$$

dargestellt werden. Analytisch kann das elementare Eckenmodell durch Faltung der idealen Grauwertecke mit einem Gaußfilter $G(\xi, \eta)$

$$I_E(\xi, \eta) = E(\xi, \eta) * G(\xi, \eta)$$

explizit ermittelt werden (z. B. Berzins, 1984, Rohr, 1992). Das Modell im lokalen System wird dann durch die Parameter (β, I_a, σ) beschrieben. Der Übergang vom lokalen zum globalen Bildkoordinatensystem erfolgt durch Rotation mit Winkel α und Translation mit (x_0, y_0) , wobei (x_0, y_0) die Position der Ecke im Bild angibt. Das elementare Eckenmodell ist damit durch sechs Parameter

$$I_E(x, y; p) = I_E(x, y; x_0, y_0, \alpha, \beta, I_a, \sigma) \quad (4-14)$$

festgelegt. Wird das Intensitätsniveau des Hintergrunds I_0 (siehe Abbildung 4.2) dem Elementarmodell I_E überlagert und als weiterer Parameter hinzugefügt, so ist die L -Ecke durch sieben Parameter beschrieben. In dieser oder ähnlicher Art und Weise haben auch Guiducci (1988), Deriche und Giraudon (1990), Giraudon und Deriche (1991) und De Micheli et al. (1989) entsprechende Modelle aufgestellt.

Die Generalisierung auf allgemeine Knotentypen hat Rohr durch Überlagerung von m keilförmigen Grauwertstrukturen gemäß

$$I_m(x, y) = \left(\sum_{i=1}^m E_i(x, y) \right) * G(x, y) = \sum_{i=1}^m (E_i(x, y) * G(x, y)) \quad (4-15)$$

vorgenommen, wobei wegen Gültigkeit des Superpositionsprinzips (die Faltung mit der Gaußfunktion ist eine lineare Operation) dies äquivalent zur Überlagerung entsprechender elementarer Eckenmodelle ist. $m = 2$ ergibt einen L -Knoten, in dem der Grauwertverlauf als Überlagerung zweier komplementärer Modellfunktionen mit den Öffnungswinkeln β und $360^\circ - \beta$ aufgefaßt werden kann. Bei $m = 3$ wird der charakteristische Grauwertverlauf von drei aneinanderstoßenden Flächen durch die Überlagerung der entsprechenden Modellfunktionen modelliert. Dazu sind zwei weitere Parameter (β_2, I_2) zum Parametervektor hinzuzufügen, so daß allgemein bei m aneinandergrenzenden Flächen $3 + 2m$ Parameter auftreten. T -, Y -, oder $Pfeil$ -Verbindungen ergeben sich aus der speziellen Parameterwahl von $I_{m=3}$, wobei die Öffnungswinkel $\beta_1 + \beta_2$ gleich, größer bzw. kleiner als 180° sein müssen. Eine entsprechende Klassifizierung der Knotentypen nach dem Waltz'schen Schema (Waltz, 1975) kann auch für $m > 3$ angegeben werden (vgl. auch Rohr, 1990).

Der Operator zur Identifikation von Grauwertecken unterschiedlichen Knotentyps einschließlich der Detektion und Lokalisierung besteht aus folgenden Schritten:

1. Detektiere Punkte hoher Grauwertvariation mit einem statistischen Verfahren oder mit $\det H$ (4-5).
2. Bestimme gerade Kantenstücke innerhalb einer lokalen Umgebung um diesen Punkt und generiere abhängig von der Anzahl dieser Kanten das entsprechende parametrische Modell (4-15). Durch die Konturdetektion liegen Näherungswerte für die Winkel α_i, β_i des Modells vor. Die Approximation für die Intensitätsniveaus kann durch Mittelung der Intensitäten in den einzelnen Sektoren gewonnen werden.
3. Schätze die Parameter \mathbf{p} durch Minimierung von

$$\Omega = \mathbf{v}^T \mathbf{v} \rightarrow \min \quad \text{mit} \quad v_i = I(\mathbf{x}_i) - I_m(\mathbf{x}_i; \hat{\mathbf{p}})$$

Die Diskrepanz zwischen Bild und Modell kommt in Ω zum Ausdruck und führt bei Übereinstimmung zur Identifikation des Grauwertverlaufs mit der Modellklasse. Durch die Schätzung der Parameter liegen die Position der Ecke, aber auch alle weiteren Parameter zur Charakterisierung des Knotens vor, so daß sie zur Identifikation des Knotentyps herangezogen werden können.

Das Elementarmodell hat Rohr (1990) durch numerische Integration berechnet; zur Schätzung der Parameter wurde ein Gradientenabstiegsverfahren verwendet. Eine wesentliche Verbesserung vor allem bezüglich der Rechenzeiten (von Stunden auf Minuten) konnte in Rohr (1992) dadurch erzielt werden, daß zur Berechnung des Elementarmodells die Gauß'sche Dichte und Verteilungsfunktion approximiert und eine explizite Linearisierung von $I_m(\mathbf{x}; \mathbf{p})$ nach den Parametern vorgenommen wurde. Der weitere Lösungsweg besteht dann in der iterativen Schätzung der unbekannt Parameter. Um den Grauwertverlauf hinreichend hoch aufzulösen, muß die Fenstergröße hinreichend groß sein, wobei Rohr in Beispielen 20×20 und 25×25 Bildausschnitte verwendet hat.

Ganz offensichtlich liegt der Schwerpunkt der Arbeit von Rohr in der Interpretation der Ecke durch die komplette Charakterisierung der Grauwertstruktur durch die geschätzten Parameter einer Modellklasse. Die geschätzten Parameter beinhalten auch die Position der Ecke, die dadurch präzise lokalisiert werden kann. Daß der lokalisierte Eckpunkt durch ein topologisches Attribut (Knotentyp) wesentlich an Semantik gewinnt, kann für nachfolgende 3D-Interpretationsaufgaben von großem Nutzen sein. Die eigentliche Schwäche des Verfahrens besteht im zweiten Schritt, der eine Bildsegmentierung voraussetzt. Der Punktfinder könnte wegen der für die Auswahl des Modells notwendigen Bildsegmentierung auch der Gruppe der indirekten, segmentierungsbasierten Verfahren zugerechnet werden. Zeichnen sich Konturen bei der Segmentierung nicht hinreichend deutlich ab, besteht die Gefahr, daß ein falscher Knotentyp modelliert wird. Im günstigsten Fall wird das Modell im 3. Schritt des Punktfinders verworfen, so daß zwar ein markanter Punkt (aufgrund Schritt 1) detektiert, dieser aber nicht als Ecke mit entsprechendem topologischen Attribut klassifiziert werden kann.

Ergänzungen

Das von Rohr vorgestellte Verfahren besteht hinsichtlich des 3. Schrittes aus einer Zuordnungsaufgabe von Bild und parametrischem Modell einer Klasse von Grauwertecken. Die geometrische Transformation zwischen Modell und Bild beinhaltet eine Rotation und Translation in der Bildebene. Weitere Parameter wie die Öffnungswinkel der Stufenkeile dienen der Beschreibung der Struktur. Die Intensitätsparameter können einer radiometrischen Modellierung zugerechnet werden und die Glättung der idealen Ecken dient der Anpassung an die bandpaßgefilterte Aufnahme.

In Experimenten hat Rohr (1992) beobachtet, daß die geschätzten Modellparameter nur wenig von einer Glättung des Bildes mit einem Gaußfilter beeinflusst werden, sieht man vom Modellparameter σ ab, der erwartungsgemäß um diese additive Glättung anwächst. Die Glättung des Bildes zur Unterdrückung von Rauschen wirkt sich erwartungsgemäß positiv auf die Schätzprozedur aus. Die weitgehende Unabhängigkeit der geschätzten Parameter von der Glättung des Bildes gilt allerdings nicht, wenn Modell und Bild nicht konsistent sind. In der Glättung des Bildes mit verschiedenen Gaußkernen sieht Rohr deshalb eine weitere Möglichkeit (neben der Residuenquadratsumme), um die Gültigkeit des identifizierten Modells zu prüfen.

Eine starke Einschränkung gegenüber manchen segmentierungsbasierten Verfahren liegt in der Grundannahme, daß die Bildintensität innerhalb von Regionen oder entlang von Konturen konstant sein muß. Bei Rohr werden konstante Niveaus innerhalb eines Keils direkt parametrisch modelliert. Bei Förstner haben wir gezeigt, daß die Konturrichtung durch $I(x(s), y(s)) = konst$ festgelegt wurde und Ausgangspunkt der geometrischen Modellbildung für die Modelle 'Grauwertecke' und 'Kreisförmige Struktur' war. In welchem Maße sich Modellverletzungen auswirken, soll an einfachen Beispielen demonstriert werden. In Abbildung 4.3 sind 6 Bildausschnitte dargestellt, die mit dem Förstner'schen Punktfinder bearbeitet wurden. Die erste Reihe zeigt einen L -Knoten, einen Y -Knoten und eine kreisförmige Bildstruktur, die jeweils exakt den Modellvoraussetzungen entsprechen. Die geschätzte Position führt erwartungsgemäß zur korrekten Lage des markanten Punktes. In der zweiten Reihe sind die Strukturen wieder aufgegriffen, wobei allerdings in zumindest einem Teilbereich des jeweiligen Bildausschnittes das konstante Intensitätsniveau durch eine geneigte Fläche ersetzt wurde. Die über die Bildgradienten bestimmte lokale Kantenrichtung weicht vom Konturverlauf beträchtlich ab, was zusammen mit den Gradienten $\nabla I \neq 0$ innerhalb der geneigten Fläche zu systematischen Positionsfehlern von mehreren Pixeln führt (vergl. Abbildung 4.3). Eine Abwandlung des Förstner-Operators zur Lokalisierung von 2.5 D-Ecken (entsprechend dem mittleren Bild in der zweiten Reihe von Abbildung 4.3), wie sie z. B. in Distanzbildern häufiger vorkommen, ist in Hahn (1994) vorgeschlagen worden.

4.2.5 Weitere Verfahren und abschließende Bewertung

Als weitere Verfahren, die nicht in das obige Klassifizierungsschema eingeordnet werden können, wollen wir den als Monotonieoperator bezeichneten Punktfinder kurz darstellen und einen zweiten Operator besprechen, der zur Schätzung des Expansionspunktes (FOE) entwickelt wurde. Der FOE-Punktfinder hat schon formal viele Parallelen zu dem von Förstner entwickelten Punktfinder, obgleich er nicht wie alle bisher besprochenen Operatoren ein Monobild prozessiert, sondern auf einem Raumzeitbild aufsetzt. Für eine weitere Richtung von Arbeiten stehen Rangarajan et al. (1989) und Mehrotra et al. (1990). Die Eckendetektion beruht hier auf der

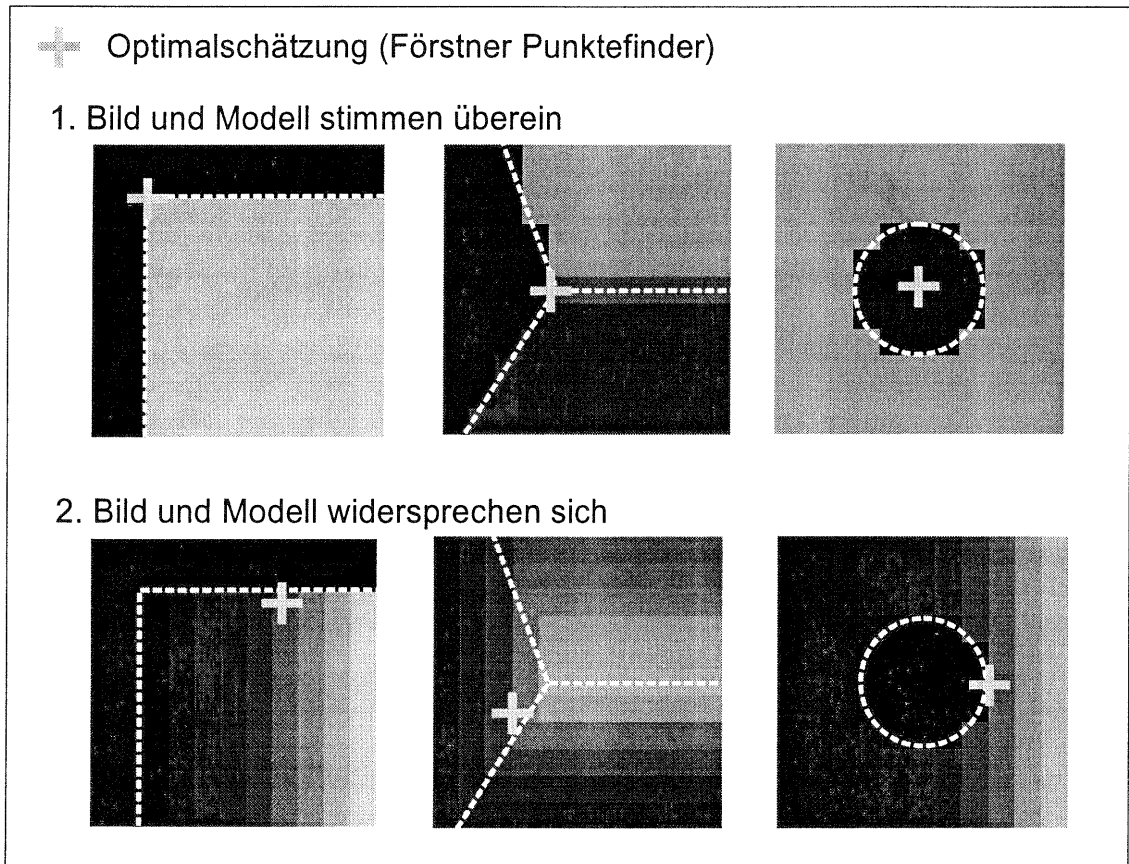


Abbildung 4.3: 1. Reihe: L -Knoten, Y -Knoten, kreisförmige Bildstruktur, 2. Reihe: Grauwertkeil vor konstantem Hintergrund, Y -Knoten als 2.5 D-Ecke, Scheibe vor Hintergrund mit Lichtabfall

Faltung mit einer Serie diskret entworfener Detektorenmasken für Ecken vom Typ L -Knoten, die dann mit einem hohen Operatorsignal antworten, wenn der Öffnungswinkel des L -Knotens β und der Drehwinkel α (Abbildung 4.2) von Maske und Bild gut übereinstimmen. Die Vorgehensweise entspricht der üblicher Kantendetektionsverfahren, in denen zur Kantendetektion das Bild mit diskreten Ableitungsmasken gefiltert wird.

Monotonieoperator

Ein in mehrfacher Hinsicht besonders einfacher Operator ist der von Kories (1985), Zimmermann und Kories (1984) vorgeschlagene Monotonieoperator, dessen Operatorsignal durch einfache Vergleichs- und Summationsoperationen äußerst effizient berechnet werden kann. Der Operator verdient, obwohl er im eigentlichen Sinn kein Punktefinder ist, aus experimenteller Sicht Beachtung, da er insbesondere bei der Schätzung von Verschiebungsvektoren erfolgreich erprobt wurde (Kories und Zimmermann, 1986).

Formal ist der Operator gemäß

$$O_{\text{Monotonie}} = \sum_w b(i, k) \quad \text{mit} \quad b(i, k) = \begin{cases} 1 & : I(x, y) > I(x+i, y+k) \\ 0 & : \text{sonst} \end{cases} \quad (4-16)$$

definiert. Die Summation über ein lokales Fenster von N Bildpunkten bildet die Bildinformation auf einen Merkmalsvektor mit Werten von 0 bis N ab. Für eine Fenstergröße von 3×3 Bildelementen ergibt sich für ein lokales Maximum (Punkt) ein Wert von 8, bei einem geraden Kantenstück ein Wert von 3 und bei einer (ein Pixel breiten) Linie ein Wert von 6. Das (konstante) Intensitätsniveau der Struktur wird hierzu höher als der Hintergrund angenommen. Neben dem geringen Rechenaufwand ergeben sich noch weitere Vorteile des Monotonieoperators: die Merkmale sind unabhängig von absoluten Intensitäten, auch skalierungs- und

rotationsinvariant. Die Klassen tendieren dazu Cluster zu bilden. Die Stabilität der Cluster ist jedoch stark vom Rauschen beeinträchtigt. Kories (1985) filtert deshalb die Bilder vorab mit einem Bandpaß, was zur Folge hat, daß das Abtastschema des Monotonieoperators angepaßt werden muß. Da die kurzen Wellenlängen infolge der Filterung eliminiert wurden, verwendet Kories acht Pixel in einem dem Glättungsradius des Filters angepaßten größeren Abstand.

In Clustern von mindestens zehn zusammenhängenden Bildelementen einer Klasse werden Schwerpunkt, Größe und Operatorresultat als Charakteristika dieses Flecks ermittelt. Als besonders stabil haben sich in Bildfolgen Flecken der Klassen 0 und 8 erwiesen. Mit den Schwerpunkten konnten selbst bei leichten Größen- und Formänderungen der Flecken dichte Verschiebungsvektorfelder mit sehr befriedigenden Resultaten bestimmt werden (siehe auch Enkelmann et al., 1986). Allerdings darf auch nicht übersehen werden, daß die Resultate in sehr stark strukturierten Szenen erzielt wurden. In strukturschwachen Gebieten täuschen leichte Grauwertvariationen Strukturen vor, die für die Zuordnung meist nicht geeignet sind.

Negahdaripour und Horn, 1987

Den Expansions- oder Kontraktionspunkt haben wir im Abschnitt 2.3.3 eingeführt als Punkt in der Bildebene, der die momentane translatorische Bewegungsrichtung des Sensors angibt. Mit dem Beobachtungsmodell für den optischen Fluß, das Konstanz der Intensitäten innerhalb eines Zeitintervalls δt zugrundelegt, ergibt sich für die starre Szene gemäß (3-31)

$$I_t + \nabla I^T L_V \omega + \nabla I^T L_Z \frac{1}{Z} d_t \mathbf{T} = 0.$$

Ist die rotatorische Komponente der Bewegung bekannt, z. B. vorab bestimmt, oder handelt es sich um eine rein translatorische Sensorbewegung, kann der rotatorische Effekt eliminiert werden

$$I_t + \nabla I^T L_Z \frac{1}{Z} d_t \mathbf{T} = 0.$$

Negahdaripour und Horn diskutieren zur Bestimmung des Expansionspunktes zwei Verfahren, wobei der attraktivere Lösungsweg mit sog. stationären Punkten argumentiert. Als stationäre Punkte werden Bildpunkte bezeichnet, deren Intensität sich innerhalb des Zeitintervalls δt nicht ändert, d. h. $I_t = 0$. Dabei sind solche Bildpunkte ausgeschlossen, die in Regionen konstanter Helligkeit liegen ($\nabla I \neq 0$). Auch Punkte, deren räumliche Entfernung sehr groß ist, sind auszuschließen, da sie praktisch keine feststellbaren Verschiebungen in der Bildebene hervorrufen. Die stationären Punkte sind nach dieser Definition Bildpunkte, die auf Konturlinien konstanter Bildintensität liegen. Aufgrund von $\omega = 0$ ist jede der Konturen radial auf den FOE zu ausgerichtet. Damit haben wir für die Lokalisierung des FOE mit den stationären Punkten exakt die Voraussetzungen, die der Lokalisierung von Ecken in (4-10) zugrunde gelegt wurde. Der genau gleiche Schätzprozeß, wie ihn Förstner für Ecken angibt, wurde von Negahdaripour und Horn (1989) unabhängig vorgeschlagen. Mit $I_t = 0$ und $1/Z \neq 0$ ergibt sich $\nabla I^T L_Z d_t \mathbf{T} = 0$. Wird die Bewegungsrichtung $d_t \mathbf{T}$ im Bildkoordinatensystem durch $(x_0, y_0, c)^T$ angegeben, führt dies für jeden stationären Punkt zu einer Gleichung der Form

$$I_x x + I_y y = I_x x_0 + I_y y_0,$$

aus der die unbekannte Position des FOE durch Lösung des überbestimmten Gleichungssystems entsprechend (4-10) direkt bestimmt werden kann. Auf ein zweites Verfahren, das zusätzlich davon Gebrauch macht, daß die Tiefe zu jedem abgebildeten Punkt positiv sein muß, sei hier lediglich verwiesen.

Zur simultanen Bestimmung von Rotationen schlagen Negahdaripour und Horn in Anlehnung an Prazdny (1981) vor, den dreidimensionalen Suchraum für ω regelmäßig aufzurastern (mit bekannt vorausgesetzten Obergrenzen $|\omega_i| \leq |\omega_{i,max}|$) und für jeden Rasterpunkt in diesem Suchraum eine FOE-Schätzung vorzunehmen. Die minimalen Residuenquadratsumme wird dann zur Auswahl der besten Schätzung in diesem Suchraum herangezogen.

Haala (1990) hat Experimente zur Schätzung des Expansionspunktes mit einer dichten Bildsequenz angestellt, bei deren Aufnahme der Sensor ausschließlich translatorisch bewegt wurde. Die experimentell verifizierte Expansionspunktbestimmung mit dem oben vorgestellten Ansatz hat in den günstigsten Fällen, in denen von den stationären Punkten mit der größten Gradientenstärke ca. 300 verwendet wurden, zu einer Genauigkeit besser

als 10 Bildelemente geführt, was für viele Anwendungen ein befriedigendes Resultat darstellt.

Abschließende Bewertung

Die statistischen Punktfinder berechnen sehr einfache Merkmale der Bildfunktion, die aber meist keine sehr leicht interpretierbaren Punkte sind. Daß sie sich dennoch potentiell zur Verschiebungsschätzung eignen, liegt daran, daß mit diesen Verfahren vorwiegend isotrope Texturen detektiert werden. Die differentialgeometrischen Punktfinder zielen vorwiegend auf einen ganz bestimmten Punkttyp eines markanten Punktes, in der Regel auf Grauwertecken. Hier ist die Interpretation schon in der Modellierung gegeben. Daß in diesen Punkten eine lokale zweidimensionale Verschiebungsschätzung gelingt, wurde insbesondere von Nagel (1983) theoretisch nachgewiesen. Die klassifizierenden Punktfinder vereinen die statistische Betrachtung mit einer geometrischen Modellierung der Intensitätsfunktion, die zur Interpretation des lokalen Bildausschnittes herangezogen wird. Da hierbei eine Optimalpunktschätzung für die markanten Punkte erfolgt, kann von diesen Verfahren ein hohes Genauigkeitpotential erwartet werden. In den nachfolgenden Experimenten wird deutlich, daß bei der Punktverfolgung über längere Abschnitte der Sequenz die Genauigkeit meist befriedigend, die Stabilität der markanten Punkte sehr kritisch ist. Die Ursache hierfür werden wir in den Experimenten zur Punktverfolgung aufzeigen und Möglichkeiten zur Verbesserung der Stabilität diskutieren.

4.3 Merkmalsbasierte Punktverfolgung - Experimentelle Untersuchung

In den Experimenten zur Punktextraktion und -verfolgung soll der von Förstner entwickelte Punktfinder verwendet werden. Die Diskussion der verschiedenen Verfahren zur Extraktion markanter Punkte hat gezeigt, daß aus theoretischer Sicht der Förstner'sche Operator durch die Klassifizierung der lokalen Bildstruktur mehr bietet als die statistischen Verfahren. Da neben Ecken auch kreisförmige Texturen und sonstige isotrop texturierte markante Punkte extrahiert werden, soll dem Operator auch der Vorzug gegenüber den auf Ecken beschränkten differentialgeometrischen Verfahren gegeben werden.

Die Verfahren zur Punktverfolgung beruhen auf den in Abschnitt 1.5 vorgestellten Heuristiken der merkmalsbasierten Zuordnung. Insbesondere wird Kompatibilität gefordert, d. h. Punkte in aufeinanderfolgenden Bildern können zugeordnet werden, wenn ihre Merkmale einander ähnlich sind. Diese Heuristik stützt sich auf die Invarianz der beiden Merkmale Rundheit und Gewicht eines markanten Punktes gegen Rotation und Translation sowie gegen eine radiometrische Verschiebung der Intensitätsfunktion. Eine radiometrische Skalierung wirkt sich linear auf das Gewicht, nicht aber auf die Rundheit aus.

Um die Problematik der Bewegungssegmentierung im Bild zu umgehen, wird von der Kontinuitätsheuristik (räumlich und temporal sind die Verschiebungsvektoren in einer lokalen Umgebung ähnlich, was durch Regularisierung oder analytische Approximation eines Verschiebungsvektorfeldes im Bildraum eingebracht werden kann) lediglich in abgeschwächter Form durch Einschränkung des Suchbereiches Gebrauch gemacht. Das Prinzip hierzu ist in Abbildung 4.4 skizziert. Dem Expansionspunkt (FOE) kommt dabei eine besondere Rolle zu (siehe auch Abschnitt 2.3.3). Wird der Anteil der Sensorrotation aus der projizierten Bewegung (2-39) eliminiert, bewegen sich die Punkte radial vom Expansionspunkt weg. Die Größe der Bewegung eines Punktes ist invers proportional zu dessen Tiefe und hängt ferner von der Lage des Punktes im Bild und von der translatorischen Bewegung des Sensors ab. In den dargestellten Suchbereichen führt das mangelnde Vorwissen über die Entfernung eines Punktes zu einer entsprechend großen radialen Ausdehnung. Die kleinere Querkomponente spiegelt alle anderen Unsicherheiten wieder, wie z. B. Einflüsse einer unzulänglich approximierten rotatorischen Bewegung. Dargestellt in Abb. 4.4 sind der FOE und dessen Lageunsicherheit, die Position der markanten Punkte P_i im vorhergehenden Bild und deren Prädiktion P_i^* in das aktuell betrachtete Bild. Die in diesem Bild extrahierten Punkte werden zur Zuordnung in Betracht gezogen, wenn sie sich jeweils innerhalb der dargestellten Suchbereiche befinden.

4.3.1 Verfahren zur Punktverfolgung

Für die Verfolgung der Punkte werden zwei Verfahren in Erwägung gezogen. Das erste Verfahren verwendet konsequent die Informationen aus dem 3D-Kalman-Filter (Kapitel 6). Durch die Prädiktion des Zustandsvektors

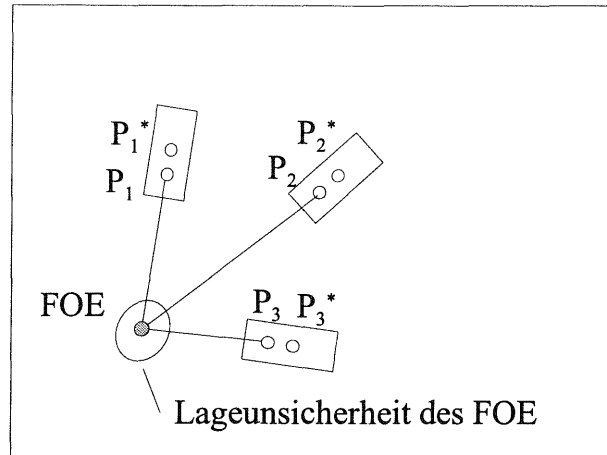


Abbildung 4.4: Suchbereiche bei Elimination der rotatorischen Bewegungsanteile

und dessen Genauigkeit liegen die äußere Orientierung, die Koordinaten der bislang schon verfolgten Punkte und deren stochastische Beschreibung durch Kovarianzmatrizen vor. Die Projektion der Punkte ins Bild liefert die prädierte Position, die zugehörige Kovarianzmatrix die Unsicherheit und damit den Suchbereich für den jeweiligen Punkt. Zur Vereinfachung wird die Fehlerellipse durch ein umschreibendes Rechteck angenähert, womit sich ein ähnliches Bild wie in Abbildung 4.4 ergibt.

Das Verfahren lehnt sich in starkem Maße an die Idee der Projektionsansätze zur 3D-Rekonstruktion (Abschnitt 1.6.2) an. Die Modellierung der Szene ist auf 3D-Punkte und deren stochastische Beschreibung beschränkt. Die 2D-3D Korrespondenz der aus dem aktuellen Bild extrahierten markanten Punkte zu den 3D-Szenepunkten wird durch die Projektion der 3D-Punkte ins Bild einfacher, da das durch die 3D-Kalman-Filterung gewonnene Wissen über die Bewegung des Sensors einbezogen ist.

In den Experimenten zur Zuordnung hat sich das Verfahren als nicht sehr vorteilhaft erwiesen. Die Ursache hierfür liegt in der Modellierung der Rotationsbewegung. Die Lenkbewegungen bei Fahrspurwechsel (die Aufnahme 40 in Abbildung 1.5 deutet dieses Manöver an) sind eher sprunghaft und müssen in der stochastischen Komponente des Bewegungsmodells repräsentiert werden. Im Bild äußert sich das in einer Unsicherheit von ca. 50 Pixeln (siehe auch Tabelle 1.1), die alle weiteren Einflüsse überdeckt und damit einen verhältnismäßig großen Suchbereich bei der Punktverfolgung festlegt.

Es wurde deshalb ein effizienteres zweites Verfahren entwickelt und weiterverfolgt. Es beruht auf dem Vorschlag von Burger (1992) einen "Fuzzy FOE" zu bestimmen und den rotatorischen Einfluß der Bewegung in der Bildebene zu approximieren und zu eliminieren. Der im wesentlichen verbleibende translatorische Einfluß führt zu Suchfenstern, wie Abbildung 4.4 zeigt.

Die Aufgabenstellung zur Übertragung des FOE aus dem Bild i in das Bild $i + 1$ ist in der Prinzipskizze 4.5 dargestellt. Der FOE in Bild i soll als Schnittpunkt der Basis des Bildpaares $i - 1/i$ mit der Bildebene von Aufnahme i festgelegt sein (der FOE $_i$ wurde in der Prinzipskizze in die Bildmitte gelegt). Dieser Punkt wird durch intensitätsbasierte Zuordnung (siehe Abschnitt 5.3.3) in das Bild $i + 1$ übertragen. Zeigt der FOE $_i$ auf einen vom Sensor weit entfernten 3D-Szenepunkt, so wird durch die Lage des zugeordneten FOE im Bild $i + 1$ die Translationsrichtung von Bild i nach Bild $i + 1$ übertragen. Folglich kann die Lagedifferenz des FOE vor und nach der Zuordnung zur Approximation der Richtungsänderung des Bildes herangezogen werden. Die Rotationsbewegung um die Translationsrichtung wird dabei als hinreichend klein vorausgesetzt.

Die gewonnene Information über die Sensorrotation wird entsprechend Abb. 4.4 zur Einschränkung des Suchraums für die Zuordnung markanter Punkte benützt. Für die Festlegung des FOE in Bild i gibt es weitere Möglichkeiten. Die Basis des Bildpaares $i/i + 1$ legt ebenso wie die Basis von Bildpaar $i - 1/i$ eine FOE-Position im Bild i . Eine andere Möglichkeit besteht darin, durch die Projektionszentren mehrerer Aufnahmezeitpunkte eine Bahnkurve zu legen und die momentane Bewegungsrichtung über die Tangente zu approximieren. Die ver-

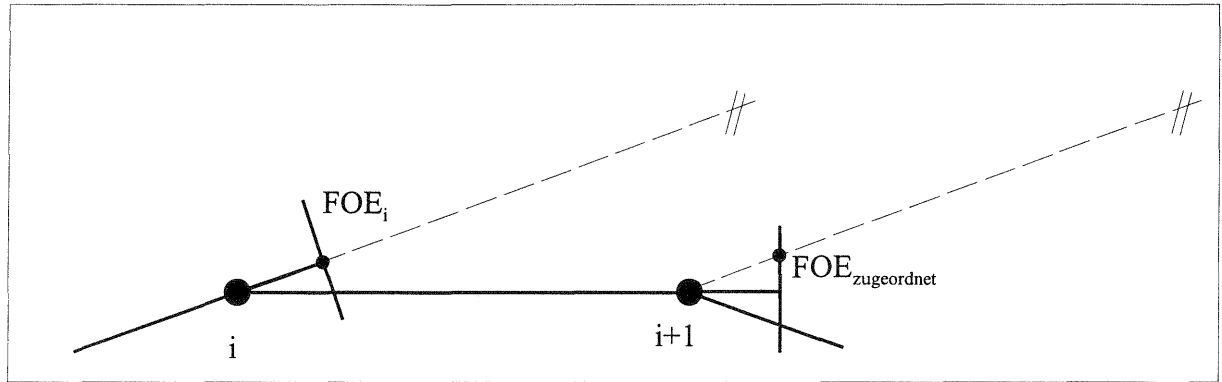


Abbildung 4.5: Übertragung des FOE zur Approximation der Sensorrotation zwischen Bild i und $i + 1$.

schiedenen Möglichkeiten zeigen, daß schon diskretisierungsbedingt der FOE mit einer Lageunsicherheit behaftet ist.

Genügt die Rotationsapproximation über die Verschiebung des FOE nicht, so kann es z. B. auch bei einem Sensor mit einem größerem Öffnungswinkel notwendig werden, daß mehrere weit entfernte Punkte zu verfolgen sind und mit diesen Punkten die dreidimensionale Rotation aus (2-39) ermittelt wird.

Für die Suche nach den korrespondierenden Punkten ist eine Vorinformation über die Entfernung der Punkte hilfreich, die bei bereits verfolgten Punkten direkt aus dem Zustandsvektor des Kalman-Filters abgeleitet werden kann. Aber auch ohne dieses Wissen genügt für die Abschätzung der Größe der in Abbildung 4.4 dargestellten Suchbereiche das sehr grobe Vorwissen über die Aufnahmebasis, die Kameraausrichtung und die minimale Entfernung der Punkte (z.B. von 10m), das bei der Charakterisierung der Sequenz in Tabelle 1.1 aufgelistet wurde.

4.3.2 Experimente und Resultate zur Punktextraktion und -verfolgung

Die gewählten Parameter und Schwellwerte für die Extraktion markanter Punkte sind in Tabelle 4.2 zusammengefaßt. Die im Detektionsschritt durch Schwellwertbildung und lokale Maxima-Suche selektierten markanten Fenster besitzen alle eine isotrope Textur, was durch den relativ hohen Schwellwert q_{min} garantiert wird. Für die Klassifizierung der lokalen Bildstruktur in Ecken, Kreise und sonstige isotrope Texturen wird ein Signifi-

| Fensterbezogene Parameter | | | Schwellwerte | | |
|---------------------------|--------------|-----------------------|--------------------|---------------------|------------------------------------|
| Faltungsfunktion ψ | Fenstergröße | Nicht-Maxima Umgebung | Rundheit q_{min} | Gewicht w_{min} | Klassifizierung $F_{79,79,99.9\%}$ |
| box | 9×9 | 9×9 | $= 0.9$ | $= \text{med}(w_i)$ | $= 2.02$ |

Tabelle 4.2: Vorgaben zur Lokalisierung und Klassifizierung markanter Punkte

kanzniveau von 99.9% festgesetzt. Die Ergebnisse der Prozessierung von 127 Bildern⁶ der Bildsequenz 'Rastatt' sind summarisch in Tabelle 4.3 angegeben. Bemerkenswert ist, daß mehr als 40% der ausgewählten Fenster (im Mittel 563) beim Übergang auf markante Punkte (323) eliminiert wurden. Die markanten Punkte sind dabei nach der Schätzung der optimalen Punktlage einer erneuten Nicht-Maxima Unterdrückung unterzogen worden.

Sehr interessant ist auch das Ergebnis der Klassifikation, wovon ähnliche Untersuchungen aus der Literatur

⁶Der letzte Abschnitt der 160 Bilder umfassenden Sequenz (siehe Abschnitt 1.7.3) wurde in den Experimenten nicht berücksichtigt, da hier das Fahrzeug durch eine stark gekrümmte Kurve fuhr. Die dadurch auf ca. 90% reduzierte Querüberdeckung hätte für die Fragestellung nach der Verfolgbarkeit von Punkten und zu einer aufnahmebedingten Obergrenze geführt, die Rückschlüsse auf die Stabilität des Punktfinders nur bedingt erlaubt.

| Anz. der prozess. Bilder | mittl. Anz. pro Bild | | | Klassifizierung der markanten Punkte | | | | | |
|--------------------------------|------------------------|-------------|------------|--------------------------------------|------------|--------|------------|--------|------------|
| | ausgewählte Fenster | markante P. | | isotrop | | Ecke | | Kreis | |
| | | Anz.P. | w_{mean} | Anz.P. | w_{mean} | Anz.P. | w_{mean} | Anz.P. | w_{mean} |
| 127 | 563 | 323 | 141.1 | 250 | 131.6 | 41 | 179.8 | 32 | 184.5 |

Tabelle 4.3: Statistik der extrahierten Punkte (w_{mean} : Mittelwert über alle Punkte und alle Bilder, Anz.P.: Anzahl Punkte)

bislang nicht bekannt sind. Verglichen mit dem Anteil allgemein isotrop texturierter Punkte (77%) ist der Anteil der Ecken und Kreise mit zusammen 23% vergleichsweise gering. Allerdings ist das durchschnittliche Gewicht der Ecken und Kreise deutlich höher als das der sonstigen isotropen Punkte.

Die experimentellen Aspekte des im vorhergehenden Abschnitt beschriebenen Verfahren zur Punktverfolgung betreffen die Festlegung der Suchbereiche, die Schwellwerte für den Merkmalsvergleich und die Auswahl des "besten" Punktepaars, wenn bei der Verfolgung eines Punktes mehrere mögliche Partner in Betracht kommen. Die Abschätzung des Suchbereichs gemäß (2-39) mit einer Basis von 1 m in Bewegungsrichtung und einer Mindestentfernung von 10 m führt für einen Punkt am Bildrand zu einem radialen Suchbereich von maximal ca 40 Pixeln. Die Querkomponente, die z. B. auch die Lageunsicherheit des FOE berücksichtigt, wurde mit 10 Pixeln festgelegt. Beim Vergleich der Merkmale wurden 30% Abweichung des Gewichtes toleriert. Da ein hohes Rundheitsmaß bei der Extraktion der Primitive gefordert wurde, ist die Ähnlichkeit bezüglich dieses Merkmals a priori gegeben. Für jeden so gefundenen möglichen Partner eines zu verfolgenden Punktes wird die lokale Umgebung des Punktepaars herangezogen um durch Korrelation mögliche Fehlkorrespondenzen zu eliminieren. Als Schwellwert hierfür wird ein minimaler Korrelationskoeffizient von $\rho_{min} = 0.75$ angesetzt. Zur Korrelation wird der mittelwertfreie, skalierte Produktmomentenkorrelationskoeffizient verwendet. Treten bei der Verfolgung eines Punktes mehrere Punkte als mögliche Partner auf, wird der bezüglich des Korrelationskoeffizienten als ähnlichster Partner bestätigte Punkt verwendet.

Aus dem in Tabelle 4.4 zusammengestellten Ergebnis der Punktverfolgung kann die Erfolgsquote der Zuordnung abgelesen werden. Die Statistik zeigt, daß fast die Hälfte aller Punkte (genau 2991 von 6823 Objektpunkten) nur

| Punktverfolgung: "K" Punkte erscheinen in genau "L" Bildern | | | | | | | | | | | | | | |
|--|------|------|-----|-----|-----|-----|-----|-----|----|----|----|----|----|----|
| L = | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| K = | 2921 | 1322 | 744 | 443 | 337 | 218 | 221 | 101 | 88 | 78 | 59 | 38 | 39 | 50 |
| L = | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| K = | 21 | 14 | 21 | 13 | 17 | 11 | 9 | 7 | 7 | 5 | 5 | 6 | 3 | 2 |
| L = | 30 | 31 | 32 | 33 | 34 | 35 | 37 | 38 | 40 | 41 | 42 | 43 | 44 | 45 |
| K = | 4 | 2 | 2 | 2 | 2 | 1 | - | 2 | 1 | 1 | 1 | 2 | 2 | 1 |
| Gesamtanzahl: # der Objektpunkte: 6823 # der Bildpunkte: 29149 | | | | | | | | | | | | | | |

Tabelle 4.4: Statistik zur Punktverfolgung

zwei aufeinanderfolgende Bilder verknüpfen. Immerhin jeder 5. Punkt wurde auch noch in einem 3. Bild wiedergefunden und jeder 9. Punkt verknüpft vier aufeinanderfolgende Bilder. Eine mögliche Ursache für die geringe Anzahl der Verknüpfungen bei den meisten Punkten ist, daß die extrahierten Punkte wenig Stabilität innerhalb der Sequenz aufweisen. Tatsächlich läßt sich bei der Verfolgung der meisten markanten Punkten beobachten, daß innerhalb der lokalen Nachbarschaft eher zufällig andere Punkte mit stärkerem Gewicht hervortreten. Dieses Resultat widerspricht natürlich dem Kriterium Stabilität, das bei Entwurf des Punktfinders gefordert wurde. Die lokale Maxima-Suche zur Selektion markanter Fenster ist deshalb ein kritischer Schritt des Punktfinders. Versuche mit variierenden Fenstergröße und variierender Auflösung im Bild auf der Grundlage von Scale-Space

Techniken deuten an, daß hierfür eine befriedigendere Lösung gefunden werden kann (Heikkilä, 1989, Schneider und Hahn, 1992).

Ein sehr positives Ergebnis der Punktverfolgung ist, daß einige Punkte auch in mehr als 40 Bildern identifiziert werden konnten. Diese Punkte werden innerhalb der 3D-Rekonstruktion geometrisch hervorragend kontrolliert und tragen viel zur Bewegungsrekonstruktion über die Sequenz bei. Eine kleine Auswahl dieser Punkte soll im nächsten Abschnitt detaillierter betrachtet werden.

Die Zuordnung der markanten Punkte von Bild zu Bild kann durch Verschiebungsvektoren visualisiert werden. Beispielhaft sind die Verschiebungsvektoren der Bildpaare 41-42, 42-43, 51-52, 53-54 in Abbildung 4.6 dargestellt. Bei genauer Betrachtung sind einige Fehlzuordnungen insbesondere aufgrund der Nachbarschaftsbeziehung zu umgebenden Verschiebungsvektoren erkennbar. Sie führen zum Großteil bei der 3D-Koordinierung der Punkte zu Widersprüchen und können deshalb innerhalb der Rekonstruktion beseitigt werden. Die Verschiebungsvektoren der beiden Bildpaare 41-42 und 42-43 zeigen beispielhaft das Ergebnis der Zuordnung über drei Bilder. Wie

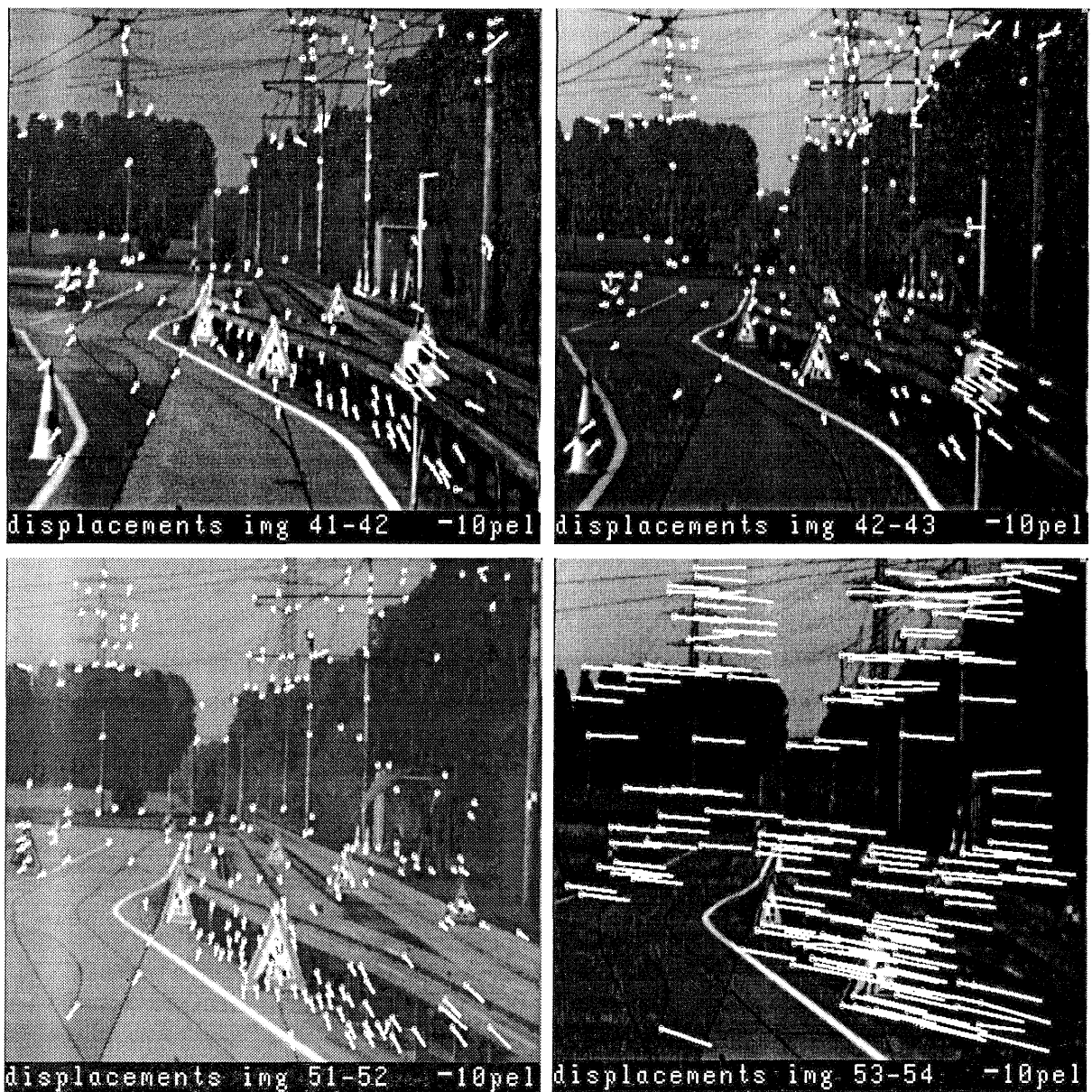


Abbildung 4.6: Verschiebungsvektoren an ausgewählten Beispielen

stark sich die Lenkbewegung in der Sequenz äußert, kann in den nächsten beiden Beispielen betrachtet werden.

Während die Verschiebungsvektoren in Bildpaar 51-52 noch vorwiegend eine translatorische Bewegung anzeigen, dominiert in Bildpaar 53-54 die durch die Lenkbewegung verursachte Rotationsbewegung des Sensors.

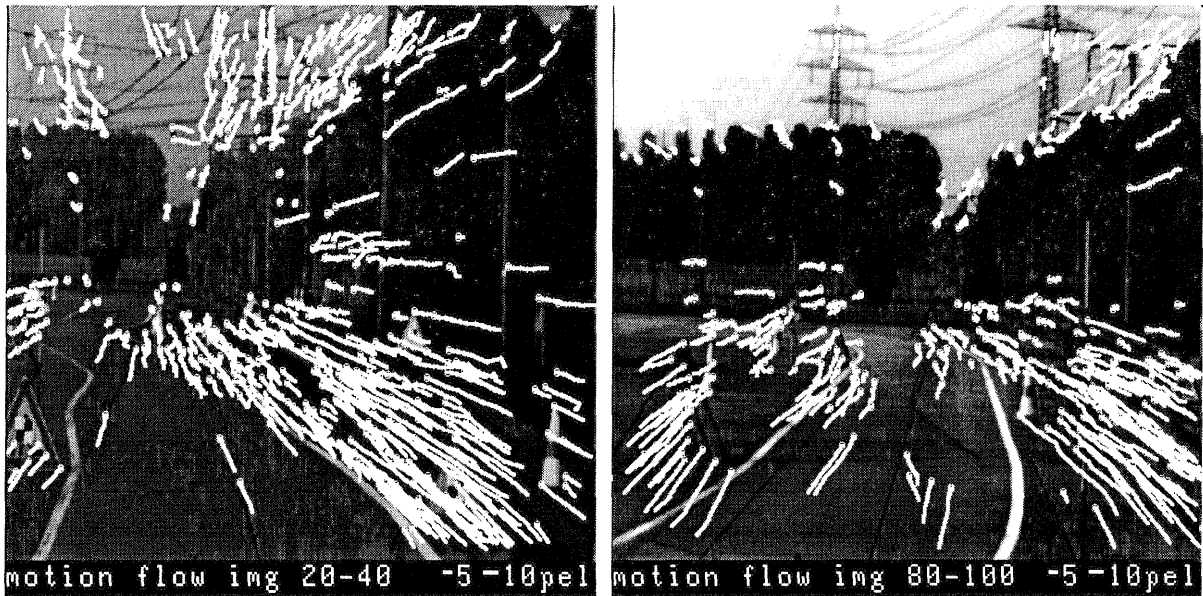


Abbildung 4.7: Punktverfolgung über Abschnitte von jeweils 20 Bildern. Dargestellt ist der translatorische Bewegungsanteil der Verschiebungsvektoren

Einen guten Eindruck von der Punktverfolgung über einen längeren Abschnitt gibt Abbildung 4.7. Die Rotationsbewegung des Sensors wurde für diese Darstellung eliminiert, d. h. aufgetragen ist nur die translatorische Bewegung der Punkte. Dargestellt sind alle verfolgten Punkte, die zwischen Bild 20 und 39 bzw. 80 und 99 extrahiert und zugeordnet wurden, unabhängig davon, wie lange sie in der Sequenz verfolgt werden konnten.

4.3.3 Beispiele verfolgter Punkte und Diskussion

Die insgesamt 6823 verfolgten Objektpunkte entsprechen 29149 extrahierten markanten Bildpunkten in der Sequenz. Natürlich können hiervon nur wenige Punkte im Detail betrachtet werden. Wir wollen uns auf 6 Punkte beschränken. Die meisten der nur eine geringe Anzahl Bilder verknüpfenden Punkte sind als allgemeine isotrope Texturen klassifiziert worden und nicht sehr aussagekräftig, weshalb sie nicht weiter dargestellt werden sollen. In den einzelnen Abbildungen ist jeweils die Optimalpunktschätzung des markanten Punktes eingetragen.

Der am häufigsten wiedergefundene Punkt 1156 (Abb. 4.8) wurde über genau 45 Bilder verfolgt. Die weiße Straßenmarkierung bildet sich zu Beginn der Verfolgung als Ecke im Bild ab. Die Maßstabsänderung und auch die Änderung der Blickrichtung führen dazu, daß die Ecke in einen ausgeprägten Kreisbogen übergeht und schließlich nicht mehr extrahiert wird. Betrachtet man die in den einzelnen Bildausschnitten eingeblendeten Bildnummern, so fällt auf, daß die Bilder 11, 14 und 17 fehlen. Dieses Fehlen ist aufnahmebedingt und mußte bei der Analyse der Sequenz berücksichtigt werden, ist aber nicht auf die Bildzuordnung zurückzuführen. Außerdem zeigen die einzelnen Bildausschnitte, daß durch Bewölkung die Bilder 4 bis 25 der Sequenz stärker abgedunkelt sind. Hiergegen erweist sich die Punktextraktion als sehr robust, was eine der Stärken der merkmalsbasierten Bildzuordnung ausmacht.

Im Gegensatz zu dem hervorragend verfolgten Punkt 1156 ist Punkt 1031 eher geeignet um Probleme aufzuzeigen. Der als isotrope Textur klassifizierte Punkt befindet sich im Schnittpunkt zweier Streben in der Nähe eines Strommastes. In Bild 4 wird allerdings ein Punkt auf einer benachbarten Straße detektiert und weiterverfolgt. Zufällig befinden sich die beiden Strebenenden dabei auf einer Geraden die zum Expansionspunkt zeigt, weshalb der um mehr als 10 Pixel verschobene Punkt im Bild 4 innerhalb des Suchbereiches liegt. Im Bild 13 wird der

virtuelle Schnittpunkt der Strebe mit dem räumlich weiter entfernten Mast detektiert, d. h. bei der Verfolgung dieses Punktes über 13 Bilder sind zwei Zuordnungsfehler aufgetreten.

Die Verfolgung von Punkt 3338 (Abb. 4.9) zeigt ebenfalls Fehler, die zudem auch bei der geometrischen Kontrolle durch die Schnittbedingung innerhalb der Rekonstruktion nicht identifiziert werden konnten, worauf in Abschnitt 6.3 noch näher eingegangen wird. Der Punkt detektiert im ersten Bild eine Ecke eines Verkehrszeichens, das an ein Stativ angelehnt ist. Er springt dann über auf die Spitze eines Kegels und verfolgt diesen über 22 Bilder bis er schließlich hinter dem Verkehrszeichen bzw. Stativkopf verschwindet. Weiterverfolgt wird dann der Stativkopf über insgesamt 14 Bilder. Zu diesem widerspruchsfreien Wechsel in Bild 50 trägt ein hell-dunkel Übergang im Bildhintergrund bei, der sich zufällig mit den Objekten im Vordergrund zur markanten Textur ergänzt. Daß der Punkt auch bei der Rekonstruktion nicht als Fehler aufgedeckt werden konnte, liegt daran, daß die erzielte Genauigkeit der Tiefenschätzung in diesem Punkt nicht ausreichte um den Tiefensprung vom Kegel auf den Stativkopf aufzulösen.

Die Verfolgung der Fahrbahnmarkierung in Punkt 1220 läßt erkennen, daß auch bei relativ einfacher Textur die optimal geschätzte Punktlage dieser Ecke im Bereich von mehreren Pixeln wandert (vgl. z. B. Bild 4 und Bild 15). Hauptsächlich verantwortlich hierfür ist ein sich um den Faktor 2 unterscheidender Maßstab zwischen Bild 0 und Bild 18. Der in Bild 16 weiterverfolgte Punkt ist ganz offensichtlich ein Zuordnungsfehler. Die Ecke der Fahrbahnmarkierung befindet sich in Bild 16 so nahe am Bildrand, daß sie vom Punktfinder nicht mehr extrahiert wird.

Punkt 1197 (Abb. 4.10) zeigt ein Verkehrszeichen, das zusätzlich mit einer Markierung versehen wurde. Diese Signalisierung wurde verwendet um eine unabhängige Kontrolle durch geodätische Messungen zu ermöglichen. Der Punkt wurde als Grauwertecke detektiert und verfolgt bis er schließlich aus dem Gesichtsfeld der Kamera verschwindet. Die Art der Markierung trägt in diesem Beispiel maßgeblich zum Erfolg der Zuordnung bei, da für diesen Punkt selbst ein Maßstabsunterschied von ca 300% zwischen Bild 0 und Bild 24 keine Schwierigkeiten bereitete.

In Punkt 1282 ist ein regelmäßiges Muster die Ursache für eine Folge von Fehlzuordnungen. In diesem Punkt ist auch die im nächsten Kapitel näher erläuterte intensitätsbasierte Punktverfolgung nicht erfolgreich. Da sich aber der Punkt relativ nah zum Projektionszentrum befindet, kann er bei der Rekonstruktion im Kalman-Filter sehr leicht als grober Fehler detektiert werden.

Abschließende Bewertung

Bei der merkmalsbasierten Verfolgung von Punkten hat sich gezeigt, daß ungefähr die Hälfte aller Verknüpfungspunkte nur in zwei Bildern identifiziert werden konnte. Als Ursache hierfür stellte sich bei der interaktiven Nachprüfung heraus, daß die Stabilität der markanten Punkte sehr viel geringer ist, als vorab (vergleiche Abschnitt 4.1) zu erwarten war. Die lokale Umgebung dieser sehr instabilen Punkte wurde vom Punktfinder nahezu ausschließlich als allgemeine isotrope Textur klassifiziert.

Für die Kalman-Filterung zur Bestimmung der Bewegungs- und Strukturparameter sind Punkte vorteilhaft, die sehr viele Bilder verknüpfen. Im Experiment konnten ca. 100 Punkte in 20 und mehr Bildern verfolgt werden. Auch hier zeigt die interaktive Nachprüfung schon bei wenigen ausgewählten Punkten, daß im Vergleich zur intensitätsbasierten Zuordnung (Abschnitt 5.3.3) ein relativ hoher Prozentsatz an Fehlzuordnung enthalten ist. Die innerhalb der Kalman-Filterung (Kapitel 6) festgestellte Fehlerquote liegt bei ca. 30 %, wobei die zwei und drei Bilder verknüpfenden Punkte bei der Ermittlung in dieser Fehlerquote nicht einbezogen wurden.

Wie in den Bildbeispielen zu sehen ist, hat wechselnde Bewölkung zu starken Helligkeitsunterscheiden innerhalb der Bildsequenz geführt. Hiergegen hat sich die Punktextraktion sehr robust erwiesen, was einer der Vorzüge der merkmalsbasierten Punktverfolgung ist.

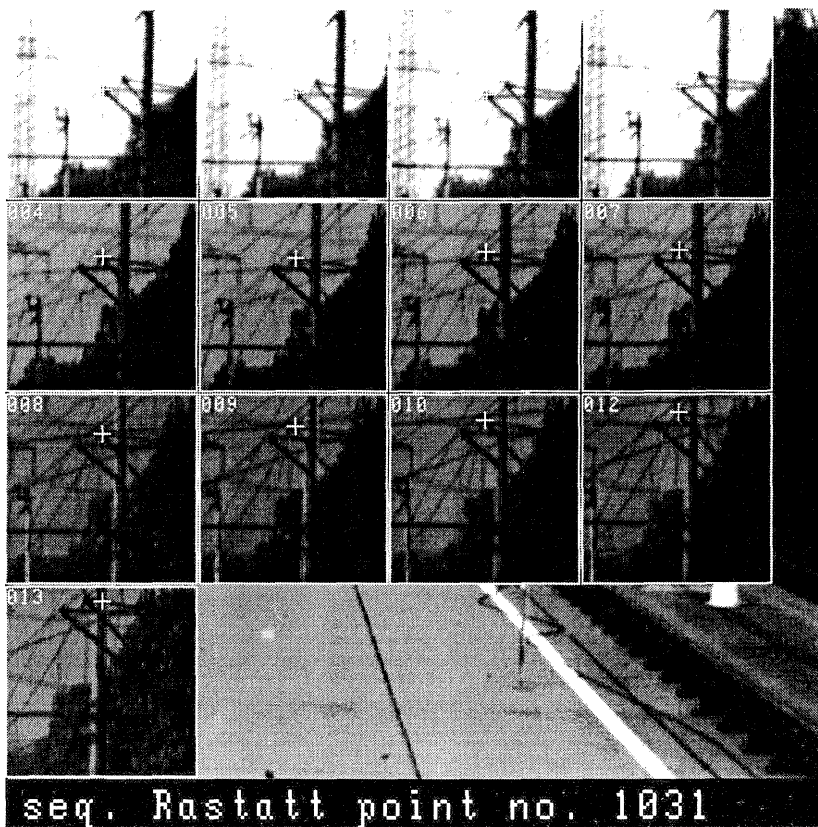
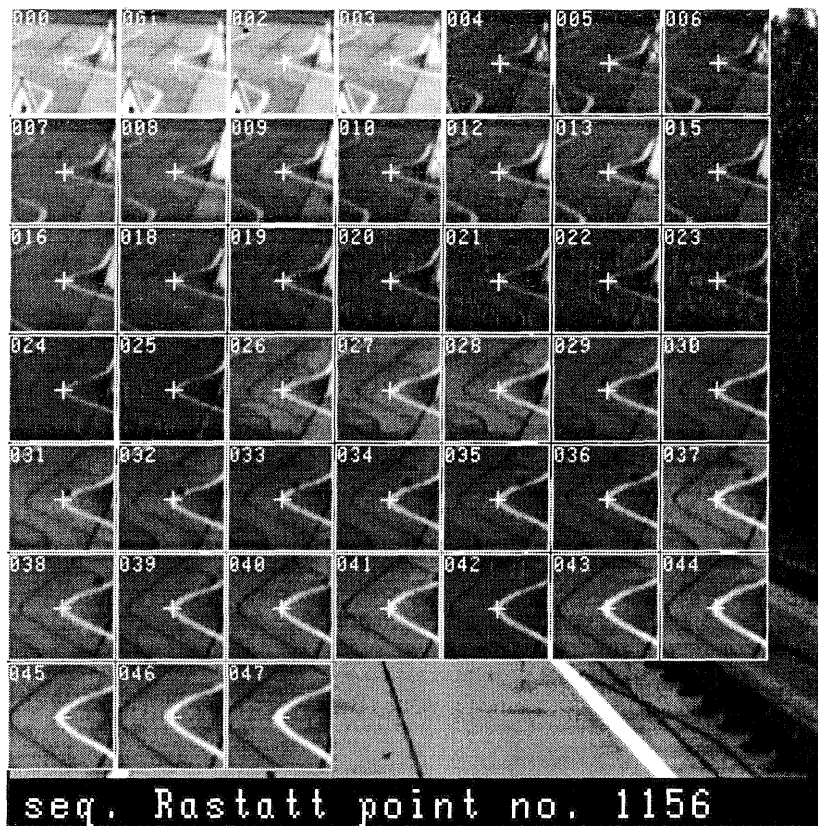


Abbildung 4.8: Beispiele verfolgter Punkte

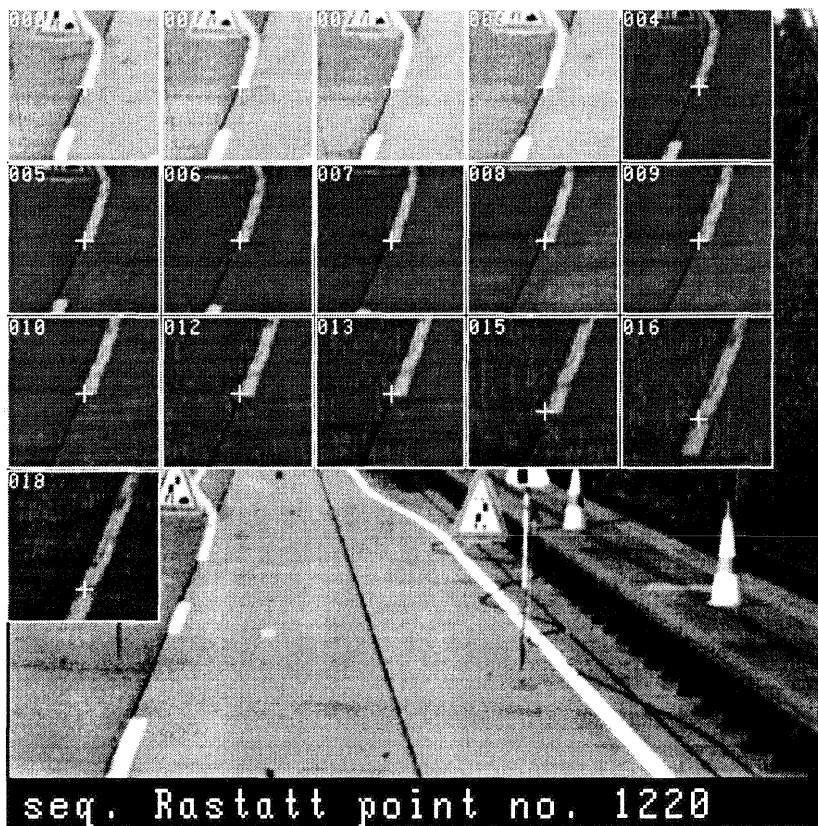
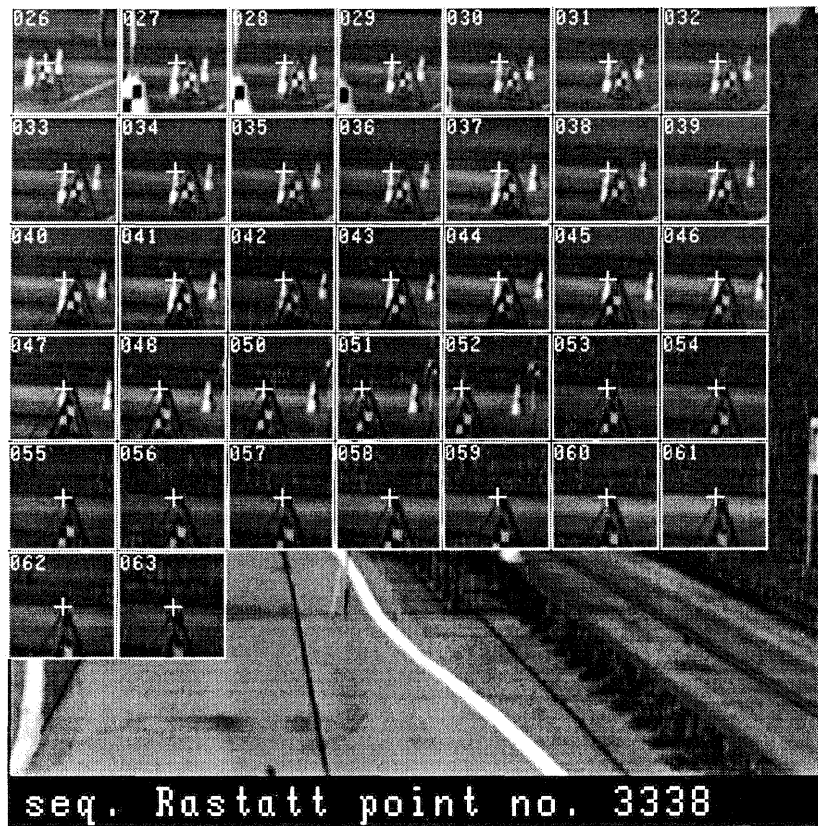


Abbildung 4.9: Beispiele verfolgter Punkte

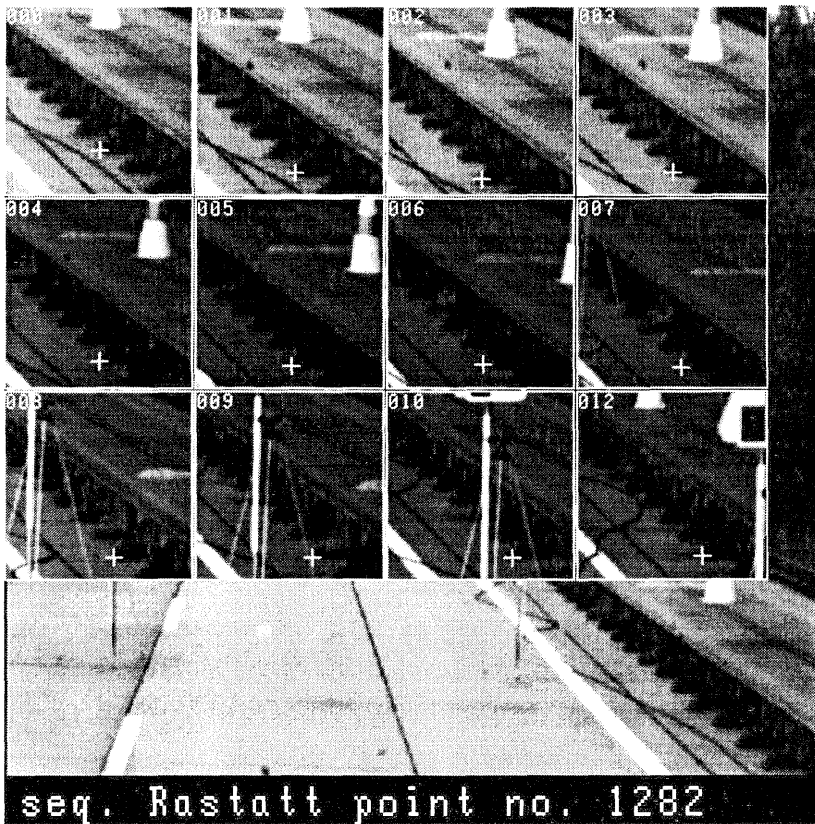
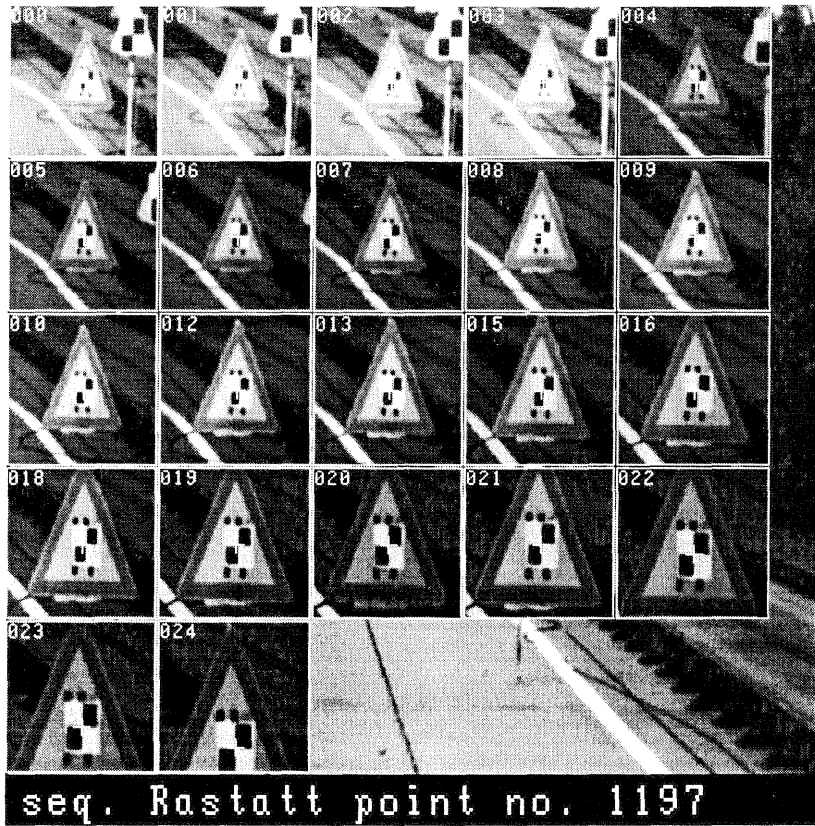


Abbildung 4.10: Beispiele verfolgter Punkte

5 Intensitätsbasierte Bildzuordnung im Scale-Space

Wissenschaftlicher Beitrag der Arbeit in diesem Kapitel:

Durch die Integration des Scale-Space in die intensitätsbasierte Bildzuordnung wird ein völlig neuer Weg zum Entwurf von Bildzuordnungsverfahren aufgezeigt. Die wesentlichen Eigenschaften des Verfahrens sind (a) ein großer Konvergenzradius, (b) schnelle Konvergenz, die sich in einer geringen Anzahl notwendiger Iterationen äußert und (c) die Möglichkeit Schärfeunterschiede zwischen den Bildern mitzuschätzen. Experimentell hat sich das Verfahren bei der Punktverfolgung in Bildsequenzen sehr bewährt. Im Mittel gelingt die intensitätsbasierte Verfolgung über sehr viel mehr Bilder als die merkmalsbasierte Verfolgung markanter Punkte. Die Quote der Zuordnungsfehler liegt bei 4%, was zeigt, daß auch die Zuverlässigkeit gegenüber der merkmalsbasierten Punktverfolgung stark verbessert ist.

In diesem Kapitel soll ein neuer Ansatz zur Bildzuordnung präsentiert werden, der den sog. Scale-Space in die Zuordnung einbezieht. Üblicherweise macht sich die Bildzuordnung Bildpyramiden zunutze, um die vom Zuordnungsverfahren benötigten Näherungswerte zu bestimmen. Das Zuordnungsverfahren wird hierfür in eine Mehrebenenprozedur eingebunden, die eine kontrollierte "grob-zu-fein"-Prozessierung bewirkt. Typisch für Bildpyramiden ist, daß die einzelnen Ebenen der Pyramide mit fest vorgegebener Reduktion der Auflösung erzeugt werden. Vorab festzulegende Auflösungsebenen sind für Aufgaben ungeeignet, bei denen eine variable Auflösung benötigt wird oder die Wahl der diskreten Ebenen datengetrieben erfolgen sollte. Für solche Aufgaben empfiehlt sich eine Bildrepräsentation mit kontinuierlicher Glättung, wobei der Raum in dem die Glättung stattfindet als Scale-Space bezeichnet wird (Witkin, 1983). Der im folgenden vorgestellte Ansatz zur intensitätsbasierten Bildzuordnung greift auf die Idee des kontinuierlichen Scale-Space zurück. Für eindimensionale Signale sind die theoretischen Grundlagen hierzu in Hahn (1990) präsentiert worden. Durch die Erweiterung des Zuordnungsmodells für lokal begrenzte 2D-Bildausschnitte kann das Verfahren in den Experimenten zur individuellen Verfolgung der markanten Punkte verwendet werden.

In den nächsten Abschnitten werden zunächst die Begriffe Bildpyramide und Scale-Space präzisiert und anschließend das Zuordnungsverfahren detailliert beschrieben. In Experimenten wird die Eignung des Zuordnungsverfahrens für die Punktverfolgung untersucht.

5.1 Bildpyramiden und Scale-Space

Unter einer *Bildpyramide* wird eine Repräsentationsstruktur verstanden, die ein Bild in einem Satz von Bildern mit gleichmäßig reduzierter Auflösung wiedergibt, wobei die unterste Ebene der Pyramide das Bild selbst enthält. Diese Repräsentationsstruktur kann sowohl für ikonische als auch für symbolische Bildinformationen definiert werden. Werden Merkmale in der Pyramide dargestellt, so wird auch von Merkmalspyramiden gesprochen. Die Anwendung von Bildpyramiden ist außerordentlich vielfältig. Zahlreiche Beispiele bezüglich der Organisation, Speicherung, Übertragung und Komprimierung von Daten, insbesondere aber auch Anwendungen in der Bildanalyse sind in Rosenfeld (1984) aufgezeigt worden. Eine Übersicht über Anwendungen in der Digitalen Photogrammetrie findet sich in Ackermann und Hahn (1991).

In der Bildanalyse werden hauptsächlich Gauß'sche Bildpyramiden verwendet. Eine solche Pyramide kann durch rekursive Anwendung folgender Operationen erzeugt werden:

$$\begin{aligned} I_{level\ i} * G^\sigma &= I_{level\ i}^\sigma \\ pick\ I_{level\ i}^\sigma &\rightarrow I_{level\ i+1} \end{aligned}$$

Die Rekursion beginnt auf der Ebene 0 mit dem Originalbild $I = I_{level\ 0}$. G^σ ist ein Gaußkern mit Standardabweichung σ , *pick* steht für das Resampling des geglätteten Bildes, bei dem die Auflösung reduziert wird.

Das theoretische Fundament für die Generierung von Bildpyramiden ist das Abtasttheorem. Durch die Faltung mit dem Gaußkern wird ein Bild geglättet, d. h. hochfrequente Bildinformation wird eliminiert. Die Stärke der Glättung hängt von der Standardabweichung σ des Gaußkerns ab. Sind im tiefpaßgefilterten Bild alle Frequenzen kleiner als die Hälfte der höchsten (Nyquist) Bildfrequenz, kann das Abtastintervall im geglätteten Bild

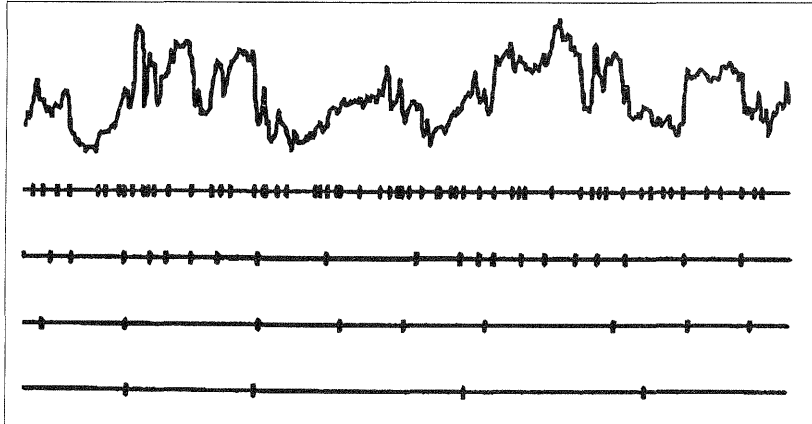


Abbildung 5.1: Nulldurchgänge der zweiten Ableitungen auf verschiedenen Glättungsniveaus (aus Witkin, 1983)

verdoppelt werden, ohne daß dabei Information verloren geht. In der Standardformulierung des Abtasttheorem erfolgt die Filterung bekanntlich mit einem idealen Tiefpaß; der Gaußfilter kann lediglich als eine Approximation eines idealen Tiefpasses betrachtet werden. Andere Filterkerne, die einen idealen Tiefpaß sehr viel besser annähern als der Gaußkern, sind in Meer et al. (1987) ausführlich untersucht worden. Typisch für das beschriebene Verfahren ist die 2:1 Reduktion der Kantenlänge eines Bildes (gemessen in der Anzahl der Pixel), d. h. die Auflösung wird beim Übergang auf die nächst höhere Pyramidenebene halbiert.

Der *Scale-Space* wurde von Witkin (1983) für eindimensionale Signale als (x, σ) -Ebene eingeführt. Generalisiert für zweidimensionale Bilder handelt es sich um einen (x, y, σ) -Raum, in dem (x, y) die Bildebene und σ den Glättungsparameter des Tiefpaßfilters angibt. Der Glättungsparameter, der von Witkin als *scale parameter* bezeichnet wird, zeigt an wie groß die in die Glättung einbezogene Umgebung ist.¹ Zur Glättung des Bildes wird wie bei der Generierung der Bildpyramide ein Gaußfilter verwendet. Die für einen Gaußkern sprechenden Argumente sind zum Teil sehr praktischen Ursprungs, z. B. daß die Integration, Differentiation und Fourier-Transformation für die Gaußfunktion sehr einfach sind. Der Gaußkern ist wie viele andere Tiefpässe auch symmetrisch und zerlegbar; das geglättete Bild strebt für $\sigma \rightarrow \infty$ gegen den Mittelwert und bleibt für $\sigma \rightarrow 0$ unverändert erhalten. Aus theoretischer Sicht wichtig ist der Nachweis, daß die Anzahl der lokalen Extrema in den Ableitungen beliebiger Ordnung mit zunehmender Glättung monoton abnimmt (Lindeberg, 1991). Mit dem so definierten Verständnis von Struktur besitzt die Faltung mit dem Gaußkern besondere Glättungseigenschaften: Es werden dabei keine neuen Strukturen erzeugt, was Witkin (1983) ursprünglich für die Nulldurchgänge der zweiten Ableitungen eines Gaußgefilterten Signals beobachtet hatte.

Die in Abbildung 5.1 dargestellten Positionen der Extrema (sie entsprechen in einem 2D-Bild Konturlinien) auf vier verschiedenen Glättungsniveaus zeigen eine grundlegende Schwierigkeit, die in der für Bildpyramiden typischen Diskretisierung der einzelnen Ebenen begründet liegt. Die Beziehung der extrahierten Punkte zwischen den Ebenen ist völlig ungeklärt. Hier wird der eigentliche Vorteil des Scale-Space offensichtlich, da bei kontinuierlichem Parameter σ ein Verfolgen von Merkmalen über den Maßstab sehr viel einfacher ist. Beispielsweise hat Bergholm (1987) den kontinuierlichen Glättungsparameter benützt, um in einem als "Kanten-Fokussierung" bezeichneten Verfahren rauschbedingte Konturen von anderen zu unterscheiden und zu eliminieren.

Witkin et al. (1987) nützen den Scale-Space zur Zuordnung eindimensionaler Signale, wobei sie Deformationen zwischen den Signalen modellieren. Die diskrete algorithmische Lösung des variational formulierten Ansatzes besteht in einer iterativen grob-zu-fein Prozessierung. Die Signale werden auf ein vorzugebendes Unschärfeniveau im Scale-Space angehoben. Von Iteration zu Iteration wird die Glättung um ein konstanten Wert $\Delta\sigma$ verringert, so daß nach einer festen Anzahl von Iterationen die Lösung erzielt wird.

¹In diesem Sinne steht der Begriff des Scale-Space durchaus in engerem Zusammenhang zum üblichen Maßstabsverständnis (z.B. werden im stark geglätteten Bild feinere Strukturen eliminiert, grobe und damit großräumigere Strukturen bleiben erhalten). Um Mißverständnisse zu vermeiden soll der Scale-Space dennoch nicht als Maßstabsraum bezeichnet werden. Auch hat sich der englische Begriff in der Literatur allgemein durchgesetzt.

Zusammenfassend soll nochmals hervorgehoben werden, daß der wesentliche Unterschied zwischen der ausschließlich auf einer Glättung (multi-scale) und der auf verschiedenen Auflösungen (multi-resolution) beruhenden Bildrepräsentation darin liegt, daß letztere die Reduktion der Anzahl der Bildpunkte von Ebene zu Ebene zum Ziel hat. Eine Bildpyramide ist deshalb bezüglich der Anzahl der zu verarbeitenden Bildpunkte sehr effektiv. Im Gegensatz hierzu wird beim diskreten Scale-Space eine konstante Anzahl von Gitterpunkten zugrundegelegt. Die Repräsentation des Bildes wird mit zunehmender Glättung immer redundanter. Der konzeptionelle Vorteil des Scale-Space liegt darin, daß auch bei diskreter Rasterung der Bildebene mit einem kontinuierlichen Glättungsparameter argumentiert werden kann, wobei dieser dann völlig frei festgelegt oder auch z. B. bei einer Zuordnung als Parameter mitgeschätzt werden kann.

5.2 Ein Verfahren zur Bildzuordnung im Scale-Space

5.2.1 Hintergründe und theoretische Grundlagen

Von den verschiedenen Aspekten der Bildentstehung (siehe Abschnitt 1.3) ist ein Aspekt bislang völlig unberücksichtigt geblieben. Er betrifft die Modellierung des optischen Systems "Kamera" als lineares verschiebungsinvariantes System, dessen Punktantwort durch eine 2D-Gaußfunktion beschrieben werden kann (Horn, 1986, S. 126). Die Antwort des Systems Kamera auf ein abzubildendes Signal $f(\mathbf{x})$ ergibt sich aus

$$I(\mathbf{x}; \beta) = h(\mathbf{x}; \beta) * f(\mathbf{x}) , \quad (5-1)$$

bzw. ausführlich dargestellt aus

$$I(x, y; \beta) = \int_{\xi=-\infty}^{\infty} \int_{\eta=-\infty}^{\infty} \frac{1}{2\pi\beta} e^{-(\xi^2+\eta^2)/2\beta} f(x-\xi, y-\eta) d\xi d\eta .$$

$\beta = \sigma^2 \geq 0$ ist die Varianz der Gaußfunktion, $h(\mathbf{x}; \beta) = (1/2\pi\beta)e^{-(|\mathbf{x}|^2)/2\beta}$ wird auch als Gauß'sche Punktverwaschungsfunktion bezeichnet. Der im vorhergehenden Abschnitt zur Erzeugung der Bildpyramide verwendete Gaußkern G^σ ist in der kontinuierlichen Form durch die Dichte $h(\mathbf{x}; \beta)$ gegeben.

Für ein vorliegendes Bild $I(\mathbf{x}; \beta_0)$ ist $\beta = \beta_0$ im allgemeinen ein fester aber unbekannter Parameter. Je größer β_0 ist, desto unschärfer ist das Bild $I(\mathbf{x}; \beta_0)$. Kann ein Objekt durch ein entsprechendes Signal $f(\mathbf{x})$ beschrieben werden, läßt sich eine Schätzprozedur formulieren, die neben Parametern zur Lokalisierung des Objekts auch den Parameter β_0 als Unbekannte beinhaltet. Thurgood und Mikhail (1982) haben eine Kleinste-Quadrate-Prozedur zur Parameterschätzung vorgeschlagen, wobei sie das Ziel verfolgten, Paßpunktsignale im Bild sehr präzise zu lokalisieren (siehe auch Mikhail, 1984). Als Parameter der Punktverwaschungsfunktion wurde von den Autoren σ_0 verwendet, was allerdings Nachteile mit sich bringt (Hahn, 1990), die auch im folgenden noch deutlich werden.

Bei einem kontinuierlichen Parameter β kann (5-1) auch als Definitionsgleichung der Scale-Space-Repräsentation eines Bildes betrachtet werden. Das Signal $f(\mathbf{x})$ ist dabei identisch mit dem Bild $I(\mathbf{x}; \beta = 0)$.² Die Definitionsgleichung (5-1) kann auch als Lösung der sogenannten zweidimensionalen Diffusionsgleichung

$$\frac{\partial I}{\partial \beta} = \frac{1}{2} \left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \right) \quad (5-2)$$

mit der Anfangsbedingung $I(\mathbf{x}; 0) = f(\mathbf{x})$ gewonnen werden (Lindeberg, 1991). Für das zu entwickelnde Zuordnungsverfahren zeigt die Diffusionsgleichung eine sehr wichtige Beziehung auf: Die Ableitung des im Scale-Space repräsentierten Bildes nach dem Glättungsparameter steht in linearem Zusammenhang zu den zweiten Ableitungen des entsprechend geglätteten Bildes nach der Lage und kann durch den Laplace-Operator berechnet werden. Der Beweis läßt sich mit den Regeln der Faltungs algebra leicht führen. Bekannt ist, daß die Ableitung eines Faltungsprodukts zweier Signale durch die Ableitung eines der beiden Signale erhalten werden kann. Angewandt auf (5-1) muß dann lediglich gezeigt werden, daß die Ableitung der Gaußfunktion h nach β : $\partial h / \partial \beta$

²Daß das Bild in den Scale-Space als Ausgangsniveau ($\beta = 0$) eingeht, impliziert nicht etwa eine Punktverwaschung mit $\beta_0 = 0$. Das Signal $f(\mathbf{x})$ wird hier lediglich mit einem (bei festem β_0) erfaßten Bild identifiziert, so daß der Scale-Space auf diesem Bild aufsetzt.

und die Laplacegefilterte Gaußdichte $\nabla^2 h$ sich um den Faktor 2 unterscheiden. Die explizite Differentiation der Gaußfunktion und der Koeffizientenvergleich sind mathematisch einfach und sollen deshalb nicht explizit dargestellt werden.

Eine weitere wichtige Eigenschaft für diskrete Berechnungen im Scale-Space ist durch

$$I(\mathbf{x}; \beta_1 + \beta_2) = h(\mathbf{x}; \beta_2) * I(\mathbf{x}; \beta_1) \quad (5-3)$$

gegeben. Mit $I(\mathbf{x}; \beta_1)$ liegt ein geglättetes Bild auf dem Niveau β_1 im Scale-Space vor, das durch einen Gaußkern mit dem Glättungsparameter β_2 gefiltert wird. Als Ergebnis wird ein stärker geglättetes Bild erhalten, das im Scale-Space der Glättungsebene $\beta_1 + \beta_2$ entspricht. Für den Beweis von (5-3) genügt es aufgrund von (5-1) zu zeigen, das sich die Varianzen bei der Faltung zweier Gaußkerne addieren. Durch die Fouriertransformation zweier gefalteter Gaußkerne gemäß

$$\begin{aligned} h(\mathbf{x}; \beta_1) * h(\mathbf{x}; \beta_2) & \quad \circ \longrightarrow \bullet \quad e^{-\frac{1}{2}|\mathbf{w}|^2\beta_1} \cdot e^{-\frac{1}{2}|\mathbf{w}|^2\beta_2} \\ = h(\mathbf{x}; \beta_1 + \beta_2) & \quad = e^{-\frac{1}{2}|\mathbf{w}|^2(\beta_1 + \beta_2)} \end{aligned}$$

und deren Rücktransformation ist der Beweis direkt erbracht.

5.2.2 Konzeption des Verfahrens

Die Aufgabenstellung für die Bildzuordnung ist in Abbildung 5.2 skizziert. I_1 und I_2 können als Bildausschnitte zweier aufeinanderfolgender Bilder einer Bildsequenz betrachtet werden. Geometrisch sollen die beiden Bildausschnitte durch einen Verschiebungsvektor \mathbf{u}_0 ineinander überführbar sein. Dies entspricht der zeitdiskreten Formulierung des innerhalb eines Fensters konstant angenommenen optischen Flusses (3-39). Selbstverständlich

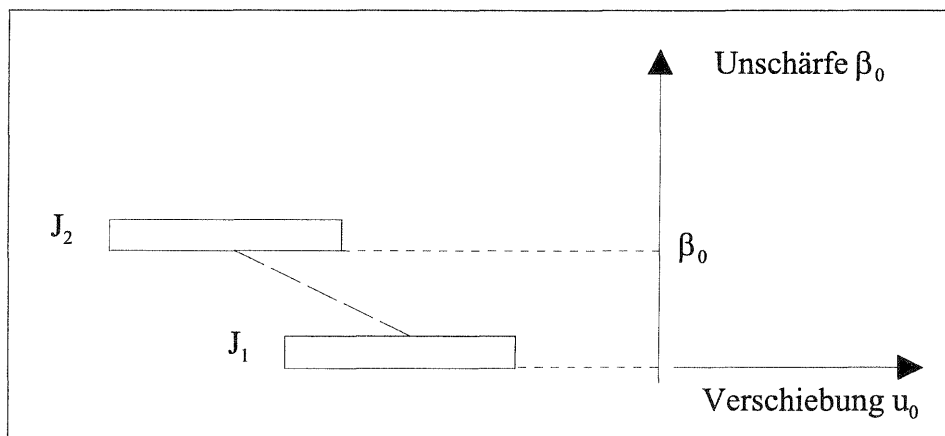


Abbildung 5.2: Bildzuordnungsaufgabe: Zu bestimmen sind der Verschiebungsvektor \mathbf{u}_0 und die Schärfedifferenz β_0 zwischen beiden Bildausschnitten

könnten anstelle von (3-39) auch alle anderen in Kapitel 3 vorgestellten Transformationsansätze einbezogen werden, ohne daß dies die Aufgabe prinzipiell ändern würde. Zusätzlich zur Verschiebung ist in Abb. 5.2 eine Schärfedifferenz β_0 der beiden Bilder berücksichtigt, die ebenfalls bestimmt werden soll. Bezüglich den in Kapitel 3 formulierten Minimierungsaufgaben (z. B. (3-35) oder (3-38)) impliziert das, daß mit β_0 ein weiterer Parameter mitgeschätzt werden muß.

Die konzeptionelle Erweiterung der Bildzuordnungsverfahren unter Einbeziehung des Scale-Space ist in Abbildung 5.3 veranschaulicht. Auf der vertikalen Achse ist der Glättungsparameter des Gaußkerns aufgetragen. Die durch Gaußfilterung mit kontinuierlichem β erhaltene Repräsentation des Bildes I_1 im Scale-Space ist in Abb. 5.3 durch Kennzeichnung des Raumes über der Bildebene angedeutet. Als Startsituation für den Zuordnungsprozeß wird ein stark geglättetes Niveau von $I_1(\mathbf{x}; \beta)$ gewählt. Die Schätzung des Glättungsunterschiedes sowie der

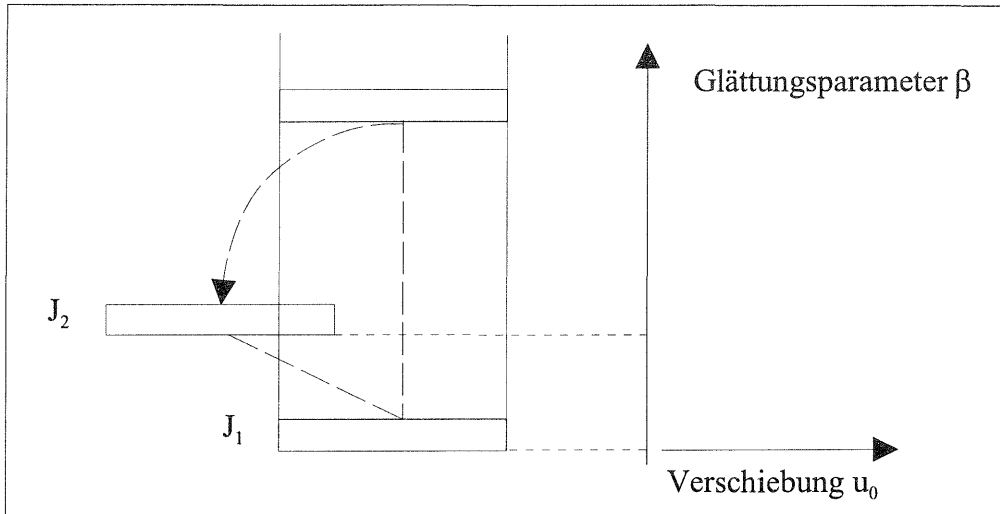


Abbildung 5.3: Bildzuordnung im Scale-Space

Verschiebung bezüglich I_2 soll nun iterativ erfolgen, wofür das Modell nachfolgend noch genau dargestellt wird. Erwartet wird von der Prozedur, daß sie einen Weg von "grob-zu-fein" findet (in der Skizze durch den gestrichelten Pfeil angedeutet) und mit dem Erreichen der Konvergenz die in Abb. 5.2 skizzierte Zuordnungsaufgabe gelöst ist.

Die eigentliche Besonderheit des neuen Verfahrens besteht in der *Integration des Scale-Space* in den Schätzprozeß. Gleichzeitig erlaubt das Zuordnungsverfahren ggf. Schärfedifferenzen zwischen beiden Bildern zu bestimmen. Ein Beispiel (Abbildung 5.4) soll einen Einblick in die "grob-zu-fein"-Schätzung geben, die sich im Konvergenzverhalten der iterativen Schätzprozedur zeigt. In diesem Beispiel wurden zwei eindimensionale Bildsignale

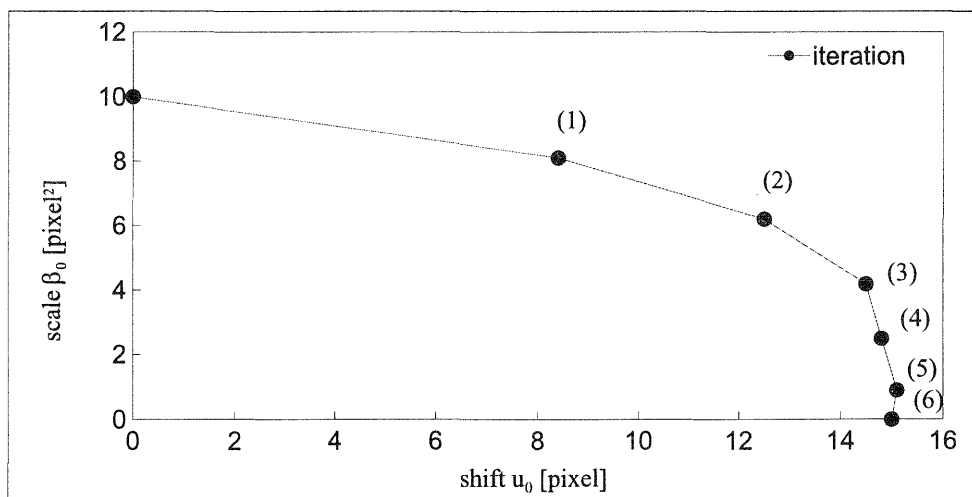


Abbildung 5.4: Konvergenzverhalten bei Scale-Space-basierter Bildzuordnung (aus Hahn, 1990)

gleicher Schärfe zugeordnet, die sich um 15 Pixel unterscheiden. Aufgetragen ist die jeweilige Schätzung in den einzelnen Iterationsschritten bis zur Konvergenz. Das Beispiel zeigt die beiden wesentlichen Vorzüge des Verfahrens auf. Zum einen wurde ein relativ großer Konvergenzradius erzielt, was bislang nur von Verfahren bekannt ist, die sich Bildpyramiden zunutze machen. Zum andern fällt auf, daß der Schätzprozeß mit relativ wenig Iterationen auskommt, d. h. daß er sehr schnell konvergiert.

5.2.3 Parametrischer Ansatz und Kleinste-Quadrate-Lösung

Auf dem dargestellten konzeptionellen Hintergrund soll nun der Zuordnungsansatz aufgebaut werden. Hierfür wird auf die in Kapitel 3 besprochenen geometrisch-radiometrischen Modelle zur Schätzung des optischen Flusses zurückgegriffen. Bei dieser Modellierung wurde bezüglich der räumlichen und zeitlichen Parameter zumeist kontinuierlich argumentiert und heuristisch die Bewegung auf kleine Verschiebungen in der Bildebene begrenzt. Da mit dem hier präsentierten Zuordnungsverfahren ein großer Konvergenzradius angestrebt wird, soll schon im Modellansatz die zeitdiskrete Betrachtung betont werden. Mit den Bildern I_1 und I_2 zweier diskreter Bezugszeitpunkte lautet das Modell

$$\begin{aligned} I_1(\mathbf{x}; \beta_1) &= f(\mathbf{x}; \beta_1) + n_1(\mathbf{x}) \\ I_2(\mathbf{x}; \beta_2) &= b_0 + b_1 f(\mathbf{x}'; \beta_2) + n_2(\mathbf{x}) \end{aligned} \quad (5-4)$$

wobei

$$\begin{aligned} \mathbf{x}' &= \mathbf{x} + \mathbf{u}_0 \\ \beta_2 &= \beta_1 + \beta_0 \\ f(\mathbf{x}; \beta) &= h(\mathbf{x}; \beta) * f(\mathbf{x}) \end{aligned}$$

n_1 und n_2 stehen für das Rauschen bei der Aufnahme der Bilder, d. h. bezüglich (5-1) wurde das Bildmodell um additives Rauschen erweitert (vgl. auch Abschnitt 1.3). Hierfür wird meist vereinfacht ein weißer Gauß'scher Rauschprozeß angenommen. Unbekannt sind in (5-4) die Parameter $b_0, b_1, u_0, v_0, \beta_0$ sowie die Funktion $f(\mathbf{x})$. Gegeben sind die Bilder I_1 und I_2 sowie die Punktverwaschungsfunktion h , die durch die Gaußfunktion modelliert ist. Grob gesprochen werden durch (5-4) die beiden Bilder bzw. Bildausschnitte durch die radiometrischen Kompensationsparameter b_0, b_1 entsprechend (3-23) angepaßt, d. h. $I_2 \cong b_0 + b_1 I_1$. Die Bewegungsmodellierung in der Bildebene wird durch eine einfache Verschiebung $\mathbf{u}_0 = \mathbf{x} - \mathbf{x}'$ der Bildkoordinatensysteme approximiert. Dieses Modell entspricht gemäß (3-39) einem konstant angenommenen optischen Fluß mit all den in Abschnitt 3.3.7 diskutierten Einschränkungen. In β_0 wird die Schärfedifferenz der beiden Bilder erfaßt, wobei das schärfere Bild I_1 dem unschärferen Bild I_2 durch die Faltung mit dem Gaußkern angeglichen wird, d. h. $I_2 \cong I_1 * G^{\beta_0}$.

Die unbekannte Funktion f in (5-4) stellt für die Zuordnungsaufgabe eine eher unbequeme Größe dar. Ist sie nicht wie hier auf die Bildebene sondern auf den Objektraum bezogen, kann f mit einem Orthobild (siehe Abschnitt 3.4.1) identifiziert werden, so daß sich die Mitbestimmung der diskretisierten Funktion lohnen kann. Ein Vorteil der gewählten Darstellungsform (5-4) ist, daß die Verallgemeinerung des Zuordnungsansatzes von 2 auf N Bilder direkt ersichtlich ist. Für eine simultane N -Bildzuordnung müssen lediglich weitere Gleichungen entsprechend der Beziehung für I_2 hinzugefügt werden.

Wird die Funktion f aus (5-4) eliminiert, kann entsprechend den in Kapitel 3 diskutierten Minimierungsaufgaben durch Summation über eine Umgebung R die Zielfunktion

$$\Omega = \sum_R (I_2(\mathbf{x}; \beta_2) - b_0 - b_1 I_1(\mathbf{x} + \mathbf{u}_0; \beta_1) * h(\mathbf{x} + \mathbf{u}_0; \beta_0))^2 \rightarrow \min \quad (5-5)$$

formuliert werden. Der Standardlösungsweg zur Bestimmung der in der Minimierungsfunktion (5-5) nichtlinear enthaltenen unbekannt Parameter besteht in der Entwicklung einer Taylor-Reihe bis zur 1. Ordnung. Mit den Näherungswerten³ für $b_0, b_1, u_0, v_0, \beta_0$, den partiellen Ableitungen $I_x = \partial/\partial x I_1(x + u_0, y + v_0; \beta_1 + \beta_0)$, $I_y = \partial/\partial y I_1(x + u_0, y + v_0; \beta_1 + \beta_0)$, $I_\beta = \partial/\partial \beta I_1(x + u_0, y + v_0; \beta_1 + \beta_0)$ sowie den Abkürzungen I_1 für $I_1(\mathbf{x} + \mathbf{u}_0; \beta_1)$, h für $h(\mathbf{x} + \mathbf{u}_0; \beta_0)$ und $\bar{I}_1 = b_0 + b_1 I_1 * h$ kann die aus (5-5) hervorgegangene linearisierte Minimierungsaufgabe gemäß

$$\Omega = \sum_R \left(\begin{pmatrix} 1 & I_1 * h & I_x b_1 & I_y b_1 & I_\beta b_1 \end{pmatrix} \begin{pmatrix} \Delta b_0 \\ \Delta b_1 \\ \Delta u_0 \\ \Delta v_0 \\ \Delta \beta_0 \end{pmatrix} + \bar{I}_1 - I_2 \right)^2 \rightarrow \min \quad (5-6)$$

³In der Regel genügt die Annahme $b_0 = 0, b_1 = 1, u_0 = 0, v_0 = 0$. Liegen geeignetere Vorinformationen vor, z.B. eine bessere Näherung für die Verschiebung, werden sie in der Regel als Pseudobeobachtung mit ihrer Unsicherheit mitberücksichtigt.

dargestellt werden. Das Rauschen der Intensitätsdifferenz $\bar{I}_1 - I_2$ ist wiederum als weißes Gauß'sches Rauschen approximiert, d. h. diskret werden die Intensitätsdifferenzen in den einzelnen Bildpunkten als gleichgenaue und unkorrelierte Beobachtungen angenommen. In der Form (5-6) ist die Minimierungsaufgabe den in Kapitel 3 kontinuierlich formulierten Minimierungsfunktionen soweit angeglichen, daß sie in einfacher Weise im direkten Vergleich betrachtet werden können. Z. B. korrespondieren die Intensitätsdifferenzen $\bar{I}_1 - I_2$ mit den temporalen Bildgradienten I_t . Die Parameterschätzung in nichtlinearen Modellen erfolgt über das Aufstellen und Lösen der Normalgleichungen nach dem üblichen iterativen Standardverfahren, bei dem die geschätzten Parameter einer Iteration jeweils als Linearisierungsstelle für die nächste Iteration verwendet werden. Die Konvergenz ist erreicht, wenn sich der geschätzte Parametervektor nicht mehr ändert oder äquivalent, wenn die aus dem nichtlinearen und dem linearisierten Modell bestimmten Residuenquadratsummen übereinstimmen.

Die Zuordnung durch den Scale-Space wird nun dadurch angestoßen, daß mit der Wahl des Startwertes für β_0 ein stark geglättetes Niveau im Scale-Space festgelegt wird. Die iterative Schätzung der Parameter endet für β_0 schließlich mit der geschätzten Schärfedifferenz der beiden Bilder. Die vorrangige Bedeutung bei der Verfolgung von Punkten in der Bildsequenz liegt aber in der durch das Zuordnungsverfahren kontrollierten grob-zu-fein Prozessierung.

Die Betrachtung der theoretischen Genauigkeit der geschätzten Größen gibt einen Anhaltspunkt für der Qualität der bestimmten Parameter. Von Interesse sind insbesondere die Verschiebung u_0 und die Schärfedifferenz β_0 . Für eine grobe Abschätzung sollen deshalb die radiometrischen Parameter außer acht gelassen werden. Die Diagonalelemente der Normalgleichungsmatrix für u_0, v_0, β_0 ergeben sich dann zu $\sum I_x^2, \sum I_y^2, \sum I_\beta^2$. Bei isotroper Bildtextur sind die Nebendiagonalelemente nahe 0, so daß sich die theoretische Genauigkeit für den Verschiebungsbetrag $|\hat{u}_0|$ zu

$$\sigma_{|\hat{u}_0|}^2 = \sigma_n^2 \frac{1}{\sum_N |\nabla I|^2}$$

und für die geschätzte Unschärfe zu

$$\sigma_{\hat{\beta}_0}^2 = \sigma_n^2 \frac{1}{\sum_N I_\beta^2} = \sigma_n^2 \frac{4}{\sum_N |\nabla^2 I|^2}$$

ergeben. σ_n ist die Rauschvarianz der in die Schätzung eingehenden Intensitätsdifferenzen. Summiert wird über die N Bildpunkte der Region R . Eine hohe Genauigkeit der geschätzten Verschiebung wird in stark texturierten Gebieten erzielt (vgl. Förstner, 1982), was auch ein sehr gewichtiger Aspekt für den Entwurf des Förstner'schen Punktefinders war (siehe Abschnitt 4.2.4). Je unschärfer ein Bild ist, d. h. je größer β_0 ist, desto stärker sind die Gradienten $\nabla I(\cdot; \beta_1 + \beta_0)$ geglättet. Folglich wird die geometrische Genauigkeit der Bildzuordnung mit wachsender Glättung zunehmend ungenauer, wobei sich der Einfluß über den Quotienten der geglätteten und ungeglätteten quadrierten Bildgradienten direkt abgeschätzt läßt. Die Genauigkeit des geschätzten Unschärfeparameters wird von den zweiten Ableitungen beeinflusst. Da sie im Mittel kleiner und sehr viel rauschanfälliger als die Gradienten sind, ist im allgemeinen die Schätzung des Unschärfeparameters sehr viel ungenauer als die geschätzte geometrische Verschiebung. Details zur algorithmischen Lösung sollen hier nicht vertieft werden. Statt dessen sei auf Hahn (1990) und Lang (1992) verwiesen.

5.3 Intensitätsbasierte Punktverfolgung - Experimentelle Untersuchung

Das im vorhergehenden Abschnitt vorgestellte Bildzuordnungsverfahren soll zur Verfolgung markanter Punkte in der Bildsequenz herangezogen werden. In erster Linie wird die intensitätsbasierte Zuordnung als ein die merkmalsbasierte Punktverfolgung (Abschnitt 4.3) ergänzendes Verfahren betrachtet. Insbesondere gilt dies für die sequentielle Übertragung des Expansionspunktes, dessen Lage in der merkmalsbasierten Zuordnung in grober Näherung benötigt wird.

Prinzipiell können die intensitätsbasierte und merkmalsbasierte Zuordnung auch als konkurrierende Verfahren zur Verfolgung markanter Punkte betrachtet werden. Typisch für die merkmalsbasierte Verfolgung von Punkten ist die Zweistufigkeit des Verfahrens, in dem zuerst die markanten Punkte in den Bildern extrahiert werden und dann in einem zweiten Schritt nach korrespondierenden Punkten gesucht wird. Die Position eines markanten

Punktes im Bild wird in der Regel im Zuge der Korrespondenz nicht verändert. Umgekehrt ist die Situation bei der intensitätsbasierten Zuordnung. Die Zuordnung kleiner Bildausschnitte führt zur Bestimmung von Transformationsparametern, mit denen die Bildausschnitte verschiedener Aufnahmezeitpunkte geometrisch aufeinander angepaßt werden. Für einen markanten Punkt ist damit wie für jeden beliebigen Punkt innerhalb des zugeordneten Bildausschnittes ein Verschiebungsvektor festgelegt, der zur korrespondierenden Punktlage im Nachbarbild zeigt. Hier wird also durch die Korrespondenz die Lage des Partnerpunktes im zweiten Bild definiert. Das erste Experiment gilt der Verfolgung des für die merkmalsbasierte Zuordnung benötigten FOE. Darüberhinaus wird in weiteren Experimenten eine Stichprobe markanter Punkte verfolgt. Hinter den Genauigkeits-, Zuverlässigkeits- und Effizienzbetrachtungen steht auch die Konkurrenzsituation von intensitätsbasierter und merkmalsbasierter Zuordnung.

5.3.1 Vorgehensweise

Bei der intensitätsbasierten Punktverfolgung sollen die in der Bildfolge lokalisierten markanten Punkte von Bild zu Bild übertragen werden. Daß markante Punkte mit ihrer lokalen Umgebung aus theoretischer Sicht für die Zuordnung besonders geeignet sind, ist schon bei der Diskussion der Punktefinder (Kapitel 4) hervorgehoben worden. Auch experimentell hat sich die auf markante Punkte gestützte intensitätsbasierte Bildzuordnung bewährt. In Untersuchungen zur Erfassung von Digitalen Höhenmodellen aus Stereobildern wurde eine Genauigkeitssteigerung und eine Verringerung von Zuordnungsfehlern sowohl gegenüber merkmalsbasierten Verfahren nachgewiesen (Hahn und Förstner, 1988), als auch gegenüber intensitätsbasierten Verfahren, die an vordefinierten Gitterpunkten zuordnen (Baltsavias, 1991). In den Grundzügen wurde diese Vorgehensweise schon in einer der frühen Arbeiten zur Punktverfolgung in Stereobildsequenzen vorgeschlagen. Moravec (1979, S.599) hat mit dem nach ihm benannten Operator (4-2) markante Punkte extrahiert und durch Korrelation in das nachfolgende Bild übertragen.

Als Verfahrensablauf zur Punktverfolgung kommen mehrere Varianten in Betracht. Allen Varianten ist gemein, daß durch die Extraktion markanter Punkte geeignete Ausgangspositionen in den Bildern für die Punktverfolgung festgelegt sind. Variante 1 besteht darin, daß durch die intensitätsbasierte Zuordnung ein solcher Punkt ins Folgebild übertragen wird und damit den Bezugspunkt für die nachfolgende Zuordnung festlegt. Sukzessive wird die Übertragung von Bild zu Bild fortgesetzt, solange eine erfolgreiche Zuordnung möglich ist. Bei Variante 2 soll die Zuordnung jeweils zwischen dem Ausgangsbild und dem aktuell verarbeiteten Bild in der Sequenz stattfinden. In Variante 3 wird die simultane Zuordnung mehrer Bilder in Betracht gezogen.

Für die simultane N -Bildzuordnung spricht die Vermutung, daß die Mehrbildzuordnung genauer und vor allem zuverlässiger ist als die Zweibildzuordnung. Die Erweiterung des Zuordnungsansatzes (5-4) auf N Bilder wurde im vorhergehenden Abschnitt bereits angesprochen. Sie führt zu einer entsprechenden Erweiterung der Minimierungsfunktion (5-6). Der zugehörige Schätzprozeß zerfällt allerdings in dieser Formulierung in $N - 1$ Teile, die $N - 1$ unabhängig bestimmten Zweibildzuordnungen entsprechen, so daß sich Variante 3 von den andern beiden Varianten praktisch nicht unterscheidet. Zudem wäre eine N -Bildzuordnung unter Navigationsgesichtspunkten nur befriedigend, wenn durch sequentielle Parameterschätzung eine mit der Erfassung der Aufnahmen fortschreitende Prozessierung gewährleistet ist. Für Variante 1 spricht, daß sich die projizierte Szene in aufeinanderfolgenden Bildern sehr ähnlich darstellt, so daß hier ein einfaches geometrisches Modell für die Anpassung lokaler Bildausschnitte gerechtfertigt ist. Allerdings besteht die Gefahr, daß infolge keiner Zuordnungsfehler bei der sequentiellen Übertragung ein Punkt zu "wandern" beginnt, was sich im Laufe der Sequenz in einem merklichen systematischen Fehler äußern kann. Die sich innerhalb der Sequenz ändernde Auflösung eines abgebildeten Objekts kann zu einer signifikanten Änderung der lokalen Bildstruktur führen und dabei maßgeblich zu systematischen Fehlern beitragen. Große Änderungen der Richtung und Distanz zum Objekt über einen längeren Abschnitt der Sequenz machen die Modellierung in Variante 2 aufwendig. Der prinzipielle Vorteil dieser Variante liegt darin, das sich die Zuordnung immer auf das Ausgangsbild und damit auf die ursprüngliche Punktlage des markanten Punktes bezieht. Allerdings beeinträchtigt die Änderung der lokalen Bildstruktur die Zuordnung bei Variante 2 sehr viel direkter als bei Variante 1, so daß bei die Zuordnung des jeweils verarbeiteten Bildes zum Ausgangsbild eine geringere Erfolgsquote erwartet werden muß. Als Vorgehensweise zur Punktverfolgung wird deshalb in den Experimenten die Variante 1 verwendet.

5.3.2 Voruntersuchungen

Um die markanten Punkte in der Bildsequenz "Rastatt" verfolgen zu können, muß der intensitätsbasierte Zuordnungsansatz (5-5) einen sehr großen Konvergenzradius von ca 50 Pixel vorweisen können. Die Grundidee, daß ein entsprechend hohes Startniveau (Abb. 5.3) gewählt wird und der Zuordnungsprozeß die Optimallösung iterativ findet, stößt an zwei Grenzen: Zum einen muß im stark geglätteten Bild genügend Struktur abgebildet sein, zum anderen impliziert ein hohes Startniveau für den Unschärfeparameter eine weit entfernte Linearisierungsstelle in (5-6). Die erste Grenze läßt sich dadurch überwinden, daß für die Zuordnung ein größerer Bildausschnitt verwendet wird. Da ein größerer Bildausschnitt aber nur auf dem Hintergrund der Näherungwertbestimmung benötigt wird, ist es in Anlehnung an eine Mehrebenenprozedur sinnvoll den Zuordnungsprozeß mit vergrößerter Fensterweite auf einer zweiten Glättungsebene zu starten. Die Glättung beider Bilder entspricht einer Anhebung beider Bilder im Scale-Space, so daß die Schätzprozedur (Prinzipiskizze 5.3) auf dieses Niveau aufsetzt.

Die Untersuchung der zweiten Grenze ist am Beispiel von 3 Bildpaaren in den Abbildungen 5.5 und 5.6 dargestellt. Bildpaar 0/1 repräsentiert einen Durchschnittstyp der Bildpaare, 3/4 zeichnet sich durch eine starke Kontraständerung aus und im Paar 132/133 dominiert eine große horizontale Verschiebung von ca 40 Pixel⁴. Dabei wird nicht der Startwert für β_0 variiert, sondern die beiden Bilder im Scale-Space angehoben, so daß ein fester Startwert für $\beta_0 (= 100)$ ⁵ auf verschiedene Glättungsniveaus $\beta_{1,[2]} = \beta_1 = \beta_2 (= 0, 0.1, 0.5, 2, 5, 50, 100)$ aufsetzt. Zugeordnet werden in diesem Beispiel relativ große Bildausschnitte mit 151×151 Bildpunkten.

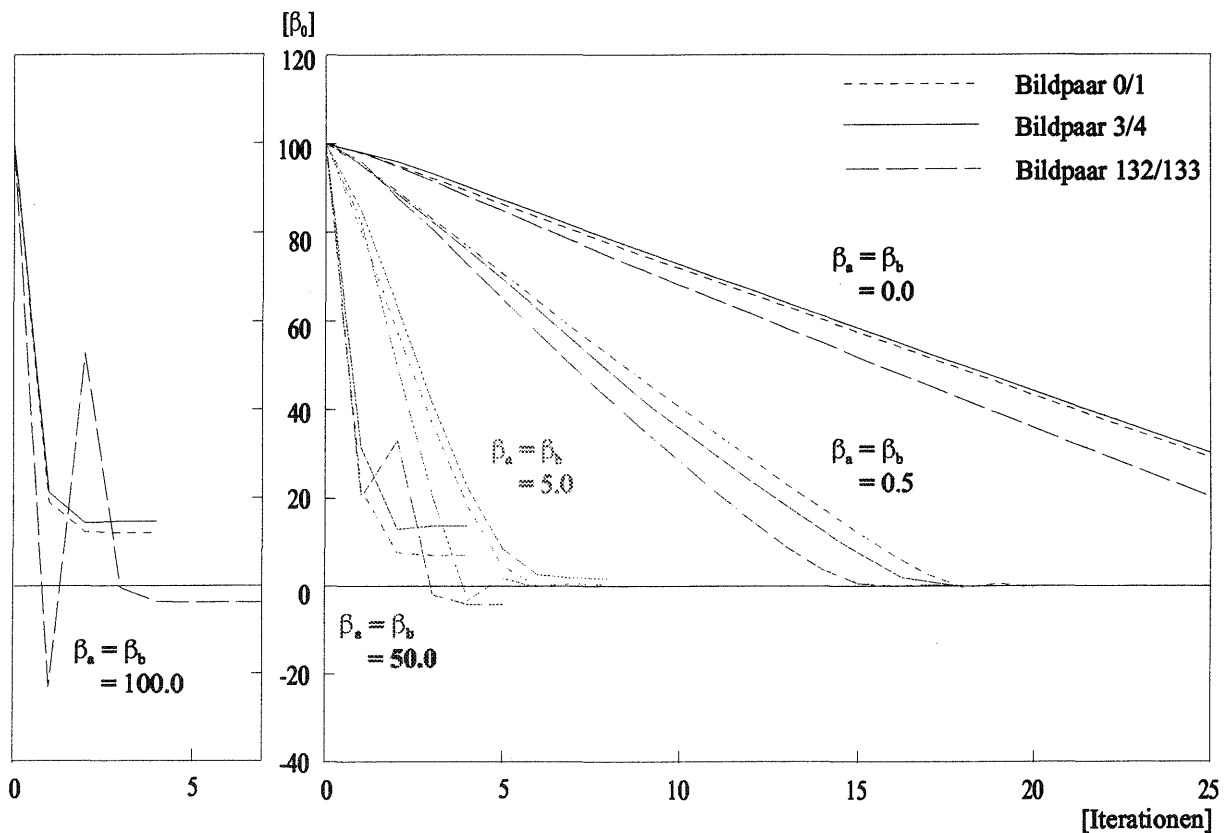


Abbildung 5.5: Konvergenzverhalten für β_0

In Abbildung 5.5 ist das Konvergenzverhalten bezüglich des Parameters β_0 aufgetragen. Bei geringer Glättung

⁴Für nähere Erläuterungen vgl. die Fußnote in Abschnitt 4.3.2.

⁵Die Einheit für β ist (Pixel)². Sie wird im folgenden nur in Ausnahmefällen mit angegeben.

($\beta_{1,[2]}$) der beiden Bilder nimmt β_0 nahezu linear mit der Zahl der Iterationen ab, wobei sehr viele Iterationen benötigt werden. Bei $\beta_{1,[2]} = 5$ werden noch ca 5 Iterationen bis zur Konvergenz benötigt, bei $\beta_{1,[2]} = 100$ genügen 3 bis 4 Iterationen. Für Bildpaar 132/133 zeigt die erste Iteration auf den stark geglätteten Ebenen ($\beta_{1,[2]} = 50$, bzw. 100) ein Überschwingen der β_0 -Schätzung in den negativen Bereich $\beta_1 + \beta_0 < 0$. Das Verfahren wird dann auf das Startniveau zurückgesetzt, wobei lediglich die nach der ersten Iteration ermittelten Verschiebungen als neue Startwerte verwendet werden.

Das entsprechende Iterationsverhalten für die geschätzten Verschiebungen \hat{u} , \hat{v} ist in Abb. 5.6 für das Bildpaar 132/133 aufgetragen. Wie schon bei der β_0 -Schätzung wird die Konvergenz für die Bilder auf höherem Scale-

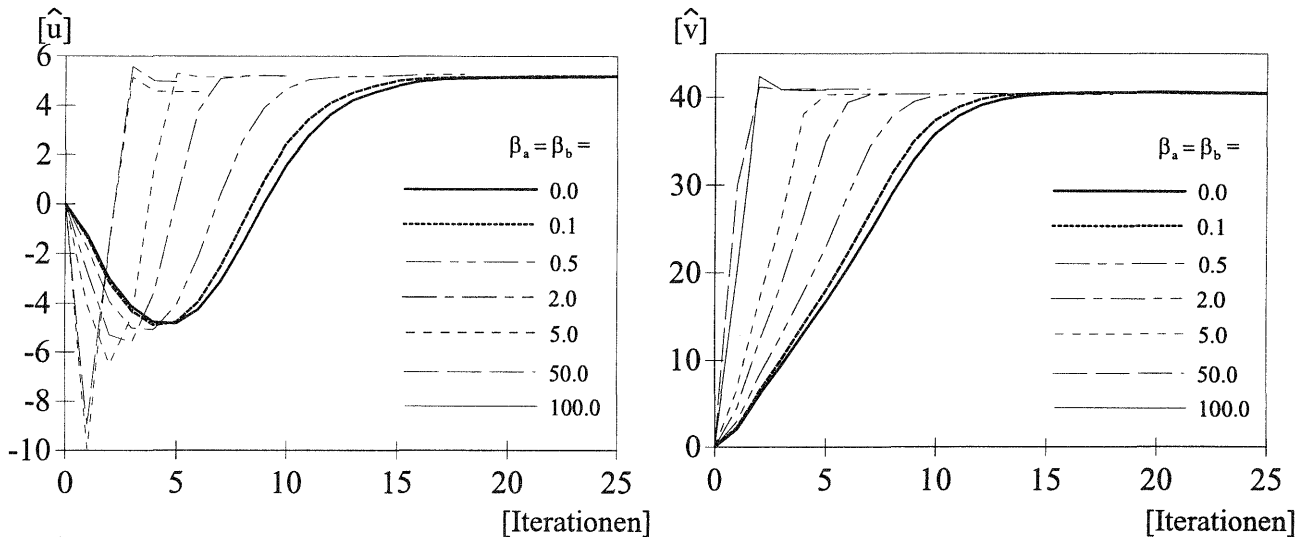


Abbildung 5.6: Konvergenzverhalten für u, v (Bildpaar 132/133)

Space-Niveau deutlich schneller erreicht. Während allerdings für $\beta_{1,[2]} = 0$ nach 20 Iterationen nur noch geringe Verschiebungsänderungen geschätzt werden, ist die β_0 -Schätzung mit einem Wert von ca 40 noch weit von der Konvergenz entfernt. Auf höheren Glättungsebenen $\beta_{1,[2]}$ sind entsprechende Konvergenzunterschiede für Scale-Space- und Geometrieparameter nicht mehr festzustellen. Daß sich die Schätzung in u -Richtung am Anfang des Iterationsprozesses mehrere Pixel von der korrekten Lösung entfernt, zeigt lediglich, daß hierbei die Schätzung des Verschiebungsvektor zunächst von störenden Strukturen beeinflusst ist, dann aber doch die korrekte Lösung erreicht wird. Weitere Experimente mit verschiedenen Fenstergrößen und mit Variationen des Startwertes β_0 sind in Lang (1992) dokumentiert.

Das Konvergenzverhalten (Abb. 5.5, 5.6) ist bei uneglätteten Bildern und hohem β -Wert sehr ungünstig. Um eine schnelle Konvergenz zu ermöglichen, sollte deshalb das Bildpaar zusätzlich auf zumindest ein hohes Scale-Space-Niveau angehoben werden. Auf dem hohen Niveau wird ein großes Zuordnungsfenster benötigt, auf dem niederen Niveau setzt der Zuordnungsansatz für eine möglichst genaue Schätzung ein kleines Fenster voraus. Experimentell hat sich folgendes Verfahren bewährt: Der Startwerte für β_0 soll zumindest gleich groß sein wie der Startwert für $\beta_{1,[2]}$, z. B. 50 oder 100. Ist die β_0 -Schätzung genügend klein (z. B. $1/3$ des Startwertes, was meist schon bei der zweiten Iteration zutrifft), so wird das Scale-Space-Niveau $\beta_{1,[2]}$ auf die Hälfte abgesenkt, dabei auch die Fenstergröße reduziert und der Iterationsprozeß in gleicher Weise wie auf dem höheren Niveau fortgesetzt. Ein typische Niveaufolge ist z. B. $\beta_{1,[2]} = 100, 50, 25, 1$. D. h. von dem unteren Scale-Space-Niveau (25) wird auf eine gering geglättete Originalbildebene ($\beta_{1,[2]} = 1$ oder 0.5) übergegangen, auf der das Zuordnungsergebnis für einen entsprechend kleinen Bildausschnitt erhalten wird. Die Fenstergröße in den verschiedenen Scale-Space-Niveaus kann sich dabei am Glättungsradius $\sqrt{\beta_{1,[2]}}$ orientieren; z. B. wird mit Fensterweiten von 100, 70, 50, 20 Pixeln der 10-fache Glättungsradius angehalten.

5.3.3 Ergebnisse und Diskussion

Das erste Experiment besteht in der Verfolgung des FOE. Der FOE wird zur merkmalsbasierten Zuordnung benützt (Abschnitt 4.3.1). Durch die Bildzuordnung wird der FOE in das nachfolgende Bild übertragen, wobei der geschätzte Verschiebungsvektor die Sensorrotation zwischen den beiden Aufnahmezeitpunkten approximiert. Nach der Rekonstruktion der äußeren Orientierung der Bilder (Kapitel 6) kann der FOE als Schnittpunkt der Geraden durch aufeinanderfolgende Projektionszentren mit der zugehörigen Bildebene geometrisch festgelegt werden. Die Positionsänderung dieses Punktes infolge der Rotation zwischen den Aufnahmen wird im Experiment als Vergleichsgröße für die in der Bildzuordnung geschätzten Verschiebung herangezogen. Ziel des Experiments ist es, sowohl Erfahrungen mit dem neuen Bildzuordnungsverfahren zu gewinnen, als auch nachzuweisen, daß es sich zur Verfolgung des FOE eignet.

In Version A werden Fensterweiten von 151, 101, 51, 15 bei Glättungsniveaus $\beta_{1,[2]}$ von 64, 16, 4, 1 gewählt, d. h. der Glättungsradius wird von Niveau zu Niveau halbiert. Dieselben Glättungsniveaus werden auch bei den anderen Versionen verwendet. In Version B werden die Fensterweiten 151, 75, 37, 19 vorgegeben. Version C verzichtet auf die kleine Fenstergröße und verwendet zur Zuordnung die Fensterweiten 151, 101, 51. In Variante D beschränkt sich die Zuordnung auf zwei Fensterweiten von 151 und 51 Pixel. Als Startniveau für β_0 wird 192 festgelegt, d. h. mit $\beta_1 + \beta_0 = 256$ ist der Glättungsradius von I_1 doppelt so groß wie der Glättungsradius von I_2 . Die Ergebnisse der FOE-Zuordnung über die gesamte Bildsequenz von 157 Bildern sind in Tabelle 5.1 aufgelistet.

| Version | Iterationen pro Bildpaar | Anz. Diff. > 4 Pixel | s_u | s_v | s_β | ρ | $\hat{\sigma}_u$ | $\hat{\sigma}_v$ | $\hat{\sigma}_\beta$ |
|---------|-----------------------------|-------------------------|-------|-------|-----------|--------|------------------|------------------|----------------------|
| A | 20.2 | 5 | 1.22 | 2.45 | 0.45 | 0.89 | 0.070 | 0.062 | 0.104 |
| B | 14.0 | 6 | 1.19 | 2.77 | 0.35 | 0.91 | 0.052 | 0.048 | 0.091 |
| C | 10.9 | 0 | 0.61 | 0.49 | 0.28 | 0.98 | 0.022 | 0.016 | 0.037 |
| D | 12.2 | 1 | 0.64 | 0.66 | 0.17 | 0.99 | 0.021 | 0.017 | 0.040 |

Tabelle 5.1: Verfolgung des FOE in der Bildsequenz

s_u, s_v, s_β sind als Wurzel der mittleren quadratischen Abweichung berechnet worden, wobei die Differenz der in der Zuordnung bestimmten Verschiebungen zu den aus der äußeren Orientierung bestimmten "Sollwerten" bestimmt wird. Für β_0 wird als Sollwert 0 erwartet. Die Anzahl der eine Schwelle von 4 Pixel überschreitenden Differenzen sind zusätzlich angegeben. $\hat{\sigma}_u, \hat{\sigma}_v, \hat{\sigma}_\beta$ geben die in der Bildzuordnung berechneten theoretischen Genauigkeiten an.

Aus der Tabelle 5.1 können folgende Schlußfolgerungen gezogen werden. Der Versuch mit kleineren Zuordnungsfenstern (Versionen A, B) eine genauere Zuordnung zu erzielen gelingt nicht, wofür vor allem Zuordnungsfehler verantwortlich sind. Die visuelle Kontrolle zeigte, daß in diesen Fällen entweder die Texturierung des Bildausschnittes um den FOE gering war, oder aber Objekte aus näherer Entfernung abgebildet waren. Ersatzweise wurde deshalb in der Umgebung des FOE (innerhalb 40 Pixel Radius) ein markanter Punkt gesucht und die Verschiebungsschätzung in diesem Punkt vorgenommen. Die Fehlerquote in Version A und B hat sich dadurch auf drei reduziert und die Differenzen s_u, s_v konnten ungefähr halbiert werden. Der zur Kontrolle mitbestimmte Korrelationskoeffizient zeigt Werte um 0.9 bzw. 0.99, die nur bei deutlichen Fehlzuordnungen geringer ausfallen. Die Unterschiede sind direkt von der Größe des Zuordnungsfensters beeinflusst.

Die Zuordnungsergebnisse von Version C und D unterscheiden sich nur geringfügig. Die Verfolgung des FOE durch die Sequenz gelingt mit einer Genauigkeit von ca. 1 Pixel. Daß die bei der Parameterschätzung ermittelten theoretischen Genauigkeiten um einen Faktor 20 - 30 zu optimistisch sind, entspricht der üblichen Größenordnung, die schon aus vielen anderen Untersuchungen zur intensitätsbasierten Zuordnung bekannt ist. Interessant ist aber der Vergleich von $\hat{\sigma}_u, \hat{\sigma}_v$ und $\hat{\sigma}_\beta$. Die theoretische Schätzgenauigkeit der Verschiebung u, v ist etwa um den Faktor 2 besser als die von β_0 , womit die Vermutung bei der theoretischen Betrachtung zur Genauigkeit in Abschnitt 5.2.3 bestätigt wird. Zusammenfassend kann festgestellt werden, daß sich in der FOE-Verfolgung die intensitätsbasierte Zuordnung im Scale-Space bewährt hat. Auch bei Verschiebungen von mehr als 40 Pixeln

gelingt die Zuordnung mit einer hinreichenden Genauigkeit von ca. 1 Pixel. Für die Zuordnung nach Version C werden im Mittel 10.9 Iterationen benötigt. Daß diese hohe Effizienz der grob-zu-fein Prozessierung sehr konstant über die ganze Sequenz erzielt wird, ist aus dem Histogramm (Abb. 5.7) ersichtlich.

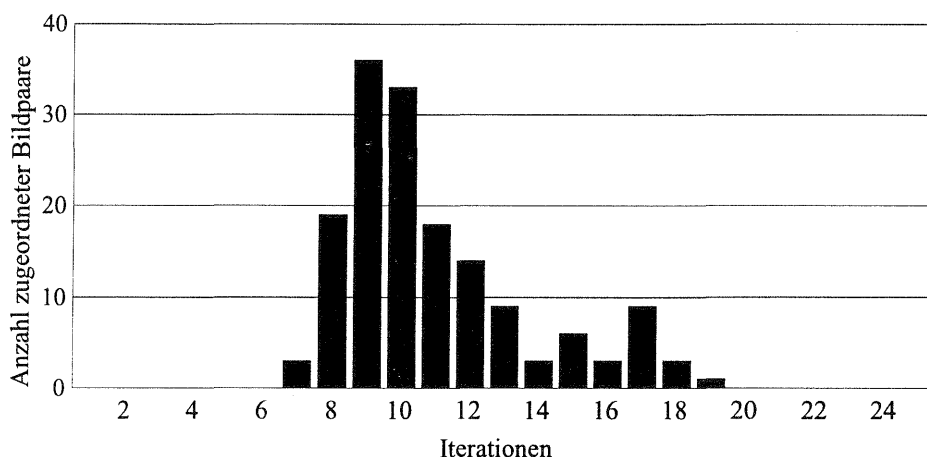


Abbildung 5.7: Histogramm: Aufgetragen ist die Anzahl der Zuordnungen über der vom Zuordnungsverfahren benötigten Anzahl Iterationen

Das zweite Experiment gilt der individuellen Verfolgung von Einzelpunkten. Die Information aus der FOE-Verfolgung wird als Vorinformation berücksichtigt, so daß in der Verschiebungsschätzung insbesondere der translatorische Einfluß der Sensorbewegung ermittelt werden muß. Für die weit entfernten Punkte beträgt die verbleibende Verschiebung nur wenige Pixel, für nahe Objektpunkte 10-20 Pixel. Zur Zuordnung werden wie bei der Verfolgung des FOE höhere Glättungsniveaus verwendet. Bei einer Verschiebung von ca. 10 Pixel genügt eine Anhebung auf ein Scale-Space-Niveau von $\beta_{1,[2]} = 20, \beta_0 = 20$. Die Zuordnung erfolgt mit einer Fenstergröße von 19×19 Pixeln. Ist das Verfahren nicht erfolgreich, so wird die Zuordnung wiederholt, wofür das Zuordnungsfenster auf dem oberen Scale-Space-Niveau auf 51×51 Pixel vergrößert wird. Diese Option soll die Verschiebungsschätzung für eine geringe Anzahl von Punkte ermöglichen, die nahe der Kamera sind und bis zu 25 Pixel verschoben sein können (vgl. Abschnitt 1.7.3 und die ersten beiden Bilder in Abb. 4.6).

Um die manuelle Einzelprüfung bei der Punktverfolgung in Grenzen zu halten, wird die Verfolgung auf eine Stichprobe von 69 Punkten beschränkt. Darunter sind 32 Punkte, die als Kontrollpunkte bei den geodätischen Messungen in Objektraum koordiniert wurden. Unter den weiteren 37 Punkten sind die 6 markanten Punkte, die als Einzelbeispiele bei der merkmalsbasierten Punktverfolgung exemplarisch herausgegriffen und diskutiert wurden. Alle anderen markanten Punkte bilden eine zufällige Stichprobe aus der Menge aller markanten Punkte in der Sequenz. Alle Punkte sind bei der Rekonstruktion durch den 3D-Kalman-Filter (Kapitel 6) im Objektraum koordiniert worden. Außerdem wurde eine simultane Bündelblocktriangulation berechnet, die zur Genauigkeitsabschätzung für die verfolgten Punkte herangezogen wird.

Die Zuordnung ist am Beispiel eines Kontrollpunktes in Abbildung 5.8 dargestellt. Diese Markierung wurde bei allen Kontrollpunkten verwendet und läßt eine gute Zuordnung erwarten. Der Übergang vom unscharfen Niveau zum dem gering geglätteten Bildniveau ist in der jeweils rechten Spalte, die Verschiebungsschätzung aus dem Vergleich der jeweils linken Spalte der Abbildung erkennbar. Die Bildausschnitte der jeweils linken Spalten zeigen das ungeglättete Bild in der Größe des Zuordnungsfensters von 19×19 Pixel.

Das Gesamtergebnis der Verfolgung dieser Kontrollpunkte ist in Tabelle 5.2 zusammengestellt. s_x, s_y sind die rms-Werte, in denen die Abweichungen der in der Punktverfolgung bestimmten Positionen zu denen über die geodätische Koordinierung (Bündelausgleich zur Bildorientierung eingeschlossen) bestimmten Referenzposition der Punkte in den Bildern ermittelt wurde. Ganz offensichtlich sind einige dieser Kontrollpunkte nicht fehlerfrei verfolgt worden, ohne daß dies durch die Selbstdiagnose des Zuordnungsverfahrens (Korrelationskoeffizient einer erfolgreichen Zuordnung muß größer als 0.75 sein, der in beiden zugeordneten Bildausschnitten berechnete (mit

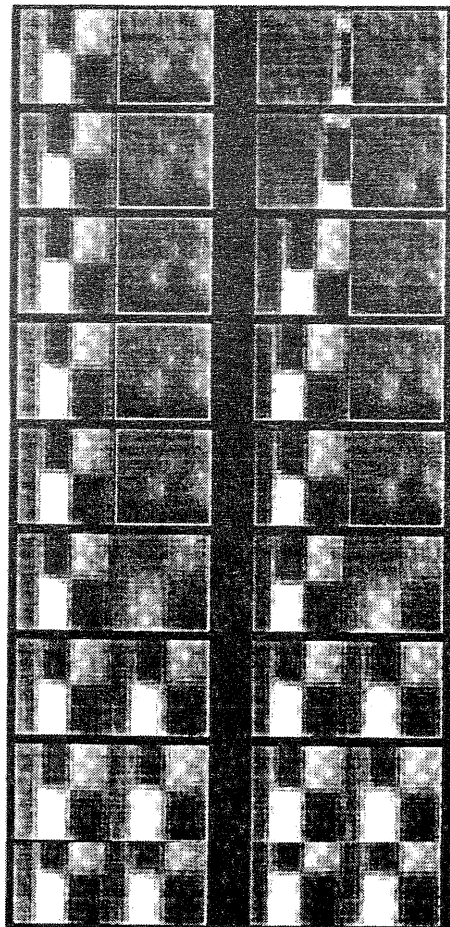


Abbildung 5.8: Iterationsverhalten im Scale-Space. Kontrollpunkt 5032, Bildpaar 35/36. Der unglättete Bildausschnitt (jeweils die linke Spalte) zeigt die jeweiligen Verschiebungen.

den Gradientenquadraten) gewichtete Schwerpunkt darf von der geschätzten geometrischen Verschiebung nicht mehr als $1/10$ Pixel abweichen) erkannt wird. Am auffälligsten ist Punkt 5076, bei dem die Fehlzurordnung ungewöhnlich groß ist. Daß überhaupt ein solcher Zuordnungsfehler möglich ist, liegt daran, daß auf dem Weg von "grob-zu-fein" der Zuordnungsprozeß durch einen Schätzfehler auf höherem Glättungsniveau fehlgeleitet wird. Andere Abweichungen, wie z. B. die der Punkte 5037 und 5048, sind nur bedingt als Zuordnungsfehler interpretierbar. In beiden Fällen stört ein Objekt aus dem Vordergrund den Zuordnungsprozeß über zwei oder drei Bilder der Sequenz, was zu einer Abweichung des Verschiebungsvektors von der korrekten Lösung führt.

Zusammenfassend zeigt Tabelle 5.2, daß bei der Zuordnung von 1002 Bildpaaren rms-Differenzen von 0.42 in x und 0.47 in y beobachtet wurden (der sehr grobe Fehler in Kontrollpunkt 5076 wurde dabei nicht einbezogen). D. h. die intensitätsbasierte Verfolgung von Punkten über die Sequenz gelingt mit einer Genauigkeit von ca $1/2$ Pixel für den Verschiebungsvektor. Wird als erreichbares Genauigkeitsniveau der intensitätsbasierten Zuordnung von Stereobildpaaren in gut texturierten Regionen ca. $1/10$ Pixel zugrunde gelegt, bedeutet dies, daß sich bei der Verfolgung über die Sequenz die Genauigkeit um einen Faktor 5 verringert.

Die Stichprobe der markanten Punkte zeigt ein ähnliches Verhalten wie die Kontrollpunkte. Einzelne Punkte, wie der Punkt 1156 (Abb. 4.8) werden noch über einige Bilder weiter verfolgt, bis sie meist das Gesichtsfeld der Kamera verlassen. In Problempunkten, wie z. B. bei Punkt 1282 (Abb. 4.10), ist die intensitätsbasierte Punktverfolgung allerdings ebenfalls nicht erfolgreich. Im Mittel ist die Zuordnung bei den markanten Punkten um ca. 50% genauer als bei den Kontrollpunkten. Hier wirkt sich aus, daß ein markanter Punkte durch seine deutliche isotrope Texturierung für die Zuordnung potentiell geeigneter ist als ein Kontrollpunkt. Die spezielle

| Kontrollpunkt | Anz. Bildpaare | s_x | s_y | max. Abweich. x/y | Anz. Zuord. mit | |
|---------------|----------------|-------|-------|---------------------|-----------------|-------------|
| | | | | | $ x/y > 0.5$ | $ x/y > 1$ |
| 5020 | 6 | 2.15 | 0.74 | 5.1 | 3 | 1 |
| 5021 | 6 | 0.48 | 0.24 | -0.9 | 3 | - |
| 5029 | 19 | 0.29 | 0.24 | 0.8 | 3 | 1 |
| 5031 | 17 | 0.25 | 0.14 | 0.8 | 1 | - |
| 5032 | 35 | 0.19 | 0.18 | -0.6 | 3 | - |
| 5033 | 13 | 0.17 | 0.18 | -0.4 | 0 | - |
| 5034 | 23 | 0.89 | 1.61 | -5.3 | 3 | 3 |
| 5035 | 49 | 0.24 | 0.18 | -0.7 | 4 | - |
| 5036 | 5 | 0.99 | 0.89 | 1.6 | 4 | 2 |
| 5037 | 67 | 0.93 | 1.08 | 8.7 | 30 | 10 |
| 5038 | 31 | 0.46 | 0.22 | 1.4 | 9 | 2 |
| 5040 | 40 | 0.18 | 0.14 | -0.4 | 0 | - |
| 5041 | 43 | 0.25 | 0.22 | 0.7 | 6 | - |
| 5042 | 51 | 0.45 | 0.95 | 6.7 | 2 | 2 |
| 5048 | 56 | 0.47 | 1.80 | 10.7 | 17 | 4 |
| 5049 | 6 | 0.31 | 0.26 | -0.6 | 1 | - |
| 5060 | 15 | 0.23 | 1.42 | 5.3 | 4 | 1 |
| 5061 | 24 | 0.21 | 0.23 | 0.6 | 1 | - |
| 5062 | 27 | 0.22 | 0.26 | -0.9 | 1 | - |
| 5063 | 15 | 0.26 | 0.16 | 0.5 | 1 | - |
| 5064 | 26 | 0.95 | 0.57 | 4.8 | 4 | 2 |
| 5065 | 42 | 0.22 | 0.16 | 0.6 | 1 | - |
| 5066 | 38 | 0.20 | 0.19 | 0.5 | 1 | - |
| 5067 | 46 | 0.38 | 0.56 | -1.7 | 11 | 4 |
| 5068 | 56 | 0.20 | 0.20 | 0.6 | 2 | - |
| 5070 | 53 | 0.28 | 0.33 | -0.8 | 11 | - |
| 5071 | 59 | 0.25 | 0.34 | -1.9 | 9 | 1 |
| 5072 | 58 | 0.22 | 0.17 | 0.5 | 3 | - |
| 5073 | 8 | 0.19 | 0.25 | -0.6 | 1 | - |
| 5074 | 32 | 0.26 | 0.31 | 1.1 | 7 | 1 |
| 5076 | 12 | 3.96 | 14.88 | -51.5 | 3 | 3 |
| 5077 | 24 | 0.30 | 0.20 | -0.7 | 1 | - |

Tabelle 5.2: Verfolgung der Kontrollpunkte in der Bildsequenz

Wahl der Signalisierung der Kontrollpunkte sorgt dafür, daß dieser Unterschied nicht noch größer ausfällt.

Abschließende Bewertung

Bei der intensitätsbasierten Verfolgung von Punkten konnten die Kontrollpunkte im Mittel über 32 Bilder, die Verknüpfungspunkte über 30 Bilder zugeordnet werden. In vielen Fällen ist der Grund für das Ende der Verfolgung, daß ein Punkt aufgrund der Bewegung des Fahrzeugs aus dem Gesichtsfeld der Kamera verschwindet. Wird als Fehlzuordnung eine Abweichung von 1 Pixel von der Sollposition definiert, so zeigt die Kontrollpunktverfolgung 37 Fehler auf 1000 Zuordnungen (vgl. Tabelle 5.2). Die Fehlerquote des intensitätsbasierten Verfahrens von 4% ist damit deutlich geringer als die Fehlerquote der merkmalsbasierten Verfolgung, die ca. 30% (Abschnitt 6.3.3) beträgt. Daß die verfolgten Punkte sehr viele Bilder verknüpfen und mit einer Fehlerquote von 4% ein für die Bildzuordnung sehr zuverlässiges Ergebnis gefunden wird ist, ein Vorteil, der bei der Rekonstruktion im Kalman-Filter besonders in Gewicht fällt.

Das in den Experimenten erzielte Genauigkeitsniveau von 1/2 Pixel liegt über der in vielen Untersuchungen zur

intensitätsbasierte Stereozuordnung experimentell gefundenen Größenordnung von $1/10$ Pixel. Eine Ursache ist darin zu sehen, daß aufgrund der fortgesetzten Übertragung eines Punktes von Bild zu Bild ein Zuordnungsfehler sich verstärken und dabei zu einem systematischen Fehler werden kann. Dabei spielt eine große Rolle, daß die Betrachtungsrichtung zum Objekt sich ändert, daß sich der lokale Bildmaßstab und damit die Auflösung eines abgebildeten Objekt ändern und dies damit einhergeht, daß sich die projizierte lokale Struktur wandeln kann. Die starke räumliche 3D-Gliederung einer Szene, wie sie bei der Bildfolge Rastatt vorliegt, schließt Verdeckungen und andere Effekte bei der Punktverfolgung nicht aus (siehe die Beispiele bei der merkmalsbasierten Punktverfolgung). Die Folge davon ist, daß Punkte nicht weiter verfolgt werden. Aber auch eine verminderte Genauigkeit und Fehlzusordnungen können hiervon herrühren.

Betrachtet man das Ergebnis der intensitätsbasierten Zuordnung und Verfolgung insgesamt, so hat sich das in diesem Kapitel präsentierte neue Verfahren zur Zuordnung im Scale-Space sehr bewährt. Auch experimentell konnte gezeigt werden, daß das Verfahren einen großen Konvergenzradius besitzt, sehr schnell konvergiert und ein hohes Maß an Zuverlässigkeit vorweist.

Teil III

3D-Rekonstruktion von Bewegung und Struktur

Der Kalman-Filter als systemtheoretisches Schätzverfahren wurde Anfang der sechziger Jahre von Kalman und Bucy entwickelt. Er hat sich sehr schnell in den Ingenieurwissenschaften als Standardverfahren etabliert, z. B. mit Anwendungen in der Satellitennavigation oder auch in der Wasserwirtschaft. Die Attraktivität des Kalman-Filters für Anwendungen im Bereich des Rechner Sehens beruht darauf, daß in dem Verfahren stochastische Messungen, das Wissen über das System und andere Informationen zur Optimalschätzung der geforderten Zustandsparameter einbezogen werden (Zhang und Faugeras, 1992). Darüber hinaus ist oft der Wunsch nach einer rekursiven Echtzeitschätzung des Systemzustandes für die Entwicklung eines Kalman-Filter-Algorithmus ausschlaggebend.

6 Entwurf und Realisierung eines 3D-Kalman-Filters

Wissenschaftlicher Beitrag der Arbeit in diesem Kapitel:

Zur Rekonstruktion von Bewegung und Struktur aus einer Bildsequenz wird ein Kalman-Filter entworfen. Der Zustandsvektor dieses Filters berücksichtigt (a) die Bewegungsparameter einer Zweibildeinheit, die durch die Position und Richtung der Kamera sowie die diskreten rotatorischen und translatorischen Geschwindigkeiten beschrieben werden und (b) die 3D-Struktur der Szene in Form der 3D-Punktkoordinaten der verfolgten Punkte. Der Schlüssel für die rekursive Rekonstruktion aus einer Monosequenz liegt darin, daß die zu den zurückliegenden Zeitpunkten bestimmte Information über die Strukturparameter zum aktuellen Zeitpunkt zur Verfügung steht. Zusammen mit der stochastischen Beschreibung eines kinematischen Bewegungsmodells und den Beobachtungen, d. h. den durch die Punktverfolgung vorliegenden Bildkoordinaten auch ggf. den beobachteten Aufnahmestandpunkten der Kamera, werden durch den Kalman-Filter die verschiedenen Informationen in eine Optimalschätzung einbezogen. In den Experimenten wird die Genauigkeit der rekursiven 3D-Rekonstruktion der Bewegungs- und Strukturparameter untersucht. Die statistische Analyse der Meßdaten zeigt, daß die Fehlerquote der merkmalsbasierten Zuordnung bei 30% liegt, so daß innerhalb des Kalman-Filter-Algorithmus viele Zuordnungsfehler detektiert und eliminiert werden müssen. Für die Rotationsparameter wurde ein Genauigkeitsniveau von 0.1-1 Gon, für die Position der Kamera eine Genauigkeit im dm-Bereich erzielt. Ganz grundsätzlich wird durch das Experiment auch empirisch der Nachweis erbracht, daß der rekursive 3D-Filter zur passiven Navigation geeignet ist.

6.1 Kalman-Filter

Kurzgefaßt kann ein Kalman-Filter als Algorithmus definiert werden, mit dem die zeitvariablen Zustände eines Systems rekursiv aus Meßdaten optimal geschätzt werden. Dabei werden sowohl Fehler in den Messungen, Fehler in der Modellierung des Systems, als auch Vorinformation über das System bei der Zustandsschätzung qualitativ berücksichtigt (Schilling, 1987). Kriterien für die Optimalität sind die Unverzerrtheit und minimale Fehlervarianz der Schätzwerte. Die Konzepte und Eigenschaften von Kalman-Filtern wurden in der Literatur seit der ursprünglichen Publikation von Kalman (1960) sehr intensiv diskutiert und können Lehrbüchern wie z. B. von Gelb (1974) oder Maybeck (1979) entnommen werden. Sehr ausführlich berichten Salzmann (1988) und Merminod und Rizos (1988) von der Verwendung des Kalman-Filters für Aufgaben der kinematischen Positionierung. Im folgenden sollen die Grundlagen der Kalman-Filterung kurz zusammengefaßt werden. Bei der Rekonstruktion von Bewegung und Struktur wird der Zustand des Systems durch Parameter des entsprechenden Modells beschrieben. Damit wird der Kalman-Filter zu einem Parameterschätzverfahren. Die Beziehungen zu anderen Schätzverfahren wie der Bayesschen, der Maximum a posteriori und der Kleinste-Quadrate-Schätzung

sowie die Bedingungen für die Äquivalenz der Methoden sind in den o.g. Berichten oder auch in Brown und Durrant-Whyte (1989) diskutiert.

Der lineare Kalman-Filter in zeitdiskreter Form kann durch ein *Systemmodell* und ein *Beobachtungsmodell* gemäß

$$\begin{aligned}\mathbf{p}_k &= \Phi_{k,k-1}\mathbf{p}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{l}_k &= A_k\mathbf{p}_k + \mathbf{v}_k\end{aligned}\quad (6-1)$$

dargestellt werden, wobei

| | |
|--------------------|---|
| $k, k-1$ | Index für die Zeit mit $k = 0, 1, 2, \dots$ |
| \mathbf{p}_k | Zustandsvektor, u -dim. Parametervektor der das System beschreibt |
| $\Phi_{k,k-1}$ | $u \times u$ Transitionsmatrix |
| \mathbf{w}_{k-1} | Systemrauschen, u -dim. |
| \mathbf{l}_k | Beobachtungsvektor, n -dim. |
| A_k | $n \times u$ Designmatrix |
| \mathbf{v}_k | Meßrauschen, n -dim. |

Im Systemmodell wurde auf einen deterministischen Kontrollinput, wie er z. B. bei gesteuerten Systemen anfällt, verzichtet.

Das System- und Meßrauschen wird üblicherweise als weißes Rauschen angenommen. Für die Zufallsvektoren¹ $\mathbf{w}_k, \mathbf{v}_k$ gilt dann

$$\begin{aligned}E[\mathbf{w}_k] &= 0, & D[\mathbf{w}_k] &= Q_k \\ E[\mathbf{v}_k] &= 0, & D[\mathbf{v}_k] &= R_k\end{aligned}$$

$E[\cdot]$ und $D[\cdot]$ stehen für den Erwartungswert- bzw. Dispersionsoperator. Ferner wird vorausgesetzt, daß zwischen dem System- und Meßrauschen sowie zwischen den Rauschprozessen verschiedener Zeitpunkte keine Korrelation besteht.

Des weiteren müssen noch die Startwerte für den Zustandsvektor spezifiziert werden. In der Regel wird zum Startzeitpunkt wiederum ein Gaußverteilter Zufallsvektor mit

$$E[\mathbf{p}_0] = \mathbf{p}_0, \quad D[\mathbf{p}_0] = P_{0|0}$$

zugrundegelegt, der ebenfalls mit dem System- und Meßrauschen nicht korrelieren soll.

Die Kalman-Filterung besteht mit der *Prädiktion* und der *Aufdatierung* aus einem zweistufigen Prozeß. Im Prädiktionsschritt (auch als zeitliche Aufdatierung bezeichnet) werden der Zustandsvektor und seine Fehlerkovarianz gemäß

$$\begin{aligned}\hat{\mathbf{p}}_{k|k-1} &= \Phi_{k,k-1}\hat{\mathbf{p}}_{k-1|k-1} \\ P_{k|k-1} &= \Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^T + Q_{k-1}\end{aligned}\quad (6-2)$$

zum Zeitpunkt k vorhergesagt. Mit den Beobachtungen erfolgt dann die Aufdatierung des prädierten Zustandes und seiner Genauigkeit gemäß

$$\begin{aligned}\hat{\mathbf{p}}_{k|k} &= \hat{\mathbf{p}}_{k|k-1} + K_k(\mathbf{l}_k - A_k\hat{\mathbf{p}}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - K_k A_k P_{k|k-1} \\ K_k &= P_{k|k-1} A_k^T (A_k P_{k|k-1} A_k^T + R_k)^{-1}\end{aligned}\quad (6-3)$$

Die Matrix K_k ist die sogenannte Kalman Gainmatrix oder Verstärkungsmatrix. Mit der Indizierung $i|j$ wird zum Ausdruck gebracht, daß in die Schätzung zum Zeitpunkt i alle Informationen bis zum Zeitpunkt j einbezogen sind.

¹In der symbolischen Darstellung von Modellparametern und Zufallsvariablen $\mathbf{w}_k, \mathbf{v}_k, \mathbf{p}_k$ wurde auf eine Unterscheidung verzichtet.

Zur Ableitung des Kalman-Filters kann (6-1) in einem Gauß-Markov Modell dargestellt werden (Salzmann, 1988, Merminod und Rizos, 1988)

$$\begin{aligned} E\left[\begin{pmatrix} \hat{\mathbf{p}}_{k-1|k-1} \\ \mathbf{w}_{k-1} \\ \mathbf{l}_k \end{pmatrix}\right] &= \begin{pmatrix} I & O \\ -\Phi_{k,k-1} & I \\ 0 & A_k \end{pmatrix} \begin{pmatrix} \mathbf{p}^{k-1} \\ \mathbf{p}^k \end{pmatrix} \\ D\left[\begin{pmatrix} \hat{\mathbf{p}}_{k-1|k-1} \\ \mathbf{w}_{k-1} \\ \mathbf{l}_k \end{pmatrix}\right] &= \begin{pmatrix} P_{k-1|k-1} & 0 & 0 \\ 0 & Q_{k-1} & 0 \\ 0 & 0 & R_k \end{pmatrix} \end{aligned} \quad (6-4)$$

In dieser Standardform eines Kleinste-Quadrate-Ansatzes zeigen sich deutlich die zum Zeitpunkt k in die Filterung einbezogenen Daten. In $\hat{\mathbf{p}}_{k-1|k-1}$ und der Kovarianzmatrix $P_{k-1|k-1}$ ist der Systemzustand mit allen Informationen bis zum Zeitpunkt $k-1$ repräsentiert. Das Wissen über das Systemverhalten \mathbf{w}_{k-1} , Q_{k-1} bildet zusammen mit den Beobachtungen \mathbf{l}_k und deren stochastische Beschreibung R_k die anderen beiden Komponenten, die in eine Kleinste-Quadrate-Minimierung eingehen und zur Ableitung von (6-2, 6-3) herangezogen werden können.

In (6-1) ist sowohl das Systemmodell als auch das Beobachtungsmodell linear formuliert. Das Systemmodell, in das die zu rekonstruierenden Bewegungs- und Strukturparameter eingehen, kann im Hinblick auf Navigationsanwendungen oft linear modelliert werden. Für das Beobachtungsmodell gilt dies meist nicht, da die beobachteten Bilddaten in nichtlinearer Beziehung zu den zu bestimmenden 3D-Parameter stehen. Für die Linearisierung bietet sich im Kalman-Filter der prädiizierte Zustandsvektor als Linearisierungsstelle an. Wird am prädiizierten Systemzustand linearisiert, so wird der Prozeß als erweitertes Kalman-Filter bezeichnet. Die Schätzung des aufdatierten Zustandes muß meist dennoch iteriert werden. Hierfür kann ausgehend vom prädiizierten Zustand die jeweilige iterativ verbesserte Schätzung als neue Näherung verwendet werden. Der Prozeß wird dann als iteriertes erweitertes Kalman-Filter (IEKF) bezeichnet.

Durch \mathbf{l}_k^0 soll der aus den nichtlinearen Beziehungen $a(\cdot)$ zu den Zustandsparametern $\hat{\mathbf{p}}_{k|k}^0$ berechnete Näherungswert der jeweiligen Iteration dargestellt werden, wobei

$$\mathbf{l}_k^0 = a(\hat{\mathbf{p}}_{k|k}^0) - A_k \hat{\mathbf{p}}_{k|k}^0.$$

Für die erste Iteration ist $\hat{\mathbf{p}}_{k|k}^0 = \hat{\mathbf{p}}_{k|k-1}$, d. h. das Iterationsverfahren startet am prädiizierten Zustand. Die Beziehung für die Aufdatierung des prädiizierten Zustandes im IEKF lautet

$$\hat{\mathbf{p}}_{k|k} = \hat{\mathbf{p}}_{k|k-1} + K_k(\mathbf{l}_k - \mathbf{l}_k^0 - A_k \hat{\mathbf{p}}_{k|k-1}). \quad (6-5)$$

Innerhalb jeder Iteration werden $\hat{\mathbf{p}}_{k|k}$, \mathbf{l}_k^0 und die Matrizen K_k und A_k neu berechnet. Nach erfolgter Konvergenz kann die Genauigkeit des aufdatierten Zustandes gemäß (6-3) bestimmt werden. Die Prädiktion im IEKF ist identisch mit (6-2), d. h. abgesehen von (6-5) sind die Beziehungen des IEKF äquivalent zu den Gleichungen des linearen Kalman-Filters.

Zur Lokalisierung grober Fehler sollen die Residuen der Beobachtungen herangezogen werden. Die im Beobachtungsmodell (6-1) durch $\mathbf{v}_k = \mathbf{l}_k - A_k \mathbf{p}_k$ definierten Residuen werden in (6-5) gemäß $\hat{\mathbf{v}}_k = \mathbf{l}_k - \mathbf{l}_k^0 - A_k \hat{\mathbf{p}}_{k|k-1}$ explizit berechnet. Die zugehörige Kovarianzmatrix ($A_k P_{k|k-1} A_k^T + R_k$) fällt bei der Berechnung der Gainmatrix K_k mit an, so daß die für einen Ausreißertest benötigten Informationen zur Verfügung stehen. Werden die geschätzten Residuen der einzelnen Beobachtungen mit \hat{v}_k , die aus der zugehörigen Kovarianzmatrix bestimmten Standardabweichungen mit $\sigma_{\hat{v}_k}$ dargestellt, so kann die normierte Verbesserung

$$\hat{v}_k / \sigma_{\hat{v}_k} \sim N(0, 1)$$

zur individuellen Prüfung jeder einzelnen Beobachtung herangezogen werden. Diese Standardvorgehensweise, die auch als data-snooping nach Baarda (1968) bezeichnet wird, haben Teunissen und Salzmann (1988) dahingehend erweitert, daß sie die Residuen mehrerer Epochen zur Prüfung heranziehen. Durch diese globalere Modellprüfung in rekursiver Form lassen sich systematische Einflüsse auf die entsprechende Beobachtung besser nachweisen.

6.2 Entwurf eines 3D-Kalman-Filters

6.2.1 Konzeption des Verfahrens

Die Grundidee zum Entwurf des 3D-Kalman-Filters für die Rekonstruktion von Bewegung und Struktur ist in den Abbildungen 6.1 und 6.2 dargestellt. Werden 3D-Punkte in der Szene als gegeben betrachtet, so besteht die Meßaufgabe im aktuell erfaßten Bild darin, diese Punkte im Bild wiederzufinden und zu lokalisieren. Sind die entsprechenden Punkte bezüglich ihrer Lage im Bild gemessen, können sie dazu herangezogen werden, um die äußere Orientierung des aktuell verarbeiteten Bildes zu bestimmen oder zu verbessern. Bei bekannter Orientierung kann dann die Koordinierung der Punkte im Raum ermittelt oder verbessert werden. Werden die Parameter der äußeren Orientierung des aktuell verarbeiteten Bildes und die Objektkoordinaten der Punkte im Zustandsvektor berücksichtigt, hat die Kalman-Filterung Ähnlichkeiten mit einer rekursiven Bündellösung, in

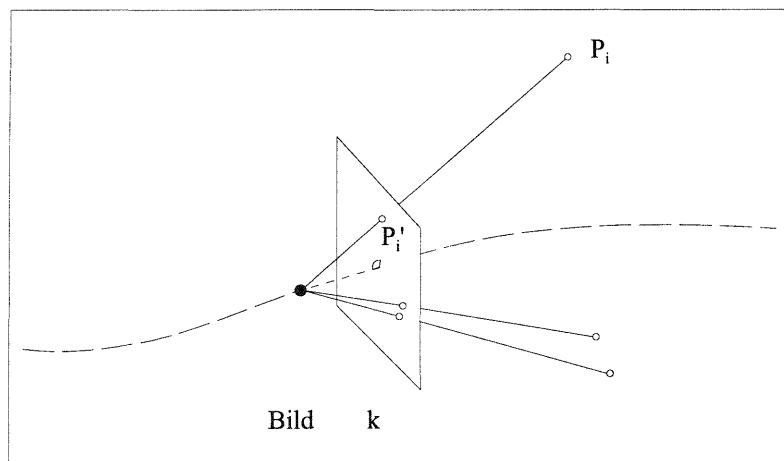


Abbildung 6.1: Systemeinheit "Einzelbild"

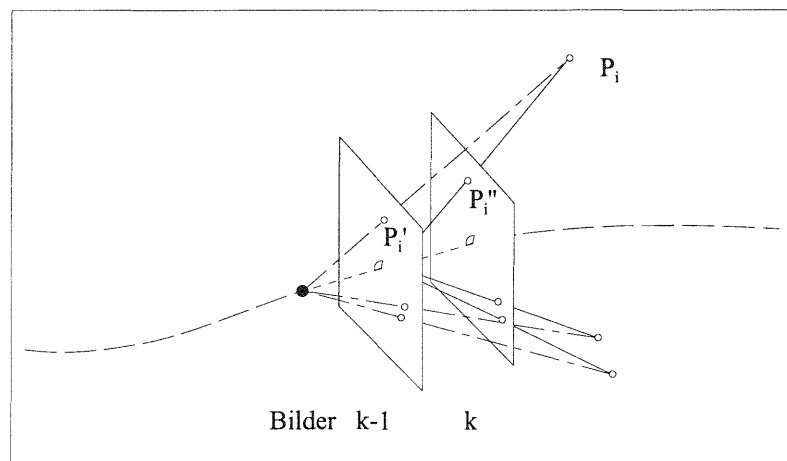


Abbildung 6.2: Systemeinheit "Bildpaar"

der durch die Beobachtungsgleichungen die Beziehungen zwischen den gemessenen Bildkoordinaten der Punkte und den Zustandsparametern hergestellt wird (Abbildung 6.1).

Neben dem Beobachtungsmodell ist gemäß (6-1) ein Systemmodell für den Kalman-Filter zu formulieren. Werden die Objektpunkte auf ein ortsfestes Koordinatensystem bezogen, so können die präzidierten Koordinaten und

die in der Aufdatierung bestimmten Koordinaten der Objektpunkte gleichgesetzt werden. Hierfür wird die Stationarität der Szene (vgl. Kapitel 2) vorausgesetzt. Die Transititionsmatrix, bzw. die auf die Objektkoordinaten beschränkte Submatrix von $\Phi_{k,k-1}$, besteht aus einer Einheitsmatrix.

Die Modellierung der Sensorbewegung soll anhand von Abbildung 6.2 erläutert werden. Der Zustandsvektor der Systemeinheit "Bildpaar" kann durch die Parameter der äußeren Orientierung eines Bildes und durch die Differenzen der Orientierungsparameter zwischen den beiden Bildern parametrisiert werden. Diese Differenzen approximieren die rotatorischen und translatorischen Geschwindigkeiten, wenn sie auf die Taktrate der Bildaufnahme bezogen werden. Genügt ein solches lineares Bewegungsmodell nicht, kann die Systemeinheit auf 3 oder mehr Bilder ausgedehnt werden, d. h. neben den Geschwindigkeiten werden auch Beschleunigungen und höhere Ableitungen der Rotations- und Translationsparameter in das kinematische Bewegungsmodell aufgenommen. Damit ist die Konzeption des Systems zur 3D-Rekonstruktion von Bewegung und Struktur aus Bildfolgen vorgestellt.

6.2.2 Realisierung

Für die experimentellen Untersuchungen wurde folgender Systementwurf algorithmisch realisiert:

1. Im Systemmodell wird die Bewegung des Sensors und die Struktur der Szene berücksichtigt. Die Systemeinheit für die Bewegungsmodellierung ist das Bildpaar. Dabei werden die 12 Parameter der äußeren Orientierung des Bildpaares im Zustandsvektor berücksichtigt. Dies entspricht der Parametrisierung mit 6 Orientierungsparametern und 6 Parametern, in denen die Orientierungsdifferenzen repräsentiert sind. Werden mit \mathbf{m} die 12 Parameter der äußeren Orientierung des Bildpaares bezeichnet, so lautet der Anteil der Bewegungsmodellierung am Systemmodell (6-1)

$$\begin{pmatrix} \mathbf{m}_k(\text{Bild } k) \\ \mathbf{m}_k(\text{Bild } k-1) \end{pmatrix} = \begin{pmatrix} 2I_6 & -I_6 \\ I_6 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{m}_{k-1}(\text{Bild } k-1) \\ \mathbf{m}_{k-1}(\text{Bild } k-2) \end{pmatrix}.$$

Das Systemrauschen berücksichtigt dabei die Abweichung der tatsächlichen Bewegung des Sensors von dem linear approximierten kinematischen Bewegungsmodell.

Die Struktur der Szene wird durch die Objektkoordinaten der ν Verknüpfungspunkte im Zustandsvektor repräsentiert. Die Übertragung auf den nächsten Bezugszeitpunkt erfolgt durch eine Einheitsmatrix, d. h. der auf die Objektmodellierung bezogene Anteil am Systemmodell (6-1) lautet

$$\mathbf{U}_k = I_{3\nu} \mathbf{U}_{k-1}.$$

Zusätzliches Systemrauschen ist für die Objektkoordinierung nicht vorgesehen.

2. Das Beobachtungsmodell beschreibt die Beziehung zwischen den in Bild k gemessenen Koordinaten der Bildpunkte und den Systemparametern. Der Ansatz ist in Kapitel 2 ausführlich hergeleitet und diskutiert worden. Das linearisierte Modell lautet gemäß (2-50):

$$d\mathbf{x}_i = (L V)_i \omega - (L R^T)_i d\mathbf{U}_0 + (L R^T)_i d\mathbf{U}_i$$

Die beiden beobachteten Bildkoordinaten \mathbf{x}_i eines Punktes sind als Funktion der 6 Parameter der äußeren Orientierung (ω bzw. $q(\omega)$, \mathbf{U}_0) und der 3 Koordinaten des Punktes (\mathbf{U}_i) dargestellt. Die Koordinaten \mathbf{x}_i gehen in den Beobachtungsvektor \mathbf{I}_k ein, die Größen ω , \mathbf{U}_0 und \mathbf{U}_i sind im Zustandsvektor \mathbf{p}_k repräsentiert. Die Belegung der Designmatrix A_k ist durch die 2×3 Blockmatrizen $(L V)_i$ und $(L R^T)_i$ (siehe Abschnitt 2.4.3) charakterisiert.

In den Experimenten werden 4 Varianten des 3D-Filters untersucht. Dabei wird das Systemrauschen auf zwei verschiedene Arten approximiert (siehe Tabelle 6.1).

Als zusätzliche Beobachtungsgruppe werden direkt beobachtete Projektionszentren \mathbf{U}_0 mit ihrer stochastischen Beschreibung bei drei Varianten im Beobachtungsmodell berücksichtigt. Hierbei ist der Gewinn für die Kalman-Filterung Gegenstand der Untersuchung, der durch diese Beobachtungen erzielt wird. Dies erfolgt vor dem Hintergrund, daß potentiell zur Positionierung einer Sensorplattform GPS herangezogen werden kann.

Weitere Gesichtspunkte, die für die Realisierung des 3D-Kalman-Filters von Bedeutung sind, betreffen:

1. Die Startwerte für den Zustandsvektor. Diese werden durch eine Bündelausgleichung für das erste Bildpaar berechnet, wobei die äußere Orientierung und die Objektkoordinaten der Verknüpfungspunkte bestimmt werden. Durch interaktiv gemessene Paßpunkte wird in der Bündellösung die Beziehung zum ortsfesten Koordinatensystem hergestellt.
2. Da im Lauf der Sequenz Verknüpfungspunkte aus dem Gesichtsfeld der Kamera verschwinden oder bei der Bildzuordnung nicht mehr erfolgreich weiterverfolgt werden, sollen solche Punkte aus dem Zustandsvektor ausgeschieden werden. Diese Punkte über die Sequenz im Kalman-Filter mitzuführen ist deshalb nicht sinnvoll, da sie zur Schätzung der 3D-Bewegung des Sensors nicht mehr beitragen. Stattdessen werden die bei der Messung im Laufe der Sequenz neu extrahierten und verfolgten Punkte in den Zustandsvektor eingefügt.

Einfügen und ausscheiden von Objektpunkten stört den für die Kalman-Filterung typischen rekursiven Prozeß und führt zu einer Suboptimalität des Filters. Für die Prozessierung langer Sequenzen ist aber wichtig, daß über die Sequenz immer eine genügend große Anzahl von Verknüpfungspunkten im Systemzustand repräsentiert wird. Im Gegensatz zu sequentiellen Verfahren der Bildtriangulation (Grün, 1985b, Grün und Kersten, 1992) ist die Dimension des Systemzustandes im Kalman-Filters über die Sequenz nahezu konstant. Sequentielle Verfahren der Bildtriangulation, bei denen der Systemzustand sich mit jedem Bild vergrößert, führen bei langen Bildsequenzen zu keiner praktikablen Lösung.

6.3 Experimentelle Untersuchungen

Das primäre Ziel der experimentellen Erprobung des entwickelten 3D-Kalman-Filters besteht im Nachweis, daß das rekursive Rekonstruktionsverfahren für die Verarbeitung langer Realwelt-Sequenzen geeignet ist.² Dabei soll die Genauigkeit der rekonstruierten Bewegungs- und Strukturparameter sowie die Lokalisierung von Fehlern in den Meßdaten untersucht werden. In den Experimenten werden die merkmalsbasiert verfolgten Punkte (Abschnitt 4.3) verwendet.

6.3.1 Kontrollinformation

Durch die geodätischen Messungen und die Netzausgleichung ist die Position der Kamera bei der Aufnahme in der Bildsequenz in einem lokalen Netz bestimmt worden. Ebenso wurden einzelne Kontrollpunkte in der Szene signalisiert und geodätisch koordiniert. Zusammen mit den verfolgten markanten Punkten sind diese Daten zur simultanen Bildtriangulation der gesamten Bildsequenz herangezogen worden. In dieser photogrammetrischen Standardtechnik (einer Bündelblockausgleichung) wurden folgende Eingangsdaten berücksichtigt:

1. Die gemessene Position der Projektionszentren einschließlich deren geodätisch bestimmte Genauigkeit, die besser als 1 cm war. Die Projektionszentren wurden zu Beginn der Sequenz bei jeder Aufnahme, später bei jeder zweiten bzw. dritten Aufnahme erfaßt.
2. Die Kontrollpunkte in der Szene als stochastische Paßpunkte, die einen mittleren Punktfehler von ca. 1 cm aufweisen.
3. Die durch einen Operateur in allen Bildern der Sequenz interaktiv gemessenen Paßpunkte, mit einer Genauigkeit von 0.4 Pixel.
4. Die merkmalsbasiert verfolgten markanten Punkte als Verknüpfungspunkte, für die in der Bildtriangulation eine Genauigkeit der Bildkoordinaten von 1.5 Pixel zugrunde gelegt wurde.

²Wie groß die Mängel der bislang vorgestellten Rekonstruktionsansätze sind, zeigt die Feststellung von Barron et al. (1990), die enttäuscht resümieren, daß "nahezu keine dieser Rekonstruktionstechniken bei der Verarbeitung realistischer Szenen erfolgreich war" (vgl. Abschnitt 1.8).

Als Ergebnis der Bildtriangulation liegen die äußere Orientierung der gesamten Sequenz sowie die Objektkoordinaten aller verfolgten Punkte vor, wobei die sehr genaue geodätische Positionsbestimmung von Projektionszentren zur genauen Bestimmung der gesamten Trajektorie der Kamera beiträgt. Die äußere Orientierung, die die Rotations- und Translationsbewegung des Sensors bezogen auf das lokale geodätische Netz repräsentiert, wird zur Kontrolle der im Kalman-Filter bestimmten Bewegungsparameter herangezogen. Für die Genauigkeitsanalyse einzelner Verknüpfungspunkte steht durch die Kontrollpunkte eine Referenz mit übergeordneter Genauigkeit zur Verfügung. Aufgrund der speziellen Markierung werden viele dieser Kontrollpunkte auch als markante Punkte detektiert, verfolgt und in 3D koordiniert. Durch die Identifikation eines solchen Punktes mit einem Kontrollpunkt, die in den Experimenten vom Operateur vorgenommen wird, ist eine unabhängige Kontrolle der räumlichen Koordinierung in der Szene gewährleistet. Für alle anderen Verknüpfungspunkte kann a priori nicht davon ausgegangen werden, daß die Bündelblockausgleichung der Bildsequenz zu einer unabhängigen und sehr viel genaueren Koordinierung führt. Die Genauigkeitsbetrachtung erfolgt hier mit den durch Fehlerfortpflanzung berechneten theoretischen Genauigkeiten.

6.3.2 Verschiedene Versionen der Filterung

Um einen Einblick in die Qualitätseigenschaften der 3D-Kalman-Filterung zu erhalten, wird die Filterung in verschiedenen Versionen durchgeführt. Diese Versionen unterscheiden sich bezüglich der Gewinnung von Näherungswerten für die Verknüpfungspunkte, bezüglich der stochastischen Modellierung des Bewegungsmodells, der Berücksichtigung der zusätzlich beobachteten Position der Kamera und der Einbeziehung von Paßpunkten. Die vier verschiedenen Varianten sind in Tabelle 6.1 charakterisiert.

| | Version A | Version B | Version C | Version D |
|--|--------------------------------------|----------------------|-----------|---|
| Verknüpfungspunkte $\sigma_{U,V,W}$ [m] | Dist. 300 m 100 | Bündelschnitt 100 | | |
| kinem. Bewegungsmodell $\sigma_{\Delta U_0}, \sigma_{\Delta V_0}, \sigma_{\Delta W_0}$ [m] $\sigma_{\Delta \omega}, \sigma_{\Delta \varphi}, \sigma_{\Delta \kappa}$ [gon] | 0.22, 0.05, 0.01 0.53, 0.36, 0.52 | | | 0.1, 0.1, 0.01 $3\sigma_{\hat{\omega}, \hat{\varphi}, \hat{\kappa}}$ |
| Position der Kamera σ_{U_0, V_0, W_0} [cm] | 5 | – | 5 | – |
| Paßpunkte $\sigma_{U,V,W}$ [cm] | 5 | 5 | – | 5 |

Tabelle 6.1: Versionen der Kalman-Filterung

Für alle Versionen wurde die Meßgenauigkeit der manuell gemessenen Paßpunkte mit 0.4 Pixel, der merkmalsbasiert verfolgten Objektpunkte (Verknüpfungspunkte) mit 1.5 Pixel zugrundegelegt.

Version A unterscheidet sich von den anderen Versionen in der Vorgehensweise für die Gewinnung von Näherungswerten für die Objektkoordinaten der Verknüpfungspunkte. Bei der algorithmischen Entwicklung und Erprobung des Kalman-Filters wurde zunächst versucht, durch relative Orientierung mit anschließender Anpassung auf bereits zum vorhergehenden Aufnahmezeitpunkt verfolgte und in 3D koordinierte Punkte das Näherungswerteproblem zu lösen. Die effektive Basis der vorwiegend translatorischen Bewegung ist allerdings zu klein, so daß dieser Weg nicht erfolgreich war. Die besten Näherungswerte werden durch räumlichen Vorwärtsschnitt gewonnen, wenn die Bildkoordinaten eines verfolgten Punktes in allen verknüpfenden Bildern simultan zur Berechnung herangezogen werden. Dieser Weg (in Tabelle 6.1 als Bündelschnitt bezeichnet) steht allerdings dem Navigationsanspruch nach einer fortlaufenden Rekonstruktion entgegen. In Version A ist deshalb eine für die Navigation geeignetere Möglichkeit vorgesehen. Die gemessene Position eines Punktes im Bild legt im Kamerakoordinatensystem eine Richtung fest, die mit Hilfe der präzidierten äußeren Orientierung in eine im szenefesten Koordinatensystem definierte räumliche Gerade transformiert werden kann. Als Näherungswert für die 3D-Position dieses Punktes wird dann ein Punkt auf dieser Geraden in einer Entfernung von 300 m vom Aufnahmestandpunkt festgelegt. Alle Punkte werden, wenn sie erstmals im Kalman-Filter erscheinen, durch

dieses Vorgehen relativ weit entfernt "in die Szene gelegt". Durch eine a priori Genauigkeitsvorgabe von 100 m wird ein geringer Einfluß auf die Koordinierung dieser Punkte im Kalman-Filter ausgeübt. Wird die Entfernung zum Punkt im Laufe der Sequenz kleiner, so kann der Punkt in der Regel räumlich präziser koordiniert werden.

Für die stochastische Beschreibung des Bewegungsmodells wird in Version D eine grobe Abschätzung vorgenommen. Bezüglich der linear prädierten Position des nächsten Aufnahmestandpunktes wird ein Systemrauschen in der Lage von 1 dm, in der Höhe von 1 cm angenommen. Das Systemrauschen bezüglich der linearen Richtungsprädiktion wurde durch die mit dem Faktor 3 multiplizierte Genauigkeit der in der Filterung bestimmten Winkel des vorhergehenden Bildes approximiert. Alternativ dazu sind für die anderen Versionen die zweiten Differenzen der tatsächlichen Bewegung ermittelt worden, die als Fehler des linearen Bewegungsmodells interpretiert das Fahrverhalten bei der Erfassung der Sequenz recht gut wiedergeben. Um den Faktor 3 verstärkt sind sie als stochastische Beschreibung des Bewegungsmodells in Tabelle 6.1 aufgeführt.

Wird die beobachtete Position der Kamera in den Kalman-Filter einbezogen, so wird hierdurch z. B. eine Messung mit einem GPS-Sensor simuliert, der zusammen mit der Kamera ein integriertes System bildet. Die Kombination von kinematischem GPS und Luftbild hat sich für die Aerotriangulation bewährt und ist z. B. in Frieß (1990) untersucht worden. Die Auswirkungen einer Positionsgenauigkeit des Aufnahmestandpunktes von 5 cm für die Rekonstruktion von Bewegung und Struktur zeigt der Vergleich der Versionen A,C mit B,D.

Durch die Paßpunkte erfolgt die Rekonstruktion im ortsfesten geodätischen System. Dabei werden dieselben 32 Paßpunkte verwendet wie bei der simultanen Bündellösung, die interaktiv in der Bildsequenz gemessen wurden. Da für viele Navigationsaufgaben kein ortsfestes Referenzkoordinatensystem benötigt wird, ist in Version C auf Paßpunkte verzichtet worden. Daß die Lagerung dennoch mit Bezug zum örtlichen geodätischen System erfolgt, wird durch die Berechnung der Startwerte des Zustandsvektors durch Bündelausgleichung der auf das erste Bildpaar bezogenen Messungen und durch die als schwaches Datum eingeführten Verknüpfungspunkte ($\sigma_{U,V,W} = 100$ [m]) gewährleistet, so daß der direkte Vergleich mit den anderen Versionen möglich ist.

Um die Datenmenge in der Kalman-Filterung einzuschränken, wird die Anzahl der verfolgten Punkte heuristisch reduziert. Dies erfolgt durch Elimination der Punkte, die sich im Szenehintergrund befinden. Diese Punkte würden nur zur Bestimmung der rotatorischen Bewegung beitragen. Als Kriterium für die Elimination wird die translatorische Bewegungsparallaxe herangezogen und eine Mindestparallaxe von 5 Pixel gefordert. Da viele Punkte nur wenige Bilder verknüpfen, der Filter aber insbesondere von Punkten profitiert, die sich über einen längeren Abschnitt verfolgen lassen, werden zusätzlich die Punkte eliminiert, die nur 2 oder 3 Bilder verknüpfen. Die Anzahl der verbleibenden und zur Kalman-Filterung herangezogenen Punkte ist in Tabelle 6.2 aufgeführt.

| Punktverfolgung: "K" Punkte erscheinen in genau "L" Bildern | | | | | | | | | | | | | | |
|--|----|----|-----|-----|-----|-----|-----|----|----|----|----|----|----|----|
| L = | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| K = | - | - | 501 | 313 | 251 | 182 | 156 | 84 | 78 | 66 | 41 | 35 | 32 | 44 |
| L = | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| K = | 18 | 15 | 14 | 12 | 17 | 10 | 9 | 7 | 5 | 7 | 5 | 5 | 2 | 2 |
| L = | 30 | 31 | 32 | 33 | 34 | 35 | 37 | 38 | 40 | 41 | 42 | 43 | 44 | 45 |
| K = | 3 | 1 | 2 | 2 | 3 | - | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 |
| Gesamtanzahl: # der Objektpunkte: 1932 # der Bildpunkte: 15290 | | | | | | | | | | | | | | |

Tabelle 6.2: Verfolgte Punkte: Eingangsdaten für die Kalman-Filterung

Im Mittel verbleiben in jedem Bild 120 extrahierte und verfolgte markante Punkte, deren Koordinaten als Meßwerte in den Kalman-Filter eingehen.

6.3.3 Bereinigung von Zuordnungsfehlern

Die interaktive Nachprüfung in Abschnitt 4.3 hat schon gezeigt, daß bei den merkmalsbasierten Punktverfolgung mit einer beträchtlichen Quote an Fehlzuordnungen zu rechnen ist. Zur Fehlerbereinigung wird die

normierte Verbesserung (Abschnitt 6.1) jeder beobachteten Bildkoordinate berechnet. Ein Punkt wird als Ausreißer betrachtet, wenn die geschätzte Verbesserung in einer Punktkoordinate das doppelte der zugehörigen Standardabweichung überschreitet, d. h. es wird die normalverteilte normierte Verbesserung mit einem Signifikanzniveau von ca. 95% getestet. Aufgrund der Fehlerbereinigung sind einige Punkte verblieben, die nur noch drei oder zwei Bilder verknüpfen. Die lediglich zweifach verknüpfenden Punkte wurden in der Kalman-Filterung eliminiert.

Eine Übersicht über die Verknüpfungspunkte nach der Bereinigung von Zuordnungsfehlern ist in Tabelle 6.3 zusammengefaßt. Der Vergleich mit der Tabelle 6.2 zeigt die Fehlerquote in den Bildpunkten. Insgesamt wurden

| Punktverfolgung: "K" Punkte erscheinen in genau "L" Bildern | | | | | | | | | | | | | | |
|--|----|----|-----|-----|-----|-----|-----|----|----|----|----|----|----|----|
| $L =$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| $K =$ | - | 45 | 333 | 222 | 177 | 100 | 106 | 57 | 51 | 38 | 27 | 28 | 24 | 33 |
| $L =$ | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| $K =$ | 10 | 6 | 9 | 14 | 13 | 10 | 8 | 3 | 3 | 8 | 4 | 2 | 1 | 2 |
| $L =$ | 30 | 31 | 32 | 33 | 34 | 35 | 37 | 38 | 40 | 41 | 42 | 43 | 44 | 45 |
| $K =$ | 1 | 2 | 1 | 3 | 1 | 1 | - | 1 | 2 | - | - | 1 | 1 | 1 |
| Gesamtanzahl: # der Objektpunkte: 1349 # der Bildpunkte: 10610 | | | | | | | | | | | | | | |

Tabelle 6.3: Die verfolgten Punkte nach der Fehlerbereinigung im Kalman-Filter

30.6% der Bildpunkte als Ausreißer lokalisiert und eliminiert. Diese Zahl weicht von der Quote der in der Bündellösung robust eliminierten Ausreißer (35%) ab, was aufgrund der höheren Zuverlässigkeit einer Simultanlösung auch erwartet werden konnte.

Daß in dem vorgeschlagenen Kalman-Filter manche Fehler nicht identifiziert und bereinigt werden können, läßt sich an dem in Abbildung 4.9 dargestellten Punkt 3338 aufzeigen. Schon in den ersten beiden Bildern ist ein Zuordnungsfehler erkennbar, der allerdings zu keinem genügend großen Widerspruch bei der Rekonstruktion führt. Einige Bilder später könnte dieser Fehler geometrisch eindeutig lokalisiert werden. Da jeweils nur die Meßdaten des aktuell erfaßten Bildes im Filter verarbeitet werden, ist eine Korrektur der Meßdaten der bereits verarbeiteten Bilder innerhalb der rekursiven algorithmischen Prozessierung nicht mehr in einfacher Weise möglich. Der Übergang der Punktverfolgung im Bild 50 auf das räumlich näherliegende Objekt führt zu keinem Widerspruch, da der Tiefensprung geometrisch nicht hinreichend genau aufgelöst werden kann.

6.3.4 Schätzung der Eigenbewegung

Da für die Parametrisierung der Bewegung des Sensors die Position und Richtung im szenefesten Koordinatensystem verwendet wird, ist die Eigenbewegung in dieser Form äquivalent zur äußeren Orientierung der ganzen Bildsequenz. Für die Beurteilung der einzelnen Versionen bietet es sich deshalb an, eine simultane Bündellösung als Referenz zu verwenden.

Die Rotationsbewegung über die Sequenz (laufende Bildnummer) ist für die Version A und für die Bündellösung aufgetragen. Das lokale geodätische System ist so ausgerichtet, daß die Bewegung des Sensors ungefähr in U -Richtung (Tiefenrichtung) zeigt. Die Bildebene (x, y) ist in erster Näherung parallel zur (V, W) -Ebene. Durch φ ist die Rotation um die W -Achse (Höhe), durch ω um die U -Achse (Längsachse, Tiefe) und durch κ um die V -Achse (Querachse) festgelegt. Der Verlauf der geschätzten Drehwinkel $\hat{\omega}$, $\hat{\varphi}$, $\hat{\kappa}$ ist in Abbildung 6.3 dargestellt. Die größten Änderungen über die Sequenz zeigen sich in $\hat{\varphi}$. Hierin spiegelt sich die Lenkbewegung bei der Aufnahme der Bildfolge wieder. Der sich im Bild 40 (Abbildung 1.5, Abschnitt 1.7.3) abzeichnende Fahrspurwechsel zeigt sich in den ermittelten $\hat{\varphi}$ -Werten zwischen Bildnummer 50 und 65. Auch die Linkskurve am Ende der Sequenz (ab Bild 130) ist noch erkennbar.

Zum Vergleich sind die in der Kalman-Filterung (Version A) geschätzten Rotationswinkel in Abbildung 6.4

dargestellt. Die durch den Filter bestimmte rotatorische Bewegung ist der Bündellösung sehr ähnlich. Selbst kleinere Richtungsänderungen können in beiden Abbildungen wiedergefunden werden.

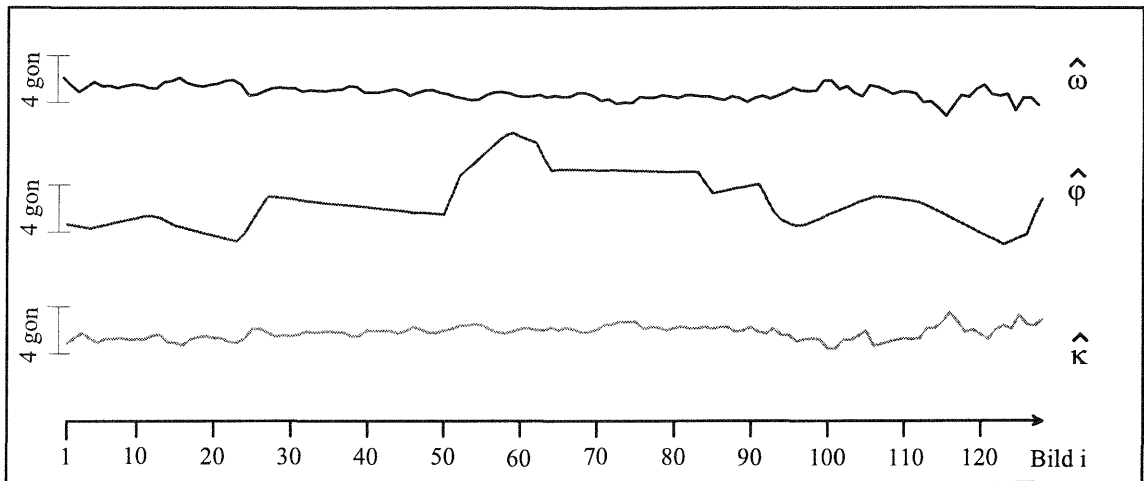


Abbildung 6.3: Bündellösung: Geschätzte Winkel

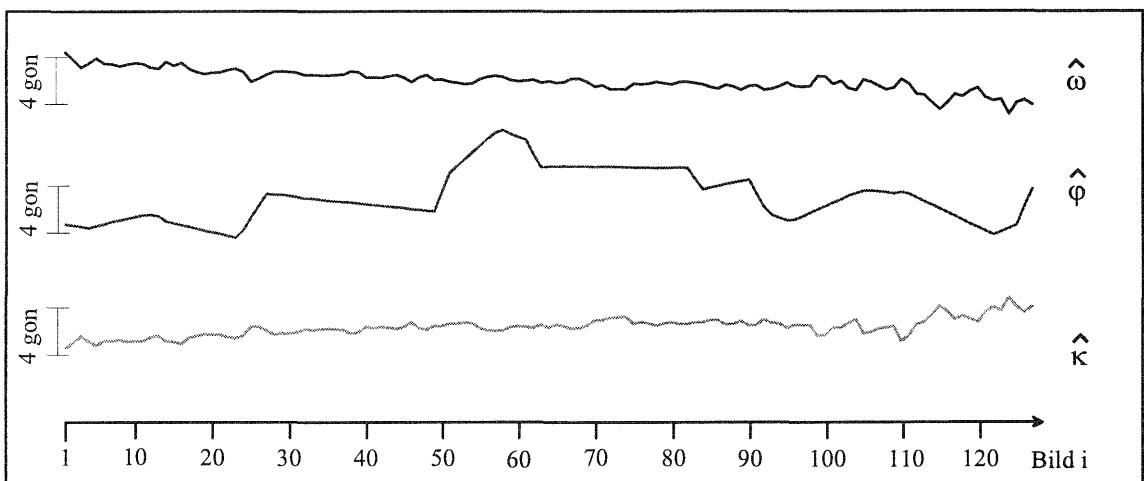


Abbildung 6.4: Kalman Version A: Geschätzte Winkel

Die Differenzen der einzelnen Versionen zur Bündellösung sind für die ermittelten Bewegungsparameter in Tabelle 6.4 zusammengefaßt. Dabei wurden die maximalen Differenzen in den Koordinaten der Projektionszentren und in den Winkeln und die rms-Differenzen (Wurzel aus der quadratischen mittleren Abweichung) berechnet. Daß die rekursive Schätzung von der Näherungswertgewinnung beeinflusst wird, zeigt der Vergleich der Version A mit den anderen Versionen. Hier zeichnet sich ein systematischer Einfluß ab, der nachfolgend noch näher untersucht werden soll. In der Tiefenrichtung wird als größte Abweichung eines Punktes eine Differenz von mehreren hundert Metern festgestellt. Insbesondere läßt dies auf eine sehr schlechte Schnittgeometrie in diesem Punkt schließen.

Vergleicht man die Ergebnisse der Versionen C und B bzw. D, so zeigt sich, daß die Beobachtung der Position der Kamera zusammen mit dem kinematischen Bewegungsmodell zu einem ähnlichen Ergebnis bezüglich der Eigenbewegung des Sensors führt, wie dies auch mit Hilfe der Paßpunkte (Versionen B und D) gefunden wird.

Die Winkeldifferenz zwischen der Bündellösung und den Versionen A bzw. D ist in den Abbildungen 6.5 und 6.6 dargestellt. Insbesondere in Version A dominiert ein systematischer Einfluß, der sich am Ende der Sequenz auf bis zu 2.5 Gon (in ω) aufsummiert hat. Für Version D ist ein sehr viel geringerer Einfluß zu beobachten.

| Äußere Orientierung: Vergleich Bündel-Kalman | | | | | | |
|--|--------------|--------------|--------------|----------|-----------|----------|
| Version | ΔU_0 | ΔV_0 | ΔW_0 | ω | φ | κ |
| maximale Differenzen [m, gon] | | | | | | |
| A | 0.28 | 0.21 | 0.26 | 2.7 | 0.88 | 2.1 |
| B | 0.17 | 0.05 | 0.03 | 0.34 | 0.07 | 0.48 |
| C | 0.03 | 0.12 | 0.07 | 0.76 | 0.51 | 0.70 |
| D | 0.17 | 0.06 | 0.03 | 0.35 | 0.10 | 0.47 |
| rms Differenzen [m, gon] | | | | | | |
| A | 0.07 | 0.11 | 0.10 | 1.3 | 0.45 | 1.0 |
| B | 0.05 | 0.02 | 0.01 | 0.11 | 0.02 | 0.19 |
| C | 0.01 | 0.04 | 0.03 | 0.31 | 0.24 | 0.24 |
| D | 0.04 | 0.03 | 0.02 | 0.11 | 0.03 | 0.18 |

Tabelle 6.4: Differenzen der Bewegungsparameter zwischen Bündellösung und Kalman-Filterung

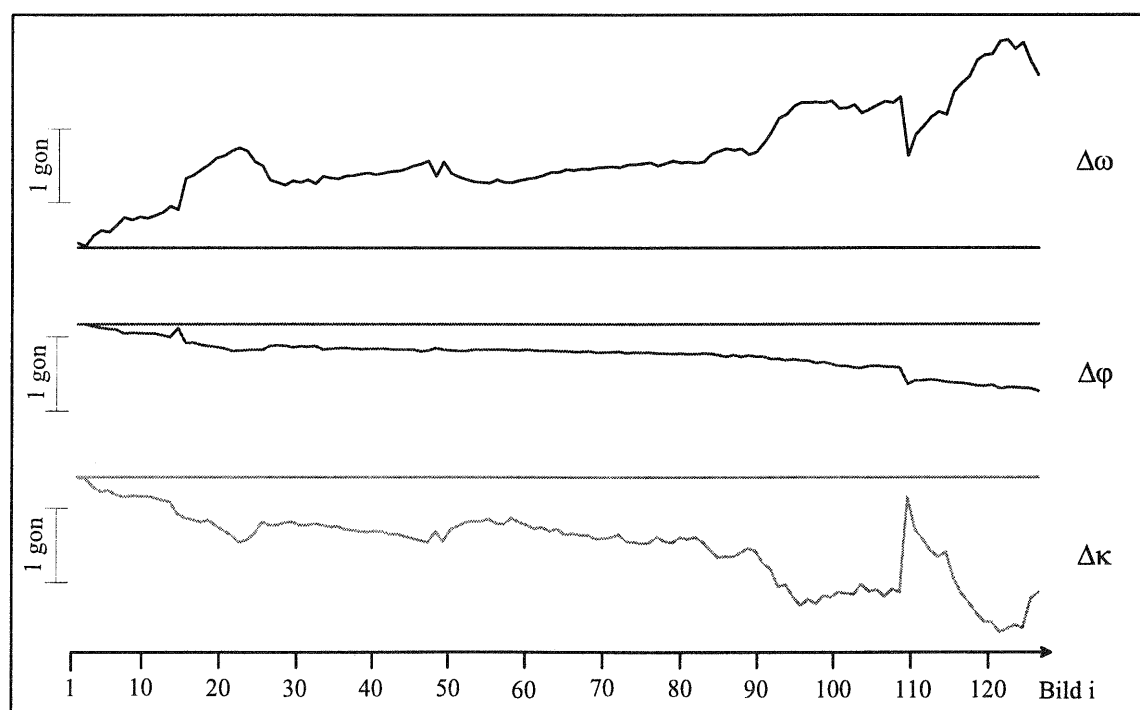


Abbildung 6.5: Winkeldifferenz zwischen Bündellösung und Kalman Version A

Die Positionsdifferenz zwischen Bündellösung und Version A ist in Abbildung 6.7 aufgetragen. Auch hier zeigt sich ein systematischer Fehler von 2-3 dm sowohl in V , d.h. der Querkomponente als auch in W , d. h. in der vertikalen Richtung. In der Bewegungsrichtung wird die Bahnkurve offensichtlich nicht deformiert. Eine Folgerung hieraus ist, daß die beobachteten Projektionszentren mit höherer Genauigkeit in die Filterung einbezogen werden müssen. Ein Versuch mit Genauigkeiten besser als 1 cm (wie sie bei der geodätischen Messung ermittelt wurden) eliminiert die Systematik nahezu komplett.

Die durch Fehlerfortpflanzung bestimmte Genauigkeit der äußeren Orientierung ist für die Bündellösung in Abbildung 6.8 dargestellt. Die Ursache für das ab Bild 80 zu beobachtende Oszillieren der theoretischen Genauigkeit liegt darin, daß ab Bild 80 nur noch jedes zweite, ab Bild 100 nur noch jedes dritte Projektionszentrum durch geodätische Messung bestimmt wurde. Für die Positionskoordinaten führt das zu einer Verschlechterung der Genauigkeit von einigen mm auf 2 bzw 3 cm. Nicht so groß ist der Einfluß auf die Richtungen, wobei

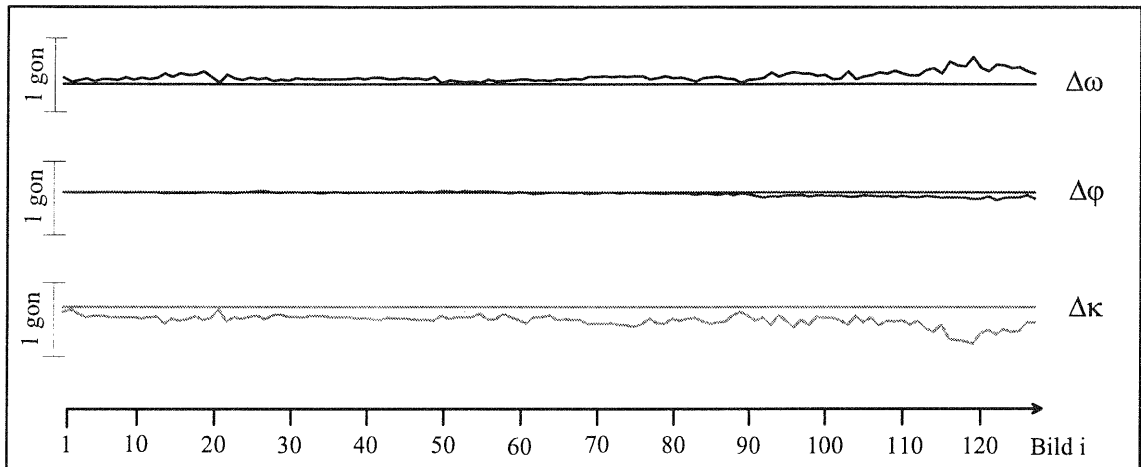


Abbildung 6.6: Winkeldifferenz zwischen Bündellösung und Kalman Version D

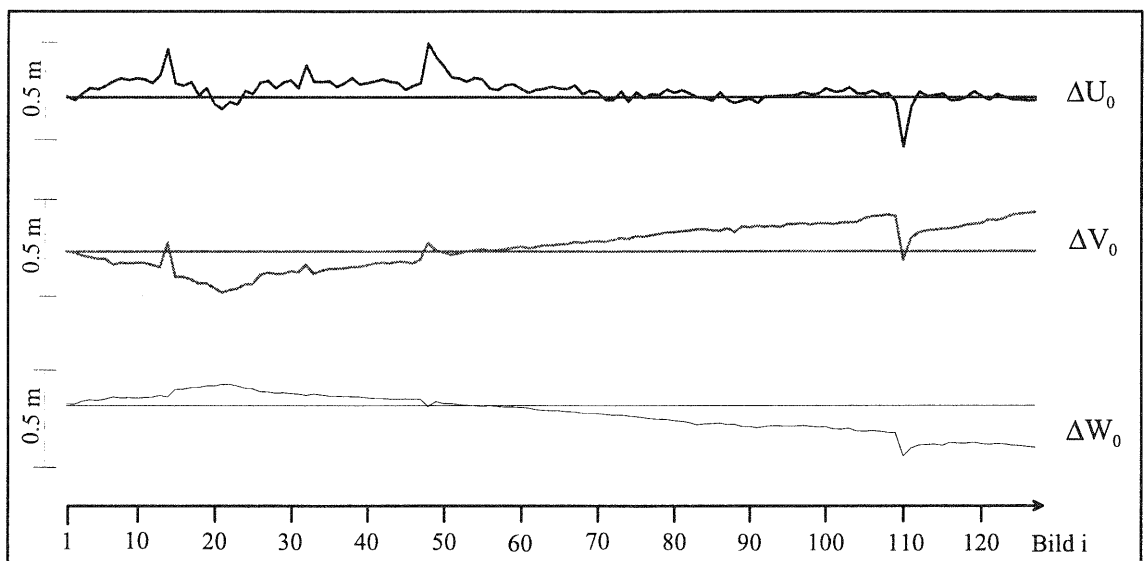


Abbildung 6.7: Positionsdiﬀerenz zwischen Bündellösung und Kalman Version A

berücksichtigt werden muß, daß diese auch nicht direkt beobachtet wurden.

In der theoretischen Genauigkeit der Kalman-Filter-Lösung (Abbildung 6.9) zeigt sich keine entsprechende Oszillation, was in der Filterwirkung des kinematischen Bewegungsmodells begründet liegt. Hier wird allerdings ein anderer Effekt deutlich. Die Genauigkeit sowohl der Rotations- als auch der Translationsparameter wird über einige Bilder (z. B. von Bild 70 bis 80) stetig genauer und verschlechtert sich dann wieder abrupt. Die Ursache hierfür ist, daß in den entsprechenden Bildern Punkte aus dem Blickfeld der CCD-Kamera verschwinden. Diese Punkte konnten lange in Sequenz verfolgt werden und haben sich dabei in kurzer Distanz (z.B. 20 m) von der Kamera befunden haben. Die effektive Basis wird für solche Punkte über viele Bilder hinweg vergleichsweise groß, so daß sie für die Schätzung der Eigenbewegung entsprechend wichtig sind.

Zusammenfassend kann festgestellt werden, daß durch die Kalman-Filterung die Schätzung der Eigenbewegung mit einer Genauigkeit von 0.1 - 1 Gon für die Rotationsparameter und im cm bis dm Genauigkeitsbereich für die Positionparameter der äußeren Orientierung gelingt. Dieses Genauigkeitsniveau ist für viele Anwendungen innerhalb der Navigation völlig ausreichend.

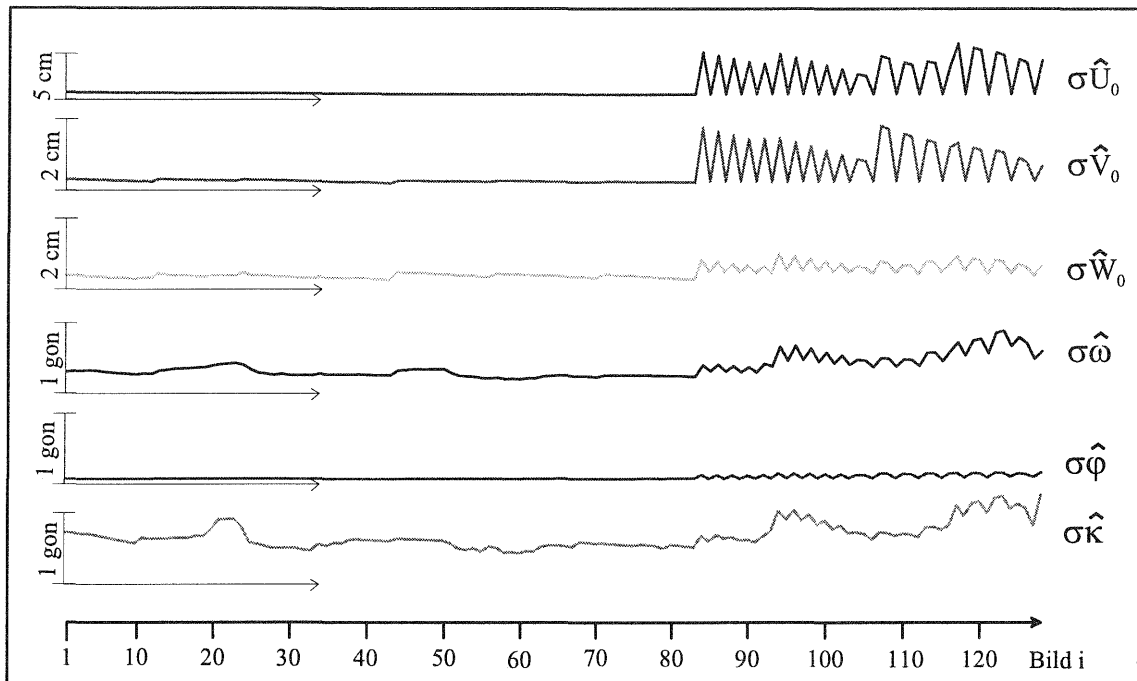


Abbildung 6.8: Theoretische Genauigkeit der äußeren Orientierung (Bündellösung)

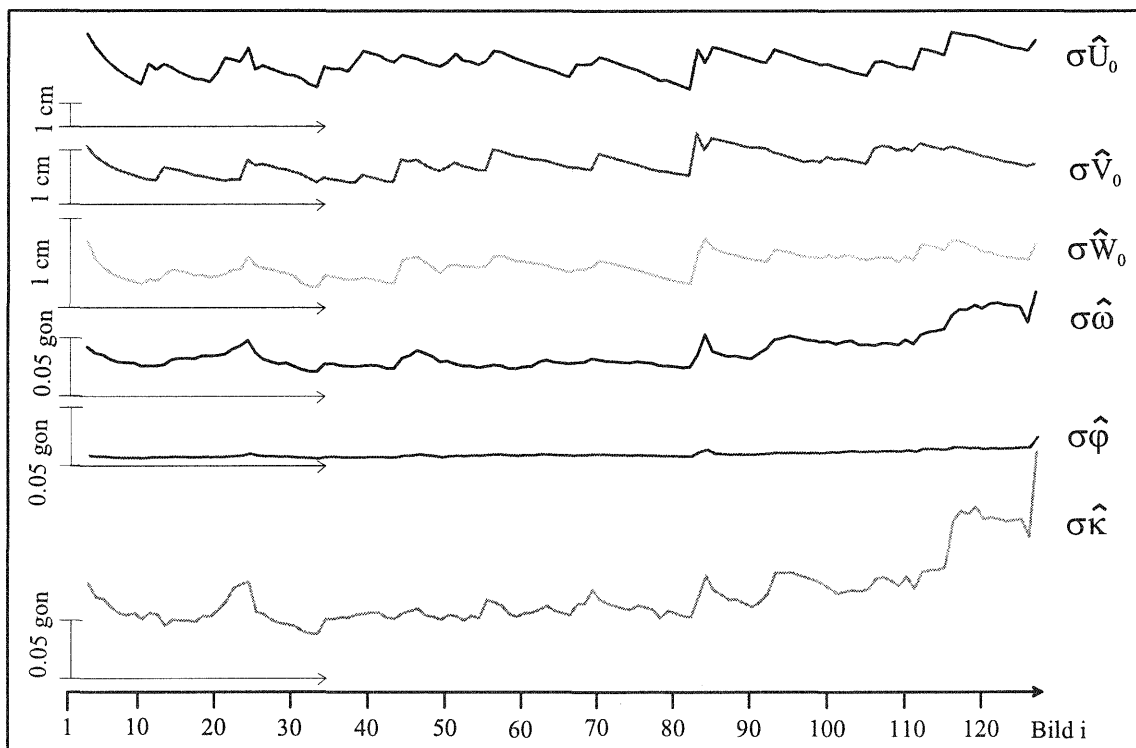


Abbildung 6.9: Theoretische Genauigkeit der äußeren Orientierung (Kalman Version A)

6.3.5 Schätzung von Strukturparametern

Die Differenzen zwischen den Objektpunktkoordinaten der Bündellösung und der Kalman-Filterung (Versionen A bis D) sind in Tabelle 6.5 aufgelistet. In Version A zeigt sich, daß das Verfahren der Näherungswertbe-

| Objektkoordinaten: Vergleich Bündel-Kalman | | | | | | |
|--|---------------------|--------------|--------------|----------------------|--------------|--------------|
| Version | ΔU_i | ΔV_i | ΔW_i | ΔU_i | ΔV_i | ΔW_i |
| | rms Differenzen [m] | | | max. Differenzen [m] | | |
| A | 94. | 21. | 4.7 | 885. | 203. | 55.6 |
| B | 6.4 | 1.6 | 0.3 | 75. | 28. | 3.1 |
| C | 7.4 | 1.7 | 0.4 | 74. | 17. | 6.3 |
| D | 7.9 | 2.1 | 0.4 | 149. | 55. | 6.2 |

Tabelle 6.5: Differenzen in den Objektkoordinaten zwischen Bündellösung und Kalman-Filterung

stimmung für die Objektpunktkoordinaten einen signifikanten Einfluß auf die Schätzung hat und dazu führt, daß die mittlere Abweichung von der simultanen Bündellösung relativ groß ist. Daß sich in der Schätzung der Tiefenkomponente (U) die größten Abweichungen zeigen, ist insbesondere auf die schlechte Schnittgeometrie zurückzuführen.

Zur Überprüfung der geschätzten Objektkoordinaten verfolgter Punkte werden die geodätisch koordinierten Kontrollpunkte herangezogen. Die Identifikation eines markanten Punktes mit einem Kontrollpunkt wird interaktiv vorgenommen. Aufgrund der speziellen Signalisierung der Kontrollpunkte (Scharz-Weiß-Muster) sind viele als markante Punkte im Bild ebenfalls detektiert und lokalisiert worden. Die Differenzen zu den geschätzten Werten der Kalman-Filterung sind in Tabelle 6.6 aufgelistet. Zum Vergleich sind die entsprechenden Differenzen zwischen den Kontrollpunktkoordinaten und den in der Bündellösung bestimmten Objektpunktkoordinaten in der Tabelle zusätzlich angegeben. Ganz offensichtlich sind die in der Bündellösung geschätzten Positionen der Verknüpfungspunkte sehr viel genauer als die in Kalman Version A bestimmten. Ist die Distanz zwischen Kamera und Objektpunkt gering, so liegt die Abweichung bei 1/10 der Entfernung. Für weit entfernte Punkte kann die Abweichung bis auf 50% der Entfernung ansteigen.

An ausgewählten Punkten sind die Koordinatendifferenzen zwischen Bündellösung und Version A bzw. Version C der Kalman-Filterung in der Abbildung 6.10 dargestellt. Die Punkte 1156, 1197 und 1220 sind bei der merkmalsbasierten Punktverfolgung ausführlich betrachtet worden. Die Verfolgung von Bild zu Bild ist in den Abbildungen 4.8, 4.9 und 4.10 (Abschnitt 4.3.2) bildhaft wiedergegeben. Bei Version A ist deutlich zu erkennen, daß die Punkte von einem Tiefenstartwert im Hintergrund der Szene (300 m entfernt) sich einer genaueren Koordinierung annähern. Für Punkt 1156 ist nach ca 20 Bildern, für Punkt 1197 nach ca 15 Bildern dieser Einschwingvorgang beendet. Werden auf diese Art und Weise die Näherungswerte gewonnen bzw. verbessert, so zeigen die Beispiele, daß die Punkte hierfür über viele Bilder in der Bildfolge verfolgt werden müssen. Ganz offensichtlich ist aber dieser Weg nicht unproblematisch.

Die durch Fehlerfortpflanzung ermittelte theoretische Genauigkeit dieser Punkte ist in Abbildung 6.11 dargestellt. Deutlich zeigt sich auch hier die Abhängigkeit von der Anzahl der verknüpften Bilder. Die Unterschiede zwischen Version A und C sind hier erwartungsgemäß sehr gering.

| Koordinatendifferenzen zwischen Kontroll- und Verknüpfungspunkten | | | | | | | | | | |
|---|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|----------------|-------------------|------------------|
| Kontroll- pkt Nr. | Verkn.- pkt Nr. | Kalman Version A | | | Bündellösung | | | Distanz [m] | letzte Bi.-Nr. | verkn. Bilder |
| | | ΔU_i [m] | ΔV_i [m] | ΔW_i [m] | ΔU_i [m] | ΔV_i [m] | ΔW_i [m] | | | |
| 5020 | 1356 | 7.54 | 2.43 | 0.29 | -0.66 | -0.21 | -0.01 | 22.44 | 8 | 7 |
| 5021 | 1363 | 1.03 | 0.07 | 0.04 | 0.03 | -0.00 | 0.00 | 13.40 | 10 | 10 |
| 5037 | 5636 | 40.45 | 9.11 | 2.03 | 1.33 | 0.29 | 0.07 | 68.09 | 60 | 4 |
| 5038 | 1923 | 21.12 | 2.37 | 0.42 | 6.09 | 0.67 | 0.14 | 82.03 | 15 | 6 |
| 5041 | 2973 | 178.25 | 27.60 | -1.12 | 22.97 | 3.35 | -0.06 | 277.60 | 25 | 5 |
| 5041 | 3311 | 10.78 | 1.44 | -0.05 | 2.52 | 0.30 | 0.01 | 95.16 | 38 | 13 |
| 5041 | 24908 | 25.58 | 2.96 | -0.20 | 2.82 | 0.31 | -0.01 | 93.16 | 55 | 6 |
| 5048 | 6324 | 11.97 | 4.21 | -0.02 | 5.38 | 1.74 | 0.02 | 138.32 | 86 | 12 |
| 5049 | 5670 | 54.19 | 23.67 | -0.35 | 19.03 | 8.07 | -0.08 | 207.82 | 63 | 6 |
| 5060 | 2669 | 17.91 | 5.47 | 0.21 | -2.26 | -0.70 | -0.03 | 50.55 | 22 | 6 |
| 5061 | 4018 | 0.67 | 0.06 | 0.04 | 0.08 | 0.00 | 0.01 | 10.37 | 44 | 11 |
| 5062 | 5128 | 7.40 | 1.00 | 0.23 | 2.32 | 0.31 | 0.08 | 45.34 | 57 | 9 |
| 5064 | 3689 | 3.71 | 0.41 | 0.08 | 0.87 | 0.09 | 0.01 | 50.67 | 42 | 14 |
| 5065 | 5535 | 57.18 | 13.88 | 0.56 | -3.81 | -0.95 | -0.03 | 97.23 | 58 | 4 |
| 5068 | 3343 | 1.18 | 0.32 | 0.08 | -2.04 | -0.71 | 0.02 | 58.65 | 46 | 21 |
| 5069 | 5989 | 75.77 | 14.18 | 1.35 | 0.13 | 0.02 | -0.00 | 141.13 | 69 | 4 |
| 5069 | 5749 | 31.86 | 6.14 | 0.53 | -2.77 | -0.54 | -0.04 | 101.42 | 64 | 5 |
| 5069 | 6293 | 54.67 | 9.37 | 1.12 | 1.59 | 0.26 | 0.03 | 111.56 | 77 | 4 |
| 5069 | 6566 | 14.21 | 2.13 | 0.35 | -0.34 | -0.04 | -0.01 | 61.53 | 86 | 6 |
| 5069 | 6900 | 1.85 | 0.20 | 0.06 | -0.16 | -0.03 | -0.00 | 35.21 | 100 | 14 |
| 5070 | 6178 | 25.63 | 7.42 | 0.73 | 5.84 | 1.62 | 0.21 | 73.34 | 86 | 15 |
| 5071 | 6053 | 99.11 | 21.95 | 1.67 | 1.88 | 0.34 | 0.11 | 175.34 | 71 | 4 |
| 5072 | 8323 | 7.75 | 1.88 | 0.37 | -0.20 | -0.06 | 0.01 | 32.81 | 129 | 10 |
| 5073 | 6861 | 18.08 | 6.47 | 0.37 | -1.26 | -0.52 | 0.00 | 80.92 | 94 | 8 |

Tabelle 6.6: Kontrolle: Differenzen zwischen den Objektkoordinaten von Kontrollpunkten und Verknüpfungspunkten (Kalman-Filter Version A)

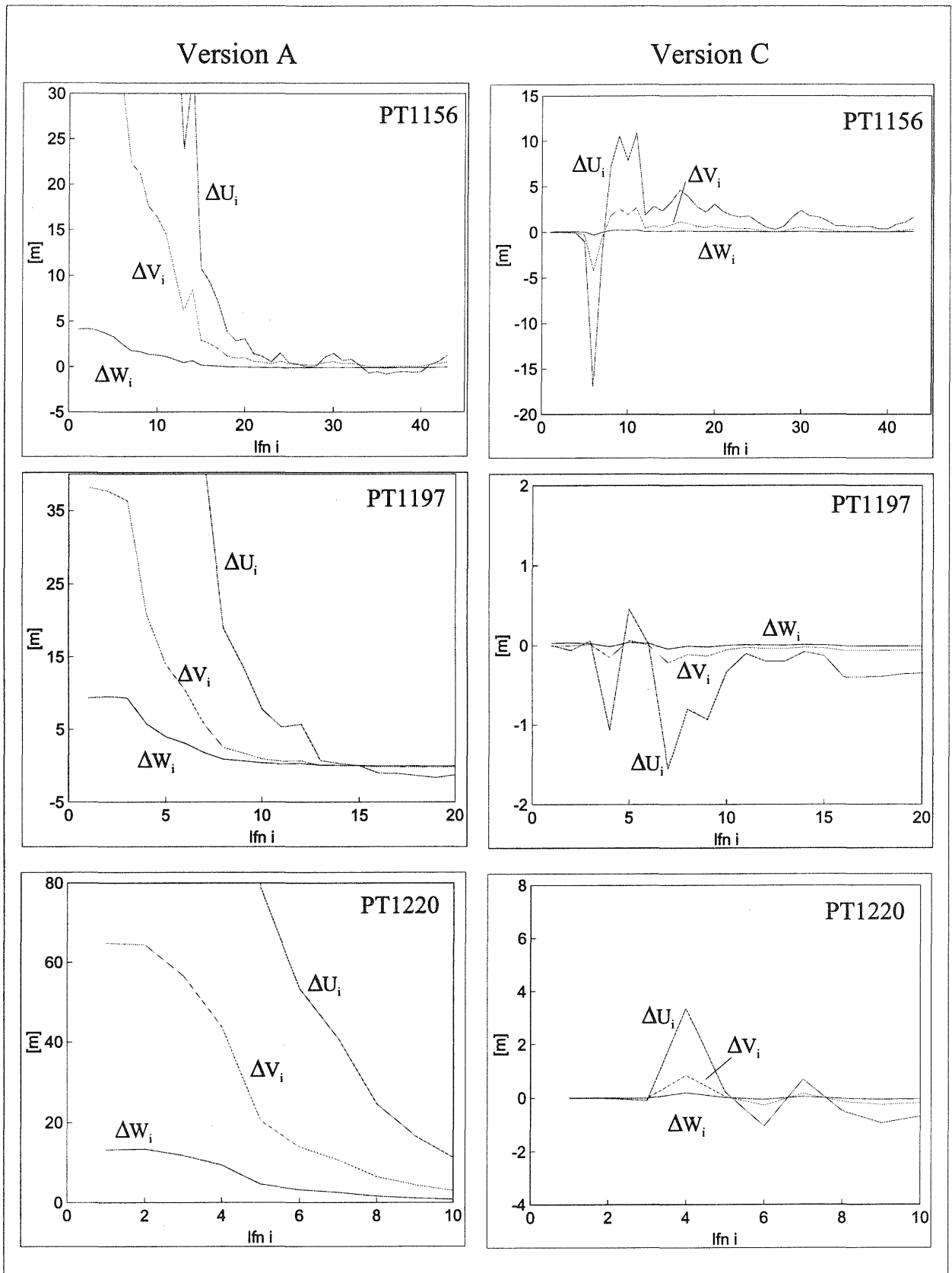


Abbildung 6.10: Koordinatendifferenzen in ausgewählten Objektpunkten zwischen der Bündellösung und der Kalman-Filterung (Versionen A und C).

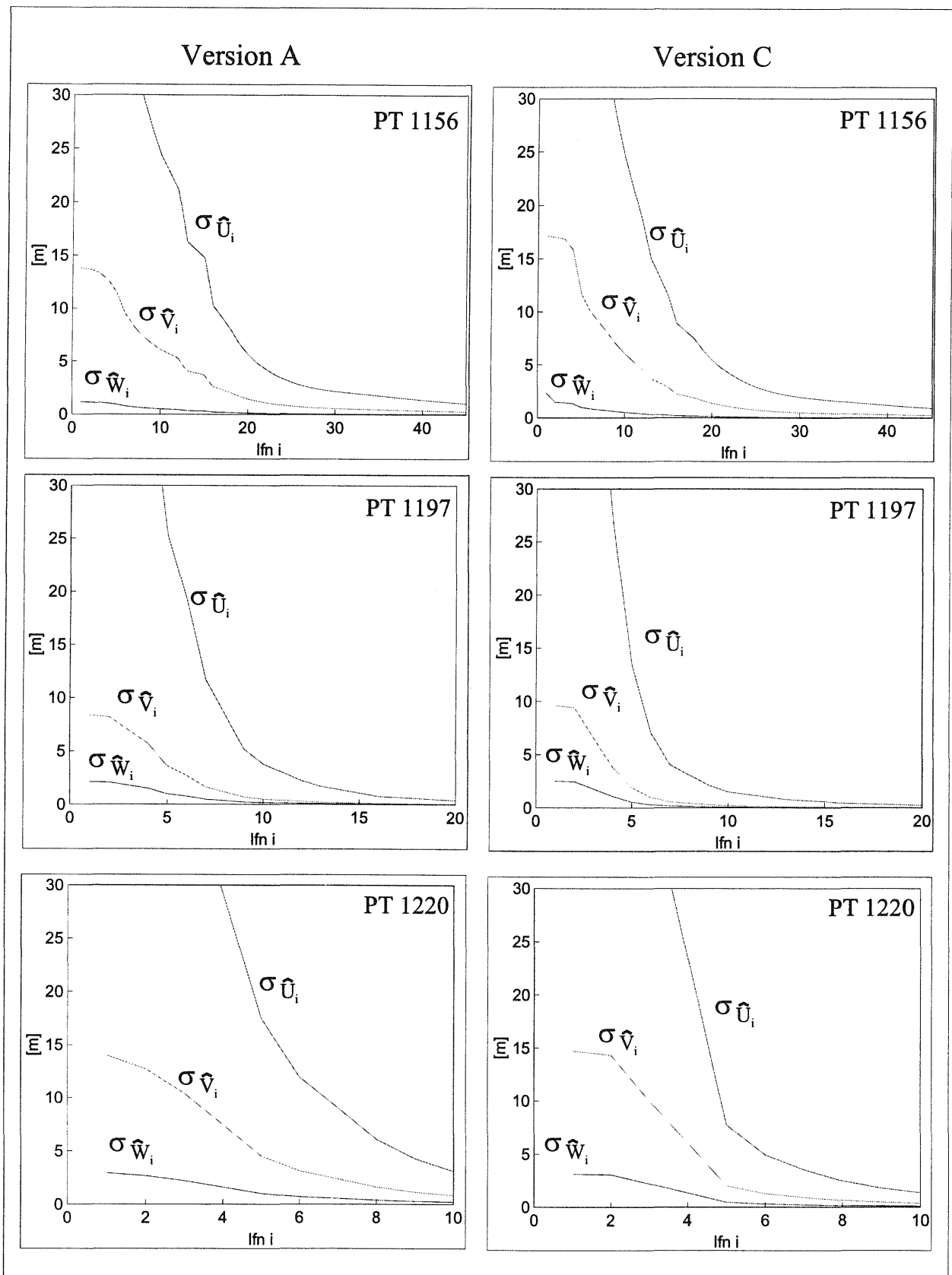


Abbildung 6.11: Theoretische Genauigkeit der geschätzten Objektpunktkoordinaten von Version A und C der Kalman-Filterung.

6.4 Zusammenfassende Beurteilung

Mit dem entwickelten 3D-Filter wird ein neuer Lösungsweg zur Rekonstruktion von Bewegung und Struktur aus langen Bildsequenzen vorgestellt. Zurückgegriffen wird dabei auf die Kalman-Filter-Technik, die vielfach auch für andere Anwendungen in der Prozessierung von Bildsequenzen vorgeschlagen wurde. Der 3D-Kalman-Filter unterscheidet sich von bekannten Verfahren der sequentiellen Bildtriangulation dadurch, daß der Systemzustand auf die 12 Orientierungsparameter eines Bildpaares und die Koordinaten der im Bild momentan verfolgten Objektpunkte beschränkt wird. Ein weiteres Unterscheidungsmerkmal von der sequentiellen Bildtriangulation ist, daß im Systemmodell des Kalman-Filters ein kinematisches Bewegungsmodell berücksichtigt wird.

Daß ein vergleichbarer Weg einer 3D-Rekonstruktion bislang aus der Literatur nicht bekannt ist, wurde bereits bei der Diskussion der Querbezüge zu anderen Arbeiten mit langen Bildsequenzen in Abschnitt 2.4 herausgestellt. Zu den wichtigsten Unterschieden zählen die Rekonstruktion in ortsfesten Objektkoordinaten, die Vorteile besitzt gegenüber der Rekonstruktion im bewegten Sensorsystem, und die simultane Berücksichtigung von Bewegungs- und Strukturparametern im Systemzustand, wodurch die für viele andere Arbeiten typische Loslösung der Bewegungs- von der Strukturrekonstruktion unter Vernachlässigung der Korrelation dieser Parameter vermieden wird.

Methodisch ist mit dem entworfenen 3D-Filter ein Weg aufgezeigt, der den Navigationsansprüchen nach einer rekursiven Aufdatierung der Bewegungs- und Strukturparameter gerecht wird. Bei der algorithmischen Realisierung des Kalman-Filters wurde die Standardform gewählt, die auch als Kovarianzfilter bezeichnet wird, da die Kovarianzmatrix des Zustandsvektors über die Zeit fortgepflanzt wird. Die Beachtung der Symmetrieeigenschaften, insbesondere bezüglich der Kovarianzmatrizen, hat sich bei der experimentellen Erprobung als hinreichend erwiesen, um eine stabile Lösung durch den Filter über die ganze Sequenz herbeizuführen. Optimierungen des Speicherplatzes oder der benötigten Rechenzeiten durch Faktorisierungs- (z.B. LU) und Orthogonalisierungsschemen (z.B. Givens Transformation) sind in Salzmann (1988) und in Grün und Kersten (1992) diskutiert.

Ein wichtiger Unterschied zu anderen Arbeiten in diesem Bereich liegt in der Durchführung eines umfangreichen, kontrollierten Tests. Aus experimenteller Sicht konnte der Nachweis erbracht werden, daß die Schätzung der Eigenbewegung bezogen auf ein ortsfestes Koordinatensystem mit einem Genauigkeitsniveau von 0.1 - 1 Gon für die Rotation und im dm-Bereich für die Position der Kamera gelingt. Bei den merkmalsbasiert verfolgten Verknüpfungspunkten wurde eine beträchtliche Anzahl an Fehlern festgestellt, die bei der Kalman-Filterung eliminiert werden konnten. Die unabhängige geodätische Kontrolle koordinierter Verknüpfungspunkte zeigt, daß die Bestimmung von Näherungswerten für die Positionsbestimmung der Objektpunkte ein kritischer Schritt bei der Analyse von Monobildsequenzen ist. Die Verfahrensversion, in der die Punkte als weit entfernte Objektpunkte mit schwachem Datum in den Filter eingeführt wurden, hat sich nur mit Einschränkungen als tauglich erwiesen. Besser ist es, Näherungswerte durch räumlichen Vorwärtseinschnitt im Sinne einer rekursiven Bündellösung zu gewinnen.

7 Zusammenfassung und Ausblick

Die Arbeit will einen Beitrag zu den theoretischen Grundlagen der Verarbeitung von Bildsequenzen leisten. Ausgerichtet sind die Analyseprozesse auf Anwendungen in der passiven Navigation, was impliziert, daß die zu ermittelnden Navigationsdaten fortlaufend rekonstruiert bzw. aufdatiert werden. Das Ziel der Arbeit ist es einen Verfahrensablauf auszuarbeiten, der sich für die Analyse langer Bildsequenzen eignet. Dabei sind zwei in der Literatur intensiv erforschte Bereiche angesprochen. Der erste Bereich betrifft den Meßprozeß, dessen Aufgabe es ist, durch Bildzuordnung die in der Bildfolge erkennbare Bewegung zu erfassen. Der zweite Bereich behandelt die Rekonstruktion von Bewegungs- und Strukturparametern, die eine Interpretation der projizierten Bewegung darstellt. Messen und Rekonstruieren fallen in dem zu entwickelnden System mit jedem neu erfaßten Bild einer Sequenz an.

Grundvoraussetzung für die Formulierung von Meßprozessen und Rekonstruktionsverfahren ist die Bewegungsmodellierung. Für die Beschreibung der projizierten Bewegung in der Bildebene einer bewegten Kamera ist das Modell-Wissen über den Objektraum von großer Bedeutung. Unter der Annahme verschiedener Objektformen wird die projizierte Bewegung abgeleitet. Sie dient als Grundlage für die Formulierung der Bildzuordnungsverfahren zur Schätzung des optischen Flusses.

Die Rekonstruktion von Bewegungs- und Strukturparametern soll in einem szenefesten Koordinatensystem erfolgen. Mit dieser für die Rekonstruktion aus langen Bildsequenzen bislang unüblichen Vorgehensweise wird ein Rekonstruktionsansatz präsentiert, der für ein rekursiv prozessierendes System sehr geeignet ist. Die Rotationsbewegung wird zur Parameterschätzung im mitbewegten System parametrisiert und in eine Quaternionen-Repräsentation überführt, die der konventionellen Winkelparametrisierung in theoretischer und algorithmischer Sicht überlegen ist. Die Beziehungen zur Winkelparametrisierung werden benötigt, um die Ergebnisse in dieser leichter interpretierbaren Form darzustellen.

Dem Meßprozeß eines navigierenden Systems kommt eine zentrale Bedeutung zu. In der Arbeit werden die verschiedenen Ansätze zur Schätzung des optischen Flusses im Hinblick auf die Zuordnung in langen Bildsequenzen diskutiert und verglichen. Diese Aufarbeitung zeigt viele Querbezüge zwischen den Verfahren auf, macht aber auch deutlich, daß kein einzelnes Verfahren als das "beste" für die Verarbeitung in Bildsequenzen identifiziert werden kann.

Bezogen auf Bildausschnitte, die als markante Punkte im Bild identifiziert werden, ist ein merkmalsbasiertes und ein intensitätsbasiertes Verfahren zur Verfolgung von Punkten in der Sequenz vorgeschlagen worden. Durch die Integration des Scale-Space in einen intensitätsbasierten Ansatz zur Bildzuordnung wird ein völlig neuer Weg zum Entwurf eines Bildzuordnungsverfahrens aufgezeigt. Hierzu wird die grob-zu-fein gerichtete Kontrollstrategie in den Schätzprozeß einbezogen.

Im Entwurf eines 3D-Kalman-Filters kommt der Systemgedanke deutlich zum Ausdruck. Der Systemzustand berücksichtigt die Bewegungsparameter einer Zweibildeinheit sowie die 3D-Struktur der Szene, die durch die 3D-Koordinaten der verfolgten Punkte parametrisiert ist. Der wesentliche Aspekt für die rekursive Rekonstruktion liegt darin, daß die zu den zurückliegenden Zeitpunkten gewonnenen Informationen über die Struktur zum aktuellen Zeitpunkt zur Verfügung stehen. Zusammen mit dem kinematischen Bewegungsmodell und den Beobachtungen, d. h. den Bildkoordinaten der verfolgten Punkte, werden durch den Kalman-Filter die verschiedenen Informationen in eine Optimalschätzung einbezogen.

In den Experimenten wird eine Bildfolge analysiert, die auf einem für die Erprobung navigierender Fahrzeuge angelegten Versuchsgelände aufgezeichnet wurde. Aus den Experimenten sollen zwei Ergebnisse hervorgehoben werden. Das neu entwickelte intensitätsbasierte Zuordnungsverfahren hat sich für die Verfolgung von markanten Punkten sehr bewährt. Es konnte nachgewiesen werden, daß das Verfahren einen sehr großen Konvergenzradius besitzt und dabei sehr schnell konvergiert.

Mit den Experimenten zur 3D-Kalman-Filterung wurde der Nachweis erbracht, daß ein solches System zur rekursiven Rekonstruktion von Bewegungs- und Strukturparametern auch bei Realweltdaten erfolgreich ist. Entsprechende erfolgreiche Experimente sind bislang aus der Literatur nicht bekannt.

Erweiterungen des entwickelten Systems können in verschiedene Richtungen erfolgen:

1. Zusätzlich zu Punkten können Linien zur Repräsentation der Struktur der Szene verwendet werden, was zu einer vollständigeren Szenebeschreibung führt.
2. Anstelle einer Kamera könnte eine Stereokamera im Kalman-Filter berücksichtigt werden, womit eine genauere Positionsbestimmung von Objektpunkten möglich wird.
3. Das System läßt sich prinzipiell auch in Bereichen der Robotik anwenden.
4. Werden die Ansätze zur merkmalsbasierten und intensitätsbasierten Zuordnung in einen Ansatz integriert, profitiert die Punktverfolgung von den Vorzügen beider Verfahren. Ein entsprechend integrierter Ansatz ist in Hahn (1993) vorgestellt worden.

Literatur

- Ackermann, F. [1984]: High Precision Image Correlation. In: *Vorträge der 39. Photogrammetrischen Woche*, Sept. 1983, Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart, Heft 9, pp. 231-243, Stuttgart, 1984.
- Ackermann, F., Ebner, H., Klein, H. [1970]: Ein Programmpaket für die Aerotriangulation mit unabhängigen Modellen. *Bildmessung und Luftbildwesen*, 38. Jahrgang, Heft 4, pp. 218-224, 1970.
- Ackermann, F., Hahn, M. [1991]: Image Pyramids for Digital Photogrammetry. In: *Digital Photogrammetric Systems*, Ebner, H. et al. (eds.), pp. 43-58, Wichmann-Verlag, Karlsruhe, 1991.
- Ackermann, F., Krzystek, P. [1991]: MATCH-T: Automatic Mensuration of Digital Elevation Models. In: *3rd Technical Seminar of the Sociedad Espanola de Cartografia, Fotogrametria y Teledeccion*, Barcelona, April 1991.
- Ackermann, F., Zheng, Y.-J. [1990]: Inverse and Ill-posed Problems in Photogrammetric Surface Reconstruction. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 28, part 3/2, pp. 22-47, Wuhan, May 1990.
- Adelson, E.H., Bergen, J.R. [1985]: Spatiotemporal Energy Models for the Perception of Motion. *Journal of Optical Society of America A*, vol. 2, no. 2, pp. 11-16, Februar 1985.
- Adiv, G. [1985]: Determining 3-D Motion and Structure from Optical Flow Generated by Several Moving Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 4, 1985.
- Aggarwal, J.K. [1986]: Motion and Time-varying Imagery - an Overview. In: *Proc. of the Workshop on Motion: Representation and Analysis*, Charleston, pp. 1-6, 1986.
- Aloimonos, J. [1989]: Unification and Integration of Visual Modules: An Extension to the Marr Paradigm. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 507-551, San Mateo, CA, 1989.
- Anandan, P. [1987a]: A Unified Perspective of Computational Techniques for the Measurement of Visual Motion. In: *Proceedings of the 1st International Conference on Computer Vision*, pp. 219-230, London, 1987.
- Anandan, P. [1987b]: Measuring Visual Motion From Image Sequences. PhD thesis, University of Massachusetts, May 1987.
- Asada, H., Brady, M. [1986]: The Curvature Primal Sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 2-14, 1986.
- Baarda, W. [1968]: A Testing Procedure for Use in Geodetic Networks. Netherl. Geodetic Commission, Publ. on Geodesy, vol. 2, no. 5, Delft, 1968.
- Bässmann, H., Besslich, Ph.W. [1991]: Bildverarbeitung Ad Oculum. Springer Verlag, Berlin, 1991.
- Baker, H.H. [1987]: Multiple-Image Computer Vision. In: *Vorträge der 41. Photogrammetrischen Woche*, Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart, Heft 12, pp. 7-19, Stuttgart, 1987. Institut für Photogrammetrie der Universität Stuttgart.
- Ballard, D.H., Brown, Ch.M. [1982]: Computer Vision. Prentice Hall, New Jersey, 1982.
- Baltsavias, E.P. [1991]: Multiphoto Geometrically Constraint Matching. Dissertation Mitteilungen Nr. 49, Institut für Geodäsie und Photogrammetrie der ETH Zürich, Zürich, December 1991.
- Barron, J.L., Jepson, A.D., Tsotsos, J.K. [1990]: The Feasibility of Motion and Structure from Noisy Time-Varying Image Velocity Information. *International Journal of Computer Vision*, vol. 5, no. 3, pp. 239-269, 1990.
- Barrow, H.G., Tenenbaum, J.M. [1978]: Recovering Intrinsic Scene Characteristics from Images. In: *Computer Vision Systems*, Hanson, A.R., Riseman, E.M. (eds.), Academic Press, New York, 1978.
- Bartsch, T. [1987]: Untersuchungen zur Detektion und Lokalisation markanter Punkte in Farbbildern. FBI-HH-B-131/87, Diplomarbeit, Universität Hamburg, Fachbereich Informatik, Hamburg, Oktober 1987.
- Battiti, R., Amaldi, E., Koch, Ch. [1991]: Computing Optical Flow Across Multiple Scales: An Adaptive Coarse-to-Fine Strategy. *International Journal of Computer Vision*, vol. 6, no. 2, pp. 133-145, 1991.
- Beaudet, P.R. [1978]: Rotationally Invariant Image Operators. In: *Proc. International Joint Conference on Pattern Recognition*, pp. 579-583, 1978.
- Beggs, J.S. [1983]: Kinematics. Hemisphere Publishing Corporation, Washington, New York, London, 1983.
- Bergholm, F. [1987]: Edge Focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 726-741, Nov. 1987.
- Bergholm, F. [1989]: On the Content of Information in Edges and Optical Flow. Dissertation TRITA-NA-P8904, Royal Institute of Technology, Department of Numerical Analysis and Computing Science, Stockholm, Sweden, May 1989.
- Berzins, V. [1984]: Accuracy of Laplacian Edge Detectors. *Computer Vision, Graphics, and Image Processing*, vol. 27, pp. 195-210, 1984.
- Bolles, R.C., Baker, H.H., Marimont, D.H. [1987]: Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion. *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7-55, June 1987.

- Brandt, A.v., Zaig, D. [1991]: Ermittlung von Eigenbewegung und Tiefeninformationen aus monokularen Bildfolgen. In: *Mustererkennung 1991, Proceedings des 13. Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung*, Radig, B. (ed.), Informatik Fachberichte, vol. 290, pp. 560-567, Springer Verlag, Berlin, Okt. 1991.
- Broida, T.J., Chellappa, R. [1986]: Estimation of Object Motion Parameters from Noisy Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 90-99, Jan. 1986.
- Broida, T.J., Chellappa, R. [1991]: Estimating the Kinematics and Structure of a Rigid Object from a Sequence of Monocular Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 497-513, June 1991.
- Bronstein, I., Semendjajew, K. [1977]: Taschenbuch der Mathematik. 17. Auflage 1977, Verlag Harri Deutsch, Zürich, 1977.
- Brooks, R.A. [1981]: Symbolic Reasoning among 3-D Models and 2-D Images. *Artificial Intelligence*, vol. 17, pp. 285-348, 1981.
- Brown, C., Durrant-Whyte, H. [1989]: Centralized and Decentralized Kalman Filter Techniques for Tracking, Navigation, and Control. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 651-675, San Mateo, CA, 1989.
- Bruss, A.R., Horn, B.K.P. [1983]: Passive Navigation. *Computer Vision, Graphics, and Image Processing*, vol. 21, no. 1, pp. 3-20, January 1983.
- Burger, W. [1992]: Inexact Visual Navigation. In: *High Precision Navigation 91*, Linkwitz, K. und Hangleiter, U. (eds.), pp. 229-240, Dümmler Verlag, Bonn, 1992.
- Burger, W., Bhanu, B. [1990]: Estimating 3-D Egomotion from Perspective Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 11, pp. 1040-1058, Nov. 1990.
- Buxton, B.F., Buxton, H. [1984]: Computation of Optic Flow from the Motion of Edge Features in Image Sequences. *Image and Vision Computing*, no. 2, pp. 59-75, 1984.
- Cafforio, C., Rocca, F. [1976]: Methods for Measuring Small Displacements of Television Images. *IEEE Transactions on Information Theory*, no. IT-22, pp. 573-579, Sept. 1976.
- Cafforio, C., Rocca, F. [1979]: Tracking Moving Objects in Television Images. *Signal Processing*, no. 1, pp. 133-140, 1979.
- Canny, J.F. [1986a]: A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.
- Canny, J.F. [1986b]: Collision Detection for Moving Polyhedra. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 200-209, March 1986.
- Castleman, K.R. [1979]: Digital Image Processing. Prentice Hall, Englewood Cliffs, New Jersey, 1979.
- Cornelius, N., Kanade T. [1983]: Adapting Optical Flow to Measure Object Motion in Reflectance and X-ray Image Sequences. In: *ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception*, pp. 145-153, Toronto, Canada, April 1983.
- Cornsweet, T.N. [1970]: Visual Perception. Academic Press, New York, 1970.
- Crowley, J.L., Stelmazsk, P. [1990]: Measurement and Integration of 3-D Structures by Tracking Edge Lines. In: *Computer Vision - ECCV 90, Proceedings of the 1st European Conference on Computer Vision*, Faugeras, O. (ed.), Lecture Notes in Computer Science, vol. 427, pp. 269-280, Springer Verlag, Berlin, 1990.
- Cui, N., Weng, J., Cohen, P. [1990]: Extended Structure and Motion Analysis from Monocular Image Sequences. In: *Proceedings of the 3rd International Conference on Computer Vision*, pp. 222-229, Osaka, 1990.
- Daniilides, K., Nagel, H.-H. [1990]: Analytical Results on Error Sensitivity of Motion Estimation from Two Views. In: *Computer Vision - ECCV 90, Proceedings of the 1st European Conference on Computer Vision*, Faugeras, O. (ed.), Lecture Notes in Computer Science, vol. 427, pp. 199-208, Springer Verlag, Berlin, 1990.
- De Micheli, E., Caprile, B., Ottonello, P., Torre, V. [1989]: Localization and Noise in Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1106-1117, 1989.
- Deriche, R., Giraudon, G. [1990]: Accurate Corner Detection: An Analytical Study. In: *Proceedings of the 3rd International Conference on Computer Vision*, pp. 66-70, Osaka, 1990.
- Dickmanns, E.D., Graefe, V. [1988]: Dynamic Monocular Machine Vision. *Machine Vision Applications*, vol. 1, pp. 223-240, 241-261, 1988.
- Dickmanns, E.D., Mysliwetz, B.D. [1992]: Recursive 3-D Road and Realtime Ego-State Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 199-213, February 1992.
- Dreschler, L. [1981]: Ermittlung markanter Punkte auf den Bildern bewegter Objekte und Berechnung einer 3D-Beschreibung auf dieser Grundlage. Dissertation, Fachbereich Informatik, Universität Hamburg, 1981.
- Dreschler, L., Nagel, H.-H. [1982]: On the Selection of Critical Points and Local Curvature Extrema of Region Boundaries for Interframe Matching. In: *International Conference on Pattern Recognition*, pp. 542-544, 1982.

- Ebner, H. [1976]: Self Calibrating Block Adjustment. In: *International Archives of Photogrammetry*, vol. 21, part 3, Helsinki, 1976.
- Ebner, H., Hofmann-Wellenhof, B., Reiss, P., Steidler, F. [1980]: A Minicomputer Program Package for Height Interpolation by Finite Elements. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 23, B 4, pp. 202-215, Hamburg, 1980.
- Ebner, H., Fritsch, D., Gillessen, W., Heipke, C. [1987]: Integration von Bildzuordnung und Objektrekonstruktion innerhalb der digitalen Photogrammetrie. *Bildmessung und Luftbildwesen*, 55. Jahrgang, pp. 194-203, 1987.
- Enkelmann, W. [1986]: Investigations of Multigrid Algorithms for the Estimation of Optical Flow Fields in Image Sequences. In: *Proc. of the Workshop on Motion: Representation and Analysis*, Charleston, pp. 81-87, IEEE Computer Society Press, Washington, 1986.
- Enkelmann, W., Kories, R., Nagel, H.-H., Zimmermann, G. [1986]: An Experimental Investigation of Estimation Approaches for Optical Flow Fields. In: *Motion Understanding: Robot and Human Vision*, Kluwer Academic Publishers, Hingham/MA, 1985.
- Faugeras, O.D. [1990]: On the Motion of 3D Curves and its Relationship to Optical Flow. In: *Computer Vision - ECCV 90, Proceedings of the 1st European Conference on Computer Vision*, Faugeras, O. (ed.), Lecture Notes in Computer Science, vol. 427, pp. 107-117, Springer Verlag, Berlin, 1990.
- Faugeras, O.D., Hebert, M. [1983]: A 3-D Recognition and Positioning Algorithm using Geometrical Matching between Primitive Surfaces. In: *Proceedings of the International Joint Conference of Artificial Intelligence*, Karlsruhe, pp. 996-1002, August 1983.
- Faugeras, O.D., Maybank, S.J.. [1990]: Motion from Point Matches: Multiplicity of Solutions. *International Journal of Computer Vision*, vol. 4, no. 3, pp. 225-246, 1990.
- Fischler, M.A., Bolles, R.C. [1981]: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the Association of Computing Machinery*, vol. 24, no. 6, pp. 381-395, June 1981.
- Fischler, M.A., Firschein, O. (Eds.) [1987]: Readings in Computer Vision: Issues, Problems, Principles, and Paradigms. Morgan Kaufmann Publishers, Los Altos, California, 1987.
- Förstner, W. [1982]: On the Geometric Precision of Digital Correlation. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 24, part 3, pp. 176-189, Helsinki, 1982.
- Förstner, W. [1986]: A Feature Based Correspondence Algorithm for Image Matching. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 26, part 3/3, Rovaniemi, August 1986.
- Förstner, W. [1991]: Statistische Verfahren für die automatische Bildanalyse und ihre Bewertung bei der Objekterkennung und -vermessung. Deutsche Geodätische Kommission, Reihe C, Nr. 370, München, 1991.
- Förstner, W., Gülch, E. [1987]: A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: *Proc. ISPRS Intercommission Workshop on "Fast Processing of Photogrammetric Data"*, Interlaken, June 1987.
- Franek, M., Müller, J. [1990]: Regularizing Visible Surface Reconstruction with Facets Stereo Vision. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 28, part 3/2, pp. 132-144, Wuhan, May 1990.
- Frieß, P. [1990]: Kinematische Positionsbestimmung für die Aerotriangulation mit dem NAVSTAR Global Positioning System. Deutsche Geodätische Kommission, Reihe C, Nr. 359, München, 1990.
- Fu, K.-S., Gonzalez R.C., Lee, C.S.G. [1987]: Robotics: Control, Sensing, Vision and Intelligence. McGraw-Hill, New York, 1987.
- Gelb, A. (Ed.) [1974]: Applied Optimal Estimation. The MIT Press, Cambridge, Massachusetts, 1974.
- Gennert, M.A. [1988]: Brightness-Based Stereo Matching. In: *Proceedings of the 2nd International Conference on Computer Vision*, pp. 139-143, Tampa, FL, 1988.
- Gennery, D.B. [1992]: Visual Tracking of Known Three-Dimensional Objects. *International Journal of Computer Vision*, vol. 7, no. 3, pp. 243-270, 1992.
- Gibson, J.J. [1950]: The Perception of the Visual World. Houghton Mifflin Co., Boston, 1950.
- Gibson, E.J., Gibson, J.J., Smith, O.W., Flock, H. [1959]: Motion Parallax as a Determinant of Perceived Depth. *Journal of Experimental Psychology*, vol. 8, no. 1, pp. 40-51, 1959.
- Giraudon, G., Deriche, R. [1991]: On Corner and Vertex Detection. In: *Proc. of the Computer Vision and Pattern Recognition Conference*, pp. 650-655, Maui, Hawaii, June 1991.
- Glazer, F. [1984]: Multilevel Relaxation in Low-Level Computer Vision. In: *Multi-Resolution Image Processing and Analysis*, Rosenfeld, A. (ed.), pp. 312-330, Springer-Verlag, Berlin, 1984.
- Goldstein, H. [1989]: Klassische Mechanik, 10. korrigierte Auflage. Aula-Verlag, Wiesbaden, 1989.

- Gonzalez, R.C., Wintz, P. [1977]: Digital Image Processing. Addison-Wesley, Reading, Massachusetts, 1977.
- Grafarend, E.W. [1983]: Reference Frame Rotation - Regularized Theory by Quaternions and Spinors. In: *Geodesy in Transition*, Schwarz, K.P., Lachapelle, G. (eds.), pp. 185-225, The University of Calgary, Calgary, Alberta, July 1983.
- Grafarend, E.W., Schaffrin, B., Lohse, P. [1989]: Dreidimensionaler Rückwärtsschnitt. *Zeitschrift für Vermessungswesen*, 114. Jahrgang, pp. 61-67, 127-137, 172-175, 225-233, 279-287, 1989.
- Grimson, W.E.L. [1981]: From Images to Surfaces: A Computational Study of the Human Early Visual System. MIT Press, Cambridge, Massachusetts, 1981.
- Grün, A. [1985a]: Adaptive kleinste Quadrate Korrelation und geometrische Zusatzinformation. *Vermessung, Photogrammetrie und Kulturtechnik*, 83. Jahrgang, Heft 9, pp. 309-312, 1985.
- Grün, A. [1985b]: Algorithmic Aspects in On-Line Triangulation. *Photogrammetric Engineering and Remote Sensing*, vol. 51, no. 4, pp. 419-436, 1985.
- Grün, A., Kersten, Th. [1992]: Sequential Estimation in Robot Vision. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 29, B5, pp. 923-931, Washington, 1992.
- Gu, W.K., Yang, J.Y., Huang, T.S. [1984]: Matching Perspective Views of a Polyhedron Using Circuits. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 3, pp. 390-400, May 1984.
- Guiducci, A. [1988]: Corner Characterization by Differential Geometry Techniques. *Pattern Recognition Letters*, vol. 8, pp. 311-318, 1988.
- Haala, N. [1990]: Vergleich von Verfahren zur Expansionspunktbestimmung und Tiefenschätzung bei dichten Bildfolgen. Diplomarbeit (unveröffentlicht), Universität Stuttgart, Institut für Photogrammetrie, 1990.
- Hahn, M. [1989]: Automatic Measurement of Digital Terrain Models by Means of Image Matching Techniques. In: *Vorträge der 42. Photogrammetrischen Woche*, Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart, Heft 13, pp. 141-151, Stuttgart, 1989.
- Hahn, M. [1990]: Estimation of the Width of the Point-Spread Function within Image Matching. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 28, part 3/2, Wuhan, May 1990.
- Hahn, M. [1993]: Measurement by Image Matching - State-of-the-Art in Digital Photogrammetry. In: *Photogrammetric Week '93*, Fritsch, D., Hobbie, D. (eds.), pp. 33-42, Wichmann Verlag, Karlsruhe, 1993.
- Hahn, M. [1994]: Corner Detection in Range Images. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 30, part 5, pp. 136-143, Melbourne, Australia, March 1994.
- Hahn, M., Jäger, R. [1987]: Zuverlässigkeitsbeschreibung bei der Partitionierung von Netzen. *Vermessung, Photogrammetrie, Kulturtechnik*, 85. Jahrgang, pp. 102-105, März 1987.
- Hahn, M., Förstner, W. [1988]: The Applicability of a Feature Based and a Least Squares Matching Algorithm for DEM-Acquisition. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 27, B9, Kyoto, 1988.
- Hahn, M., Heck, B., Jäger, R., Scheuring, R. [1989]: Ein Verfahren zur Abstimmung der Signifikanzniveaus für allgemeine $F_{m,n}$ - verteilte Teststatistiken. Teil I: Theorie. *Zeitschrift für Vermessungswesen*, 114. Jahrgang, Heft 5, pp. 234-248, Mai 1989.
- Hahn, M., Heck, B., Jäger, R., Scheuring, R. [1991]: Ein Verfahren zur Abstimmung der Signifikanzniveaus für allgemeine $F_{m,n}$ - verteilte Teststatistiken. Teil II: Anwendungen. *Zeitschrift für Vermessungswesen*, 116. Jahrgang, Heft 1, pp. 15-26, Januar 1991.
- Hahn, M., Proß, E. [1992]: Bildzuordnung nach dem Variationsprinzip - ein Vergleich mit Kleinste Quadrate Verfahren. *Zeitschrift für Photogrammetrie und Fernerkundung*, 60. Jahrgang, Heft 4, pp. 116-123, Juli 1992.
- Hannah, M.J. [1974]: Computer Matching of Areas in Stereo Imagery. PhD thesis, Stanford University, AI-Memo 219, October 1974.
- Hannah, M.J. [1980]: Bootstrap Stereo. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 201-208, 1980.
- Haralick, R.M. [1980]: Edge and Region Analysis for Digital Image Data. *Computer Graphics and Image Processing*, vol. 12, no. 1, pp. 60-73, 1980.
- Harries, C.G. [1987]: Determination of Ego-Motion from Matched Points. In: *Proceedings of the 3rd Alvey Vision Conference*, pp. 189-192, Cambridge, UK, 1987.
- Hartl, Ph. [1989]: High Precision Navigation with Satellites. In: *High Precision Navigation*, Linkwitz, K. und Hangleiter, U. (eds.), pp. 3-14, Springer Verlag, Berlin, 1989.
- Heeger, D.J. [1986]: Depth and Flow from Motion Energy. In: *Proceedings of the Fifth National Conference on Artificial Intelligence*, Philadelphia, pp. 567-663, 1986.
- Heeger, D.J. [1988]: Optical Flow using Spatio-temporal Filters. *International Journal of Computer Vision*, vol. 1, no. 4, pp. 279-302, 1988.

- Heeger, D.J., Jepson, A.D. [1992]: Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation. *International Journal of Computer Vision*, vol. 5, no. 2, pp. 95-117, 1992.
- Heikkilä, J. [1989]: Multiscale Representation with Förstner Operator. *The Photogrammetric Journal of Finland*, vol. 11, no. 2, pp. 40-59, 1989.
- Heipke, C. [1990]: Integration von Bildzuordnung, Punktbestimmung, Oberflächenrekonstruktion und Orthoprojektion innerhalb der digitalen Photogrammetrie. Deutsche Geodätische Kommission, Reihe C, Nr. 366, München, 1990.
- Helava, U.V. [1988]: Object Space Least Squares Correlation. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 27, B3, Kyoto, Japan, 1988.
- Hildreth, E. [1984]: Computations Underlying the Measurement of Visual Motion. *Artificial Intelligence*, vol. 23, no. 3, pp. 309-354, August 1984.
- Hinsken, L. [1987]: Algorithmen zur Beschaffung von Näherungswerten für die Orientierung von beliebig in Raum angeordneten Strahlenbündeln. Deutsche Geodätische Kommission, Reihe C, Nr. 333, München, 1987.
- Hinsken, L. [1988]: A Singularity Free Algorithm for Spatial Orientation of Bundles. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 27, B5, Kyoto, 1988.
- Hirt, U. [1990]: Entwicklung eines 3D-Kalmanfilters für die Analyse einer CCD-Bildfolge. Diplomarbeit (unveröffentlicht), Universität Stuttgart, Institut für Photogrammetrie, 1990.
- Hochberg, J. [1987]: Machines should not see as people do, but must know how people see. *Computer Vision, Graphics, and Image Processing*, vol. 37, pp. 221-237, 1987.
- Hoff, W., Ahuja, N. [1986]: Surfaces from Stereo. In: *Proc. of the 8th International Conference on Pattern Recognition*, pp. 516-518, Paris, 1986.
- Horn, B.K.P. [1977]: Image Intensity Understanding. *Artificial Intelligence*, vol. 8, no. 2, pp. 201-231, 1977.
- Horn, B.K.P. [1986]: Robot Vision. The MIT Press, Cambridge, Massachusetts, 1986.
- Horn, B.K.P. [1987]: Closed-form Solution of Absolute Orientation using Unit Quaternions. *Journal of Optical Society of America A*, vol. 4, no. 4, pp. 629-642, April 1987.
- Horn, B.K.P. [1990]: Relative Orientation. *International Journal of Computer Vision*, vol. 4, pp. 59-78, 1990.
- Horn, B.K.P., Schunk, B. [1981]: Determining Optical Flow. *Artificial Intelligence*, vol. 17, pp. 185-203, August 1981.
- Horn, B.K.P., Weldon, E.J.(Jr.) [1988]: Direct Methods for Recovering Motion. *International Journal of Computer Vision*, vol. 2, pp. 51-76, 1988.
- Huang, T.S. (Ed.) [1981]: Image Sequence Analysis. Springer-Verlag, Berlin, 1981.
- Huang, T.S. (Ed.) [1983]: Image Sequence Processing and Dynamic Scene Analysis. Springer-Verlag, Berlin, 1983.
- Huang, T.S. [1990]: Computer Vision. In: *Tutorial on Computer Vision and Dynamic Scene Analysis*, International Society for Photogrammetry and Remote Sensing, Zurich, Sept. 1990.
- Huang, T.S., Faugeras, O.D. [1989]: Some Properties of the E-matrix in Two-view Motion Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1310-1312, 1989.
- Huang, T.S., Netravali, A.N. [1990]: Motion and Structure from Feature Correspondences: A Review. In: *Tutorial on Computer Vision and Dynamic Scene Analysis*, 55 p., International Society of Photogrammetry and Remote Sensing, Zurich, September 1990.
- ICCV1: First International Conference on Computer Vision [1987], J.M. Brady and A. Rosenfeld (eds.), London, 8-11 June 1987, IEEE Computer Society Press, Washington.
- ICCV2: Second International Conference on Computer Vision [1988], R. Bajcsy and S. Ullman (eds.), Tampa, 5-8 December 1988, IEEE Computer Society Press, Washington.
- ICCV3: Third International Conference on Computer Vision [1990], M. Nagao (ed.), Osaka, 4-7 December 1990, IEEE Computer Society Press, Los Alamitos, CA.
- ICCV4: Fourth International Conference on Computer Vision [1993], H.-H. Nagel (ed.), Berlin, 11-14 May 1993, IEEE Computer Society Press, Los Alamitos, CA.
- Jähne, B. [1989]: Digitale Bildverarbeitung. Springer Verlag, Berlin, 1989.
- Jerian, C., Jain, R. [1988]: Polynomial Methods for Structure from Motion. In: *Proceedings of the 2nd International Conference on Computer Vision*, pp. 197-206, Tampa, FL, 1988.
- Julesz, B. [1971]: Foundations of Cyclopean Perception. The University of Chicago Press, Chicago, 1971.
- Kalman, R.E. [1960]: A new Approach to Linear Filtering and Prediction Problems. *ASME Journal of Basic Engineering*, vol. 82, pp. 34-45, 1960.
- Kanade, T. [1978]: Region Segmentation: Signal vs. Semantics. In: *Proc. of the 4th International Joint Conference on Pattern Recognition*, pp. 95-105, Kyoto, 1978.

- Kanatani, K. [1990]: Group-Theoretical Methods in Image Understanding. Springer Series in Information Sciences, vol. 20, Springer-Verlag, 1990.
- Kitchen, L., Rosenfeld, A. [1982]: Gray-level Corner Detection. *Pattern Recognition Letters*, vol. 1, pp. 95-102, December 1982.
- Koch, K.R. [1980]: Parameterschätzung und Hypothesentests in linearen Modellen. Dümmler Verlag, Bonn, 1980.
- Koenderink, J.J., van Doorn, A.J. [1976]: Local Structure of Movement Parallax of the Plane. *Journal of the Optical Society of America*, vol. 66, no. 7, pp. 717-723, July 1976.
- Koller, D., Nagel, H.-H. [1990]: Ein robustes Verfahren zur Detektion und Verfolgung bewegter Objekte in Bildfolgen. In: *Mustererkennung 1990, Proceedings des 12. Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung*, Großkopf, R.E. (ed.), Informatik Fachberichte, vol. 254, pp. 625-633, Springer Verlag, Berlin, September 1990.
- Kories, R.R. [1985]: Maschinelles Bewegungssehen in natürlichen Szenen: Die Auswertung von Bildfolgen gestützt auf Bildmerkmale. Dissertation an der Universität Mainz, Fachbereich Biologie, 1985.
- Kories, R.R. [1988]: Analysis and Interpretation of Image Sequences. Tutorial notes, Berlin, 1988.
- Kories, R.R., Zimmermann, G. [1986]: A Versatile Method for the Estimation of Displacement Vector Fields from Image Sequences. In: *Proc. of the Workshop on Motion: Representation and Analysis*, Charleston, pp. 101-106, May 1986.
- Korn, A. [1986]: Invariante Formbeschreibung in verschiedenen Auflösungsebenen. In: *Mustererkennung 1986, Proceedings des 8. Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung*, Hartmann, G. (ed.), Informatik Fachberichte, pp. 89-93, Springer Verlag, Berlin, 1986.
- Korn, G. A., Korn, T. M. [1968]: Mathematical Handbook for Scientists and Engineers, 2nd Edition. McGraw-Hill, New York, N.Y., 1968.
- Kreiling, W. [1976]: Automatische Herstellung von Höhenmodellen und Orthophotos aus Stereobildern durch digitale Korrelation. Dissertation, TU Karlsruhe, 1976.
- Kuan, D., Phipps, G., Hsueh, A.-Ch. [1987]: Autonomous Land Vehicle Road Following. In: *Proceedings of the 1st International Conference on Computer Vision*, pp. 557-566, London, 1987.
- Lang, F. [1992]: Bildzuordnung im Scale-Space. Diplomarbeit (unveröffentlicht), Universität Stuttgart, Institut für Photogrammetrie, 1992.
- Li, M. [1989]: Hierarchical Multi-point Matching with Simultaneous Detection and Location of Breaklines. Dissertation, Royal Institute of Technology, Department of Photogrammetry, Stockholm, Sweden, 1989.
- Limb, J.O., Murphy, J.A. [1975]: Estimating the Velocity of Moving Images in Television Signals. *Computer Graphics and Image Processing*, vol. 4, pp. 311-327, 1975.
- Lin, X., Zhu, Z. [1990]: Detecting Height from Constrained Motion. In: *Proceedings of the 3rd International Conference on Computer Vision*, pp. 503-506, Osaka, 1990.
- Lindeberg, T. [1991]: Discrete Scale-Space Theory and the Scale-Space Primal Sketch. Dissertation TRITA-NA-P9108, Royal Institute of Technology, Department of Numerical Analysis and Computing Science, Stockholm, Sweden, May 1991.
- Lindeberg, T., Gårding, J. [1993]: Shape from Texture from a Multi-Scale Perspective. In: *Proceedings of the 4th International Conference on Computer Vision*, pp. 683-691, Berlin, 1993.
- Liu, Y., Huang, T.S., Faugeras, O.D. [1988]: Determination of Camera Location from 2D to 3D Line and Point Correspondences. In: *Proc. of the Computer Vision and Pattern Recognition Conference*, Ann Arbor MI, 1988.
- Longuet-Higgins, H.C. [1981]: A Computer Algorithm for Reconstructing a Scene from Projections. *Nature*, vol. 293, pp. 133-135, Sept. 1981.
- Longuet-Higgins, H.C. [1984]: The Visual Ambiguity of a Moving Plane. *Proc. Royal Society London, Series B*, vol. 223, no. 1, pp. 165-170, 1984.
- Longuet-Higgins, H.C., Prazdny, K. [1980]: The Interpretation of a Moving Retinal Image. *Proc. Royal Society London, Series B*, vol. 208, pp. 385-397, July 1980.
- Lowe, D.G. [1991]: Fitting Parameterized Three-dimensional Models to Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 5, pp. 441-450, May 1991.
- Lowe, D.G. [1992]: Robust Model-based Motion Tracking Through the Integration of Search and Estimation. *International Journal of Computer Vision*, vol. 8, no. 2, pp. 113-122, 1992.
- Marr, D. [1978]: Representing Visual Information. In: *Computer Vision Systems*, Hanson, A.R., Riseman, E.M. (eds.), Academic Press, New York, 1978.
- Marr, D. [1982]: Vision. W.H. Freeman, San Francisco, California, 1982.
- Marr, D., Nishihara, H.K. [1978]: Visual Information Processing: Artificial Intelligence and the Sensorium of Sight. *Technology Review*, vol. 81, no. 1, pp. 2-23, 1978.

- Marr, D., Poggio, T. [1979]: A Computational Theory of Human Stereo Vision. *Proc. of the Royal Society of London, Series B*, vol. 204, pp. 301-328, 1979.
- Marr, D., Hildreth, E.C. [1980]: Theory of Edge Detection. *Proc. Royal Society of London, Series B*, vol. 207, pp. 187-217, 1980.
- Marzan, G.T., Karara, H.M. [1975]: A Computer Program for the Direct Linear Transformation Solution of the Collinearity Condition, and some Applications of it. In: *ASP Symposium on Close-Range Photogrammetric Systems*, 1975.
- Matthies, L.H. [1989]: Dynamic Stereo Vision. PhD Thesis, Carnegie Mellon University, CMU-CS-89-195, October 1989.
- Matthies, L.H., Szelisky, R., Kanade, T. [1988]: Incremental Estimation of Dense Depth Maps from Image Sequences. In: *Proc. Computer Vision and Pattern Recognition Conference*, Ann Arbor MI, 1988.
- Matthies, L.H., Kanade, T., Szelisky, R. [1989]: Kalman Filter-based Algorithms for Estimating Depth from Image Sequences. *International Journal of Computer Vision*, vol. 3, pp. 209-236, 1989.
- Maybeck, P.S. [1979]: Stochastic Models, Estimation, and Control. Vol. 1, Academic Press, New York, 1979.
- Mayhew, J.E.W., Frisby, J.P. [1981]: Psychophysical and Computational Studies toward a Theory of Human Stereopsis. *Artificial Intelligence*, vol. 17, pp. 349-385, 1981.
- Meer, P., Baugher, E.S., Rosenfeld, A. [1987]: Frequency Domain Analysis and Synthesis of Image Pyramid Generating Kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 6, pp. 512-522, 1991.
- Mehrotra, R., Nichani, S. [1990]: Corner Detection. *Pattern Recognition*, vol. 23, no. 11, pp. 1223-1233, 1990.
- Merminod, B., Rizos, C. [1988]: A Surveyor's Perspective of Kalman Filtering. *Austr. J. Geod. Photogram. Surv.*, no. 49, pp. 1-38, Dec. 1988.
- Meygret, A., Thonnat, M. [1990]: Segmentation of Optical Flow and 3D Data for the Interpretation of Mobile Objects. In: *Proceedings of the 3rd International Conference on Computer Vision*, pp. 238-245, Osaka, 1990.
- Mezger, K. [1988]: Automatische Führung von Binnenschiffen durch Bordradar-Bildauswertung. Dissertation an der Universität Stuttgart, VDI-Verlag, 1988.
- Michael, D.J. [1990]: Direct Non-Linear Methods for Recovering Structure and Motion. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 292-298, San Mateo, CA, Sept. 1990.
- Mikhail, E.M. [1984]: Photogrammetric Target Location to Subpixel Accuracy in Digital Images. In: *Vorträge der 39. Photogrammetrischen Woche*, Sept. 1983. Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart, Heft 9, Stuttgart, 1984.
- Moravec, H.P. [1977]: Towards Automatic Visual Obstacle Avoidance. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, p. 584, 1977.
- Moravec, H.P. [1979]: Visual Mapping by a Robot Rover. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, Tokio, pp. 598-600, 1979.
- Moravec, H.P. [1981]: Rover Visual Obstacle Avoidance. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, Vancouver, B.C., pp. 785-790, 1981.
- Müller, B., Hahn, M. [1992]: Parallel Processing - The Example of Automatic Relative Orientation. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 29, B2, pp. 623-630, Washington, 1992.
- Nagel, H.-H. [1983]: Displacement Vectors Derived from Second-Order Intensity Variations in Image Sequences. *Computer Vision, Graphics, and Image Processing*, vol. 21, no. 1, pp. 85-117, January 1983.
- Nagel, H.-H. [1985]: Analyse und Interpretation von Bildfolgen. *Informatik Spektrum*, vol. 8, pp. 178-200, 312-327, 1985.
- Nagel, H.-H. [1987]: On the Estimation of Optical Flow: Relations between Different Approaches and some New Results. *Artificial Intelligence*, vol. 33, pp. 299-324, 1987.
- Nagel, H.-H. [1989]: On a Constraint Equation for the Estimation of Displacement Rates in Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 13-30, January 1989.
- Nagel, H.-H., Enkelmann, W. [1986]: An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 565-593, Sept. 1986.
- Negahdaripour, S., Horn, B.K.P. [1987]: Direct Passive Navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 168-176, January 1987.
- Negahdaripour, S., Horn, B.K.P. [1989]: A Direct Method for Locating the Focus of Expansion. *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 3, pp. 303-326, June 1989.
- Negahdaripour, S., Lee, S. [1992]: Motion Recovery from Image Sequences Using Only First Order Optical Flow Information. *International Journal of Computer Vision*, vol. 9, no. 3, pp. 163-184, 1992.

- Neumann, B. [1988]: Szenenrekonstruktion aus Bildsequenzen. In: *Bildsignalverarbeitung für Aufgaben der Flugführung*, Kolloquium des Instituts für Flugführung, Braunschweig, Mai 1988.
- Neumann, B. [1989]: Wissensbasierte Bewegtbildanalyse. Vortrag beim Arbeitsgespräch "Wissensbasierte Bildanalyse" des SFB 228, Stuttgart, Mai 1989.
- Neumann, B., Novak, H.-J. [1986]: NAOS: Ein System zur natürlichsprachlichen Beschreibung zeitveränderlicher Szenen. *Informatik Forschung und Entwicklung*, vol. 1, pp. 83-92, 1986.
- Niemann, H. [1985]: Wissensbasierte Bildanalyse. *Informatik Spektrum*, vol. 8, pp. 201-214, 1985.
- Niemcz, W. [1963]: Anwendung der Quaternionen auf die allgemeinen Bewegungsgleichungen der Flugmechanik. *Zeitschrift Flugwiss.*, vol. 11, no. 9, pp. 368-372, 1963.
- Noble, J.A. [1988]: Finding Corners. *Image and Vision Computing*, vol. 6, no. 2, pp. 121-128, 1988.
- Novak, K. [1991]: Integration von GPS und digitalen Kameras zur automatischen Vermessung von Verkehrswegen. *Zeitschrift für Photogrammetrie und Fernerkundung*, 59. Jahrgang, Heft 4, pp. 112-120, Juli 1991.
- Paderes, F.C., Mikhail, E.M., Förstner, W. [1984]: Rectification of Single and Multiple Frames of Satellite Scanner Imagery using Point and Edges as Control. In: *Proc. NASA Symposium on Mathematical Pattern Recognition and Image Analysis*, Houston, 1984.
- Pauli, J. [1992]: Erklärungs-basiertes Computer-Sehen von Bildfolgen. Dissertation an der Technischen Universität München, Institut für Informatik, 1992.
- Pentland, A.P.(Ed.) [1986]: From Pixels to Predicates. Ablex Publishing Corporation, Norwood, New Jersey, 1986.
- Pertl, A. [1984]: Digital Image Correlation with the Analytical Plotter Planicomp C100. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 25, part 3, pp. 874-881, Rio de Janeiro, Brazil, 1984.
- Philip, J. [1991]: Estimation of Three-Dimensional Motion of Rigid Objects from Noisy Observations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 61-66, January 1991.
- Poggio, T., Torre, V., Koch, Ch. [1985]: Computational Vision and Regularization Theory. *Nature*, vol. 317, pp. 314-319, September 1985.
- Pope, A [1970]: An Advantageous Alternative Parametrization of Rotations for Analytical Photogrammetry. Technical Report, Coast and Geodetic Survey CaGS39, US Department of Commerce ESSA, Rockville, Maryland, September 1970.
- Pope, A. [1974]: Two Approaches to Nonlinear Least Squares. *The Canadian Surveyor*, vol. 28, no. 5, pp. 663-669, Dec. 1974.
- Prazdny, K. [1981]: Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinearly Moving Observer. *Computer Graphics and Image Processing*, vol. 17, no. 3, pp. 238-248, November 1981.
- Rangarajan, K., Shah, M., van Brackle, D. [1989]: Optimal Corner Detector. *Computer Vision, Graphics, and Image Processing*, vol. 48, pp. 230-245, 1989.
- Rauhala, U. [1986]: Compiler Positioning of Array Algebra Technology. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 26, part 3/3, pp. 173-198, Rovaniemi, August 1986.
- Rauhala, U. [1988]: Compiler Positioning System: An Array Algebra Formulation of Digital Photogrammetry. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 27, B9, pp. 151-161, Kyoto, 1988.
- Rinner, K. [1963]: Studien über die allgemeine, voraussetzungslose Lösung des Folgebildanschlusses. *Österreichische Zeitschrift für Vermessungswesen*, Sonderheft 23, 1963.
- Roberts, B., Bahnu, B. [1990]: Inertial Navigation Sensor Integrated Motion Analysis for Autonomous Vehicle Navigation. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 364-375, San Mateo, CA, Sept. 1990.
- Roberts, L.G. [1965]: Machine Perception of Three-Dimensional Solids. In: *Optical and Electrooptical Information Processing*, Tippelt, J.T. (ed.), pp. 159-197, The MIT Press, Cambridge, Massachusetts, 1965.
- Rohr, K. [1990]: Über die Modellierung und Identifikation charakteristischer Grauwertverläufe in Realweltbildern. In: *Mustererkennung 1990, Proceedings des 12. Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung*, Großkopf, R.E. (ed.), Informatik Fachberichte, vol. 254, pp. 217-224, Springer Verlag, Berlin, September 1990.
- Rohr, K. [1992]: Recognizing Corners by Fitting Parametric Models. *International Journal of Computer Vision*, vol. 9, no. 3, pp. 213-230, 1992.
- Rohr, K., Nagel, H.-H. [1990]: Modellgestützte Bestimmung des Bewegungszustandes von Fußgängern in Realweltbildfolgen. In: *Mustererkennung 1990, Proceedings des 12. Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung*, Großkopf, R.E. (ed.), Informatik Fachberichte, vol. 254, pp. 52-58, Springer Verlag, Berlin, 1990.
- Rosenfeld, A. (Ed.) [1984]: Multiresolution Image Processing. Springer-Verlag, Berlin, 1984.
- Rosenholm, D. [1986]: Accuracy Improvement of Digital Matching for Evaluation of Digital Terrain Models. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 26, part 3, pp. 573-587, Rovaniemi, 1986.

- Rosenholm, D. [1988]: Multi-point Matching Along Vertical Lines in SPOT Images. *International Journal of Remote Sensing*, vol. 9, no. 10 & 11, pp. 1687-1703, 1988.
- Rutkowski, W.S., Rosenfeld, A. [1978]: A Comparison of Corner Detection Techniques for Chain-coded Curves. Technical Report 623, MD 20742, Computer and Science Center, University of Maryland, College Park, January 1978.
- Salzmann, M.A. [1988]: Some Aspects of Kalman Filtering. Technical Memorandum E3B5A3, Dept. of Surveying Engineering, Univ. of New Brunswick, 1988.
- Sanso, F. [1973]: An Exact Solution of the Roto-translation Problem. *Photogrammetria*, vol. 29, pp. 203-216, 1973.
- Sawhney, H.S., Oliensis, J., Hanson, A.R. [1990]: Description and Reconstruction from Image Trajectories of Rotational Motion. In: *Proceedings of the 3rd International Conference on Computer Vision*, pp. 494-498, Osaka, 1990.
- Schaffrin, B. [1985]: Das geodätische Datum mit stochastischer Vorinformation. Deutsche Geodätische Kommission, Reihe C, Nr. 313, München, 1985.
- Schenk, T., Toth, C., Li, J.-C. [1990]: Zur automatischen Orientierung von digitalen Bildpaaren. *Zeitschrift für Photogrammetrie und Fernerkundung*, 58. Jahrgang, Heft 6, pp. 182-189, 1990.
- Schilling, W. (Ed.) [1987]: Anwendungsmöglichkeiten des Kalman-Filter-Verfahrens in der Wasserwirtschaft. Deutsche Forschungsgemeinschaft, VCH Verlagsgesellschaft, Weinheim, 1987.
- Schneider, F., Hahn, M. [1992]: Matching Images of Different Geometric Scale. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 29, B3, pp. 295-302, Washington, 1992.
- Schnörr, C. [1992]: Computation of Discontinuous Optical Flow by Decomposition and Shape Optimization. *International Journal of Computer Vision*, vol. 8, no. 2, pp. 163-165, 1992.
- Schut, G.H. [1957/58]: An Analysis of Methods and Results in Analytical Aerotriangulation. *Photogrammetria*, pp. 16-32, 1957/58.
- Schut, G.H. [1958/59]: Construction of Orthogonal Matrices and their Application in Analytical Photogrammetry. *Photogrammetria*, no. 4, pp. 149-162, 1958/59.
- Schut, G.H. [1961]: On Exact Linear Equations of Rotational Elements in Absolute Orientation. *Photogrammetria*, vol. 17, no. 1, pp. 34-37, 1961.
- Schwarz, K.P. [1992]: Inertial Techniques in Geodesy - State of the Art and Trends. In: *High Precision Navigation 91*, Linkwitz, K. und Hangleiter, U. (eds.), pp. 423-440, Dümmler Verlag, Bonn, 1992.
- Schwidersky, K., Ackermann, F. [1976]: Photogrammetrie, 7. Auflage. B.G. Teubner Verlag, Stuttgart, 1976.
- Selfridge, O.G. [1955]: Eyes and Ears for the Computer. In: *Proceedings of the Western Joint Computer Conference*, pp. 91-93, 1955.
- Shih, T.-Y. [1990]: The Duality and Critical Condition in the Formulation and Decomposition of a Rotation Matrix. *Photogrammetric Engineering and Remote Sensing*, vol. 56, no. 8, pp. 1173-1179, Aug. 1990.
- Smith, G.B. [1986]: Stereo Integral Equation. In: *Proceedings of the Fifth National Conference on Artificial Intelligence*, pp. 689-694, Los Altos, California, 1986.
- Snyder, M.A. [1989]: On the Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 1004-1011, San Mateo, CA, 1989.
- Snyder, M.A. [1991]: On the Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1105-1114, 1991.
- Sommer, G. [1992]: Wissensverarbeitung in der Bildanalyse - Teile I-III. *Bild und Ton*, Band 45, Hefte 5/6, 7/8, 11/12, pp. 103-110, 151-156, 314-319, 1992.
- Spetsakis, M.E., Aloimonos, J. [1990]: A Unified Theory for Structure from Motion. In: *Proc. of the DARPA Image Understanding Workshop*, pp. 271-283, San Mateo, CA, Sept. 1990.
- Spetsakis, M., Aloimonos, J. [1991]: A Multi-frame Approach to Visual Motion Perception. *International Journal of Computer Vision*, vol. 6, no. 3, pp. 245-255, 1991.
- Stefanovic, P. [1973]: Relative Orientation - A New Approach. *ITC-Journal*, vol. 3, 1973.
- Stuelpnagel, J. [1964]: On the Parametrization of the Three-Dimensional Rotation Group. *SIAM Review*, vol. 6, no. 4, pp. 422-430, October 1964.
- Szelisky, R. [1990]: Bayesian Modelling of Uncertainty in Low-Level Vision. *International Journal of Computer Vision*, vol. 5, no. 3, pp. 271-301, 1990.
- Taalebinezhaad, M.A. [1990]: Direct Recovery of Motion and Shape in the General Case by Fixation. In: *Proceedings of the 3rd International Conference on Computer Vision*, pp. 451-455, Osaka, 1990.

- Tempfli, K. [1982]: Genauigkeitsschätzung digitaler Höhenmodelle mittels Spektralanalyse. Dissertation, Geowissenschaftliche Mitteilungen Heft 49, Institut für Photogrammetrie der Technischen Universität Wien, Mai 1982.
- Terzopoulos, D. [1986]: Regularization of Inverse Visual Problems Involving Discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 413-424, 1986.
- Terzopoulos, D. [1988]: The Computation of Visible Surface Representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 417-438, 1988.
- Teunissen, P.J.G., Salzmann, M.A. [1988]: Performance Analysis of Kalman Filters. Paper presented at HYDRO88, Amsterdam, November 1988.
- Thompson, E.H. [1958/59]: An Exact Solution of the Problem of Absolute Orientation. *Photogrammetria*, no. 4, pp. 163-178, 1958/59.
- Thompson, E.H. [1959]: A Method for the Construction of Orthogonal Matrices. *Photogrammetric Record*, vol. 3, no. 13, pp. 55-59, April 1959.
- Thorpe, Ch.E. [1984]: An Analysis of Interest Operators for FIDO. In: *Proc. of the Workshop on Computer Vision: Representation and Control*, pp. 135-140, IEEE Computer Society Press, Washington, 1984.
- Thorpe, Ch., Amidi, O., Gowdy, J., Hebert, M., Pomerleau, D. [1992]: Integrating Position Measurement and Image Understanding for Autonomous Navigation. In: *High Precision Navigation 91*, Linkwitz, K. und Hangleiter, U. (eds.), pp. 639-660, Dümmler Verlag, Bonn, 1992.
- Thurgood, J.D., Mikhail, E.M. (1982): Subpixel Mensuration of Photogrammetric Targets in Digital Images. Technical Report, School of Civil Engineering, Purdue University.
- Tikhonov, A.N., Arsenin, V.Y. [1977]: Solutions of Ill-Posed Problems. Winston, Washington, D.C., 1977.
- Tsai, R.Y., Huang, T.S., Zhu, W.L. [1982]: Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, II: Singular Value Decomposition. *IEEE Transactions on Acoustics, Speech, Signal Proc.*, vol. 30, pp. 525-534, 1982.
- Tsai, R.Y., Huang, T.S. [1984]: Uniqueness and Estimation of 3-D Motion Parameters of Rigid Bodies with curved Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 13-27, 1984.
- Tsotsos, J.K. [1987]: A 'Complexity Level' Analysis of Vision. In: *Proceedings of the 1st International Conference on Computer Vision*, pp. 346-355, London, 1987.
- Ullman, S. [1979]: The Interpretation of Visual Motion. The MIT Press, Cambridge, Massachusetts, 1979.
- Ullman, S. [1987]: Analysis of Visual Motion by Biological and Computer Systems. In: *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, Fischler, M.A. and Firschein, O. (eds.), pp. 132-144, Morgan Kaufmann Publishers, Los Altos, California, 1987.
- Umeyama, S. [1991]: Least-Squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376-380, April 1991.
- Verri, A., Poggio T. [1987]: Against Quantitative Optical Flow. In: *Proceedings of the 1st International Conference on Computer Vision*, pp. 171-180, London, 1987.
- Verri, A., Poggio T. [1989]: Motion Field and Optical Flow: Qualitative Properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 490-498, May 1989.
- Vosselman, G. [1992]: Relational Matching. Lecture Notes in Computer Science, vol. 628, Springer Verlag, Berlin, 1992.
- Waltz, D.L. [1975]: Understanding Drawings of Scenes with Shadows. In: *The Psychology of Computer Vision*, P.H. Winston (ed.), pp. 19-91, McGraw-Hill, New York, 1975.
- Waxman, A.M. [1987]: An Image Flow Paradigm. In: *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, Fischler, M.A. and Firschein, O. (eds.), pp. 145-168, Morgan Kaufmann Publishers, Los Altos, California, 1987.
- Waxman, A.M., Wohn, K. [1985]: Contour Evolution, Neighborhood Deformation, and Global Image Flow: Planar Surfaces in Motion.. *The International Journal of Robotics Research*, vol. 4, no. 3, pp. 95-108, 1985.
- Waxman, A.M., Kamgar-Parsi, B., Subbarao, M. [1987]: Closed-Form Solutions to Image Flow Equations. In: *Proceedings of the 1st International Conference on Computer Vision*, pp. 12-24, London, 1987.
- Wehr, A. [1991]: Entwicklung und Erprobung von opto-elektronischen Entfernungsmesssystemen mit CW-Halbleiters Lasern. Dissertation, Universität Stuttgart, Institut für Navigation, Stuttgart, 1991.
- Weinstock, R. [1974]: Calculus of Variations: With Applications to Physics and Engineering. Dover Publications, New York, 1974.
- Weisensee, M. [1991]: Modelle und Algorithmen für das Facetten-Stereosehen. Deutsche Geodätische Kommission, Reihe C, Nr. 374, München, 1991.
- Weng, J., Huang, T.S., Ahuja, N. [1987]: 3-D Motion Estimation, Understanding, and Prediction from Noisy Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 3, pp. 370-389, May 1987.

- Weng, J., Huang, T.S., Ahuja, N. [1989]: Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 451-476, May 1989.
- Weng, J., Huang, T.S., Ahuja, N. [1992]: Motion and Structure from Line Correspondences: Closed-form Solution, Uniqueness, and Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 318-336, May 1992.
- Weng, J., Ahuja, N., Huang, T.S. [1992]: Matching Two Perspective Views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 806-825, August 1992.
- Witkin, A.P. [1981]: Recovering Surface Shape and Orientation from Texture. *Artificial Intelligence*, vol. 17, nos. 1-3, pp. 17-45, August 1981.
- Witkin, A.P. [1983]: Scale Space Filtering. In: *Proceedings of the International Conference of Artificial Intelligence*, Karlsruhe, pp. 1019-1022, 1983.
- Witkin, A.P., Terzopoulos, P., Kass, M. [1987]: Signal Matching Through Scale Space. *International Journal of Computer Vision*, vol. 1, pp. 133-144, 1987.
- Wittenburg, J. [1977]: Dynamics of Systems of Rigid Bodies. B.G. Teubner Verlag, Stuttgart, 1977.
- Wohn, K.Y., Wu, J., Brockett, R.W. [1991]: A Contour-Based Recovery of Image Flow: Iterative Transformation Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 746-760, August 1991.
- Wrobel, B. [1987]: Digitale Bildzuordnung durch Facetten mit Hilfe von Objektraummodellen. *Bildmessung und Luftbildwesen*, 55. Jahrgang, Heft 3, pp. 93-101, Mai 1987.
- Wrobel, B.P. [1988]: Least Squares Methods for Surface Reconstruction from Images. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 27, B3, pp. 806-821, Kyoto, Japan, 1988.
- Wrobel, B.P. [1989]: Geometrisch-physikalische Grundlagen der digitalen Photogrammetrie. In: *Vorträge der 42. Photogrammetrischen Woche*, Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart, Heft 13, pp. 223-241, Stuttgart, Sept. 1989.
- Wrobel, B., Klemm, D. [1984]: Über die Vermeidung singulärer Fälle bei der Berechnung allgemeiner räumlicher Drehungen. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 25, part A3b, pp. 1153-1163, Rio de Janeiro, 1984.
- Wünsche, H.-J. [1988]: Bewegungssteuerung durch Rechnersehen. *Fachberichte Messen, Steuern, Regeln*, vol. 20, Springer Verlag, Berlin, 1988.
- Xiao, J., Liu, J., Chu, L. [1988]: Digital Matching of SPOT Stereo Images by Finite Elements Least Squares Techniques. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 27, B9, pp. 373-382, Kyoto, 1988.
- Zhang, Z., Faugeras, O.D. [1992]: Three-Dimensional Motion Computation and Object Segmentation in a Long Sequence of Stereo Frames. *International Journal of Computer Vision*, vol. 7, no. 3, pp. 211-241, 1992.
- Zhang, Z., Faugeras, O. [1992]: 3D Dynamic Scene Analysis. *Springer Series in Information Sciences*, vol. 27, Springer-Verlag, Berlin, 1992.
- Zheng, Y.-J. [1991]: Inverse und schlecht gestellte Probleme in der photogrammetrischen Rekonstruktion. *Deutsche Geodätische Kommission, Reihe C, Nr. 390*, München, 1991.
- Zheng, Y.-J., Hahn, M. [1990]: Surface Reconstruction from Digital Images in Presence of Discontinuities, Occlusions and Deformations. In: *International Archives of Photogrammetry and Remote Sensing*, vol. 28, part 3/2, pp. 1121-1144, Wuhan, May 1990.
- Zimmermann, G., Kories, R.R., [1984]: Eine Familie von Bildmerkmalen für die Bewegungsbestimmung in Bildfolgen. In: *Mustererkennung 1984, Proceedings des 6. Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung*, Kropatsch, W. (ed.), *Informatik Fachberichte*, vol. 87, pp. 147-153, Springer Verlag, Berlin, 1984.
- Zuniga, O.A., Haralick, R.M. [1983]: Corner Detection Using the Facet Model. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 30-37, 1983.

Danksagung

Mein besonderer Dank gilt Prof. Dr. Dr.h.c.mult. Friedrich Ackermann für die Betreuung dieser Arbeit. Seine Ratschläge haben die Arbeit inhaltlich wesentlich beeinflusst, seine kritischen Anmerkungen und Verbesserungsvorschläge dazu beigetragen, die Arbeit in die endgültige Form zu bringen. Dankbar bin ich Prof. Ackermann darüberhinaus für die Zusammenarbeit, in der ich viele der von ihm wesentlich mitgestalteten Themen der modernen Photogrammetrie kennenlernen konnte. Die 4-jährige Assistententätigkeit am Institut beinhaltete für mich eine solide Ausbildung im gesamten Bereich der analytischen Photogrammetrie, aber auch der GPS-Photogrammetrie und der Erfassung von Digitalen Geländemodellen durch Stereobildzuordnungsverfahren und Laserabtastung. Meine Aufgaben in der Folgezeit als Leiter der Gruppe digitale Photogrammetrie führten zur Konzentration auf die von Prof. Ackermann sehr intensiv geförderten Forschungsthemen der Bildanalyse und des Bildverstehens. Einer dieser Bereiche, die Bildsequenzanalyse, ist Gegenstand der vorliegenden Arbeit.

Ebenso herzlich danke ich Prof. Dr. Dieter Fritsch und Prof. Dr. Wolfgang Förstner für die Übernahme der Korreferate. Prof. Förstner hat noch während der gemeinsamen Zeit in Stuttgart die Aufgabenstellung mitdefiniert und sich auch danach stets am Thema interessiert gezeigt. In seiner Funktion als neuer Institutsleiter in Stuttgart hat mir Prof. Fritsch die notwendigen Freiräume zur Fertigstellung der Arbeit gewährt. Wertvolle Literaturhinweise, insbesondere zur Parametrisierung von Rotationen und Modellierung von Rotationsbewegungen, verdanke ich Prof. Dr. Dr.h.c. Erik W. Grafarend.

Für die Unterstützung bei der experimentellen Auswertung der Bildsequenz gilt mein besonderer Dank Dipl.-Ing. Werner Schneider. Durch seine umfangreichen Erfahrungen mit Standardsoftware der analytischen Photogrammetrie konnten die sehr spezifischen Probleme bei der Bündeltriangulation einer langen Bildsequenz erfolgreich gelöst werden.

Abschließend gilt mein Dank meiner Frau und meinen beiden Kindern, die mich auf ihre Weise unterstützt und ermutigt haben.

Lebenslauf

| | |
|-------------|---|
| 21.6.57 | geboren in Stuttgart |
| 1964 - 1976 | Schulbesuch in Merklingen, Weil der Stadt, Heimsheim und Stuttgart |
| 8.6.76 | Reifeprüfung |
| 1977 - 1978 | Grundwehrdienst |
| 1978 - 1983 | Studium des Vermessungswesens an der Universität Karlsruhe |
| 1984 | Vorbereitungsdienst für den höheren vermessungstechnischen Verwaltungsdienst in Baden-Württemberg |
| 1984 - 1986 | Wissenschaftlicher Mitarbeiter am Geodätischen Institut der Universität Karlsruhe |
| 1986 - 1990 | Assistent für Photogrammetrie am Institut für Photogrammetrie der Universität Stuttgart |
| seit 1990 | Akademischer Rat am vorgenannten Institut |
| seit 1994 | stellvertretender Institutsdirektor |