Rekursive Schätzung der relativen 3D-Bewegung einer Ebene aus längeren monokularen Bildfolgen*

Konstantinos Daniilidis
Institut für Informatik und Praktische Mathematik
Christian-Albrechts-Universität zu Kiel, Preusserstr. 1-9, 24105 Kiel
email: kd@informatik.uni-kiel.d400.de

Zusammenfassung

In diesem Beitrag wird die Fehlerempfindlichkeit der rekursiven Schätzer von 3D-Bewegung und Struktur aus längeren monokularen Bildfolgen untersucht. Ein neuer Algorithmus zur Schätzung der relativen Geschwindigkeit und der Normalen einer Ebene wird vorgeschlagen. Die Experimente aus synthetischen sowie auch Realweltbildfolgen zeigen, daß sich der Schätzfehler bei der Einführung eines Bewegungsmodells abschwächt, während die bekannte Empfindlichkeit an der Unterscheidung zwischen Translation und Rotation vorhanden bleibt.

1 Einführung

Zur Führung von autonom mobilen Fahrzeugen muß man aus bildgebenden Sensoren die Struktur der Umgebung und die relativen Bewegungen der Szenenkomponenten zur Kamera ermitteln. Die Durchführung dieser Aufgabe ist durch die Auswertung von monokularen Grauwertbildfolgen unter ganz allgemeinen Annahmen (wie Starrheit der Szenenkomponenten) im Prinzip möglich. Jedoch leidet eine Lösung dieses Problems unter Mehrdeutigkeiten. Existierende Algorithmen weisen eine hohe Empfindlichkeit gegen Meßrauschen auf, was deren Einsatz bei Realweltsituationen zur Zeit verbietet.

In [5, 4, 6] wurden analytische Nachweise für den Zusammenhang zwischen der Fehlerempfindlichkeit und der relativen Bewegung sowie der Geometrie der abgebildeten Szenenkomponenten bei der Berechnung der Bewegungsparameter aus einem Bildpaar gegeben. Die Benutzung eines Bildpaars dient als Fundament zum Verständnis der bei längeren Bildfolgen auftretenden Empfindlichkeiten und gibt direkt Leistungshinweise für die Ansätze, die eine längere Bildfolge als eine Summe von aufeinander folgenden Bildpaaren verarbeiten. In der Praxis einzusetzende Systeme müssen dreidimensionale Interpretationen während einer sich dauernd ändernden Umgebung anhand der ständig aufgenommenen Grauwertbilder ermitteln. Deshalb ist das eigentliche Ziel die Verwendung der gesamten Information, die sich während der Aufnahmezeit durch die andauernden relativen Bewegungen zwischen der Kamera und den Objekten der Umgebung erschließen läßt. Insbesondere ist der zwangsmäßig rekursive Charakter dieser Ermittlung von Bedeutung.

Die vorliegende Studie widmet sich der Ausnutzung der zeitlichen Kohärenz der Bewegung während einer längeren Bildfolge durch die Einführung von a priori Wissen über Bewegungsmodelle der Szenenkomponenten. Wir schlagen ein neues Verfahren zur rekursiven Berechnung der relativen Bewegung und der Neigung einer Ebene vor. Anhand dieses Verfahrens untersuchen wir, unter welchen Bewegungs- und Geometriekonfigurationen sich die Fehlerempfindlichkeit abschwächt oder unbeeinflußt bleibt. Die Untersuchung wird an synthetischen Experimenten sowie auch an einer Bildfolge durchgeführt, die von einer am Greifer eines Roboterarms befestigten Kamera aufgenommen wurde. Das vorgeschlagene rekursive Verfahren für Punktmerkmale, angewandt an mehreren Ebenen eines polyedrischen Objektes, in Kombination mit dem in [4]

*Diese Arbeit wurde während meiner Tätigkeit am Institut für Algorithmen und Kognitive Systeme an der Universität Karlsruhe (TH) unter finanzieller Unterstützung des Deutschen Akademischen Austauschdienstes (DAAD) durchgeführt. Mein Dank gilt meinem Betreuer Hans-Hellmut Nagel für die kritische Durchsicht der entsprechenden Teile meiner Dissertation. Ich bedanke mich bei Sami Atiya für unsere inspirationsvollen schätzungstheoretischen Diskussionen und bei Volker Gengenbach für die Hilfe bei der Durchführung der Experimente.

vorgestellten Bildpaarverfahren für geradlinige Merkmale kann die Basis für einen verallgemeinerten Algorithmus zur Schätzung der relativen Bewegung von polyedrischen Objekten bilden.

Der Leser sei für einen erschöpfenden Überblick über Verfahren zur Auswertung von längeren Grauwertbildfolgen auf [4] verwiesen. Die Ansätze unterscheiden sich nach der Art der verwendeten Bildbereichshinweise, die Annahmen über die Geometrie der Szene sowie die Bewegung und die verwendete Schätzmethode. Wir beschränken uns auf Ansätze zur Auswertung von *monokularen* Bildfolgen, die eine starre Bewegung unterstellen, und unterteilen die Algorithmen in vier Gruppen in Bezug auf die Annahmen und die verwendeten Sensoren. Die allgemeinste Problemstellung findet sich in der ersten Gruppe von Ansätzen, die eine beliebige Bewegung [8, 21, 3, 18, 10, 11, 1] oder eine Glattheit bzw. eine besondere Form der Bewegungstrajektorie [2, 15, 20, 19] unterstellen. Die zweite Gruppe besteht aus Ansätzen, die die Bewegungsinformation aus anderen Sensoren entnehmen (*Bewegungsstereo*), während die dritte Gruppe Modellinformation über die Szenenkomponenten einsetzt. Die vierte Gruppe beinhaltet Ansätze, die das Modell der orthographischen statt der perspektivischen Projektion verwenden. Aus Platzgründen wird leider auf die Literatur aus diesen Gruppen von Ansätzen, die mit dem hier vorgestellten Ansatz nicht direkt vergleichbar sind, verzichtet.

Bei der Auswertung einer längeren Bildfolge ist man gezwungen, die Schätzung der relativen Bewegung zwischen bildgebendem Sensor und einer Szenenkomponente *rekursiv* durchzuführen, d.h. zu jedem Zeitpunkt t_k den Zustandsschätzwert anhand nur der zu diesem Zeitpunkt aufgenommenen Messungen zu aktualisieren. Das Problem der Bewegungs- und Strukturschätzung leidet unter einer nichtlinearen Meßfunktion (perspektivische Abbildung) und gegebenfalls unter nichtlinearer Übertragungsfunktion. Dadurch ist man auf suboptimale Schätzer verwiesen: Die Mehrheit der obengenannten Ansätze verwenden den Erweiterten Kalman-Filter (EKF), dessen Leistung sehr stark von der Abweichung der Startwerte von den tatsächlichen Werten abhängt. Wenn die Abweichung groß ist, tritt eine Verzerrung auf, die sich nicht in der vom Filter berechneten Fehlerkovarianz erfassen läßt. Der EKF bildet nur den ersten Schritt des Iterierten Erweiterten Kalman-Filters (IEKF), der an einem bestimmten Zeitpunkt optimal im Sinne einer MAP-Schätzung ist, falls Meß- und Systemrauschen normalverteilt sind. Eine bessere nicht-iterative Approximation zum optimalen Filter ist der Modifizierte Gaußfilter zweiter Ordnung (MGSO)[12].

Das Ziel unserer Studie ist ein zweifaches: Erstens ist uns von Interesse, die Bewegungs- und Geometriekonstellationen zu ermitteln, bei denen die Einführung von a priori Wissen in Form eines Bewegungsmodells eine Abschwächung der Fehlerempfindlichkeit verursacht. Zweitens wollen wir die Eignung von jedem der obgenannten Schätzer untersuchen. Bei längeren Bildfolgen erlauben die Fülle der Bildbereichshinweise, die Anzahl der Unbekannten und die Einführung von a priori Wissen keine analytische Untersuchung wie beim Bildpaar, deshalb beschränken wir uns auf die funktionalen Zusammenhänge für den Verlauf des Schätzfehlers, die sich bei Realweltexperimenten und Simulationen beobachten lassen.

2 Relative Bewegung einer Ebene

Die Anzahl der Unbekannten, die der Struktur einer Szenenkomponente entprechen, ist proportional zur Anzahl der charakteristischen Szenenbereichshinweise (Vertizes, Kanten, markierte Punkte). Damit wir aber eine Einsicht in den Fehlerverlauf der Struktur bekommen und Rückschlüsse auf die Fehlerempfindlichkeitsergebnisse im Fall des Bildpaars [6] machen können, werden wir die grundlegende Annahme machen, daß wir nur die Menge der Szenenbereichshinweise untersuchen, die sich in einer Ebene befinden. So reduziert sich die Anzahl der Unbekannten auf zwei, den Polar- und Azimutwinkel der Normalen der Ebene. Der Abstand der Ebene vom Projektionszentrum wird wegen der Skalierungsmehrdeutigkeit im Betrag der Translation mitberücksichtigt.

Diese Annahme steht im Einklang mit aktuellen Anwendungen der Bewegungsschätzung in Realweltsituationen. Nehmen wir an, daß eine Kamera auf einem autonom geführten Fahrzeug befestigt ist, so ist die befahrbare Fläche eine Ebene. Der hier vorgestellte Algorithmus berechnet die momentane Translations- und Winkelgeschwindigkeit der Egobewegung in Bezug auf das Kamerakoordinatensystem, sowie auch die Normale der befahrbaren Ebene. Damit erspart man sich die Ermittlung der Transformation vom Straßenkoordinatensystem zum Kamerakoordinatensystem [7].

Eine zweite Anwendung ist die Auswertung von monokularen Bildfolgen aufgenommen von

einer Kamera, die am Greifer eines Roboterarms befestigt ist. Eine weit verwendete Annahme ist, daß die zu manipulierenden Objekte polyedrisch sind. Der hier vorgestellte Algorithmus ermittelt die Egobewegung in Bezug auf das Kamerakoordinatensystem und die Normale der berücksichtigten Ebene. Ist die Transformation vom Kamerakoordinatensystem zum Greiferkoordinatensystem (Hand-Auge-Kalibrierung) bekannt, so läßt sich die zum beabsichtigten Zugriff auf das Objekt benötigte Orientierung bezüglich des Roboterkoordinatensystems ermitteln. Eine dritte Anwendung ist die Ermittlung der Bewegung von Objekten, die man als polyedrische Strukturen modellieren kann.

Das eingeführte a priori Wissen betrifft die Glattheit der Bewegung. Wir nehmen an, daß die Translations- sowie auch die Winkelgeschwindigkeit in Bezug auf das Kamerakoordinatensystem konstant bis auf eine Unsicherheit bleiben, die sich mittels der Kovarianz des Systemrauschens darstellen läßt. Das trifft im Fall eines autonom geführten Fahrzeugs sowie auch eines Roboterarms zu, wenn die Bewegung zwischen zwei Aufnahmen gering ist. Die Annahme ist für den Fall der Verwendung von Verschiebungsraten im Bildbereich geeignet, weil die Ermittlung von Verschiebungsraten kleine Bewegungen im Abtastintervall unterstellen. Unser Ansatz ähnelt der Methode von [17], wobei aber auf die Eignung des rekursiven Schätzers (EKF) nicht eingegangen wird. Weiterhin vereinfachen wir die Meßgleichungen durch Einführung von Hilfsparametern und modellieren korrekt die Kollisionszeit.

Wir gehen auf die Beschreibung der vorgestellten Methode im Zustandsraum ein. Wir bezeichnen mit $\boldsymbol{v}$, $\boldsymbol{\omega}$ und $\boldsymbol{N}$ jeweils die Translations-, die Winkelgeschwindigkeit und die Einheitsnormale der Ebene in Bezug auf das Kamerakoordinatensystem. Der Abstand des Kamerazentrums von der Ebene ist d, so daß die Gleichung der Ebene im Kamerakoordinatensystem lautet $\boldsymbol{N}^T\boldsymbol{X} = d$. (siehe Abb. 1). Unter der Annahme von konstanten Geschwindigkeiten erhalten wir folgende Differentialgleichungen für den zeitlichen Verlauf des Systems:

$$\begin{array}{llll} \dot{\boldsymbol{v}} & = 0 & \dot{\boldsymbol{\omega}} & = 0 \\ \dot{\boldsymbol{N}} & = -\boldsymbol{\omega} \times \boldsymbol{N} & \dot{d} & = -\boldsymbol{v}^T\boldsymbol{N} \end{array} \tag{1}$$

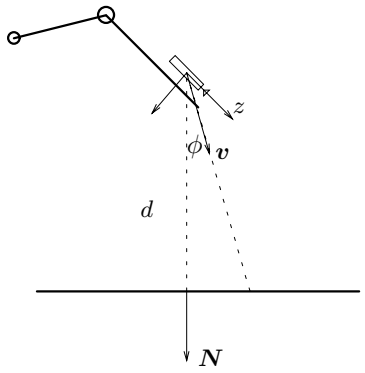

Abbildung 1: Bewegung einer an einem Greifer befestigten Kamera

Wir führen als unbekannten Vektor die durch den Abstand zur Ebene skalierte Translationsgeschwindigkeit $\boldsymbol{u} = \boldsymbol{v}/d$ ein. Der Betrag dieses Vektors wird von [17] als das Inverse der Kollisionszeit betrachtet. Wie man aber aus der Abb. 1 erkennen kann, stimmt das nicht, falls man die Kollisionszeit als die Zeit definiert, die bei konstanter Geschwindigkeit bis zum Treffen der Ebene verläuft. Nennt man den Winkel zwischen der Normalen und der Translationsgeschwindigkeit ϕ, läßt sich die Kollisionszeit als

$$\frac{1}{\rho_{ttc}} = \frac{d}{\|\boldsymbol{v}\| \cos\phi} = \frac{d}{\boldsymbol{v}^T\boldsymbol{N}} = \frac{1}{\boldsymbol{u}^T\boldsymbol{N}} \tag{2}$$

beschreiben. Wir werden das Inverse der Kollisionszeit $\rho_{ttc} = \boldsymbol{u}^T\boldsymbol{N}$ als Nebenprodukt aus unserem Algorithmus erhalten und es als Bewertungsmaß verwenden.

Wir definieren den Zustandsvektor des Systems als den neunelementigen Vektor

$$\boldsymbol{s} = \begin{pmatrix} \boldsymbol{u}^T & \boldsymbol{\omega}^T & \boldsymbol{N}^T \end{pmatrix}, \tag{3}$$

wobei $\boldsymbol{N}$ ein Einheitsvektor ist. Die entsprechende Zwangsbedingung $\|\boldsymbol{N}\| = 1$ wird in den Meßgleichungen berücksichtigt.

Zur Durchführung des Prädiktionsschrittes benötigen wir die Übergangsgleichung des Systems vom Zeitpunkt t_k auf den Zeitpunkt t_{k+1}, die wir aus den Differentialgleichungen (1) herleiten müssen. Die Übergangsgleichung für die Winkelgeschwindigkeit ist recht einfach:

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k. \tag{4}$$

Die Integration der Differentialgleichung für die Normale ist bekannt [14]. Die Lösung ist die Rotation der Normalen um eine Achse parallel zu $\boldsymbol{\omega}$ um den Winkel $-\|\boldsymbol{\omega}\|T$, wobei $T = t_{k+1} - t_k$:

$$\boldsymbol{N}_{k+1} = \boldsymbol{N}_k - \frac{\sin(\|\boldsymbol{\omega}_k\|T)}{\|\boldsymbol{\omega}_k\|}\boldsymbol{\omega}_k \times \boldsymbol{N}_k + \frac{1-\cos(\|\boldsymbol{\omega}_k\|T)}{\|\boldsymbol{\omega}_k\|^2}\boldsymbol{\omega}_k \times (\boldsymbol{\omega}_k \times \boldsymbol{N}_k). \tag{5}$$

Die Ermittlung der Übergangsgleichung für die skalierte Geschwindigkeit $\boldsymbol{u}_{k+1} = \frac{\boldsymbol{v}_{k+1}}{d_{k+1}}$ ist aufwendiger. Für den Übergang der skalierten Geschwindigkeit läßt sich beweisen [4], daß

$$\boldsymbol{u}_{k+1} = \frac{\boldsymbol{u}_k}{1 - \boldsymbol{u}_k^T\boldsymbol{N}_k T + \frac{\boldsymbol{u}_k^T(\boldsymbol{\omega}_k \times \boldsymbol{N}_k)\,(1-\cos(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^2} - \frac{(\boldsymbol{u}_k \times \boldsymbol{\omega}_k)^T(\boldsymbol{\omega}_k \times \boldsymbol{N}_k)\,(\|\boldsymbol{\omega}_k\|T - \sin(\|\boldsymbol{\omega}_k\|T))}{\|\boldsymbol{\omega}_k\|^3}}. \tag{6}$$

Zur Vervollständigung der Beschreibung des Systemübergangs benötigen wir die Jacobische Matrix der Übergangsfunktion, d.h. die Ableitung der Zustandsgrößen zum Zeitpunkt t_{k+1} nach den Zustandsgrößen zum Zeitpunkt t_k, um den Prädiktionsschritt bei einem rekursiven Schätzer durchzuführen. Die Berechnung dieser Jakobischen Matrix findet sich in [4]. Als Messung verwenden wir das Feld der Verschiebungsraten zum Zeitpunkt t_k. Aus der Annahme, daß es sich um die relative Bewegung einer Ebene handelt, lassen sich aus den Verschiebungsraten acht Hilfsparameter [16] herleiten.

$$\boldsymbol{q}_k = \begin{matrix} (-u_{kx}N_{kz} - \omega_{ky} & -u_{kx}N_{kx} + u_{kz}N_{kz} & -u_{kx}N_{ky} + \omega_{kz} & -u_{ky}N_{kz} + \omega_{kx} \\ -u_{ky}N_{kx} - \omega_{kz} & -u_{ky}N_{ky} + u_{kz}N_{kz} & -\omega_{ky} + u_{kz}N_{kx} & u_{kz}N_{ky} + \omega_{kx})^T \end{matrix} \tag{7}$$

Die acht Parameter können aus $m \geq 4$ Verschiebungsraten durch Lösung eines linearen Systems ermittelt werden, das aus folgender Minimierung folgt:

$$\sum_{i=1}^{m}(\dot{\vec{\boldsymbol{x}}}_{ki} - B_{ki}\boldsymbol{q}_k)^T C_{ki}^{-1}(\dot{\vec{\boldsymbol{x}}}_{ki} - B_{ki}\boldsymbol{q}_k) \Longrightarrow \min_{\boldsymbol{q}_k}. \tag{8}$$

Die Matrix C_{ki} ist die Kovarianz des Rauschens in den Verschiebungsraten, die abhängig vom Verfahren zur Ermittlung des optischen Flusses ist. Man kann durch die Einführung der Pseudoinversen statt der Inversen in (8) die Gewichtung nur der Projektion des optischen Flusses auf eine stabile Richtung erzwingen. Ist z.B. nur die Komponente entlang des Grauwertgradienten meßbar, so bekommen wir in (8) direkt die Minimierung der Quadrate dieser Komponenten. Die Ermittlung der Parameter $\boldsymbol{q}_k$ ist eine lineare Operation, daher bleibt der Fehler normalverteilt mit der Kovarianzmatrix $(\sum_{i=1}^{m} B_{ki}^T C_{ki}^{-1} B_{ki})^{-1}$. Zusätzlich zu den acht Parametern $\boldsymbol{q}_k$ wird als neunte Messung die Zwangsbedingung $\|\boldsymbol{N}_k\| = 1$ mit verschwindender Fehlerkovarianz berücksichtigt. Der Meßvektor zum Zeitpunkt t_k lautet dann

$$\boldsymbol{z}_k = \begin{pmatrix} \boldsymbol{q}_k \\ \|\boldsymbol{N}_k\|^2 - 1 \end{pmatrix}. \tag{9}$$

Wie man aus den Meßgleichungen erkennen kann, ist die Messung eine nichtlineare Funktion der Zustandsgrößen. Präziser gesagt sind die Messungen bilinear bezüglich der Translationsgeschwindigkeit und der Normalen und linear bezüglich der Winkelgeschwindigkeit. Zum Prädiktionsschritt der rekursiven Schätzung verwenden wir die Gleichungen (6), (4) und (5). Zum Aktualisierungsschritt werden wir drei Verfahren gegenüberstellen: den Erweiterten Kalman-Filter (EKF), den Modifizierten Gaußschen Filter zweiter Ordnung (MGSO) und den Iterierten Erweiterten Kalman-Filter (IEKF)[1]. Der rekursive Schätzer benötigt die Einstellung der Startwerte $\hat{\boldsymbol{s}}_0$ und der entsprechenden Kovarianz P_0^- sowie auch der Meß- und Prozeßrauschwerte.

[1] mit der Modifikation, daß die Minimierung nach dem Levenberg-Marquardt und nicht dem Gauß-Newton Schema abläuft (siehe auch [13]).

3 Experimentelle Untersuchung der Fehlerempfindlichkeit

Die hier zu auswertende monokulare Grauwertbildfolge besteht aus zwanzig Aufnahmen und wurde schrittweise von einer am Greifer eines Manipulators befestigten Kamera aufgenommen. Bei jedem Schritt hat der Manipulator eine Bewegung entprechend den sechs Stellgrößen unternommen, die von uns im voraus ausgerechnet wurden, so daß die erwünschte Bewegung (konstante Translations- und Winkelgeschwindigkeit) in Bezug auf das Kamerakoordinatensystem erreicht wird.

Die Stellgrößen in dem Experiment wurden so eingestellt, daß eine bezüglich des Kamerakoordinatensystems rein translatorische Bewegung mit $\boldsymbol{v} = (0, 0, 20)\,\mathrm{mm}$ erfolgt. Ob der Roboter tatsächlich die erwünschte Bewegung durchgeführt hat, hängt von zwei Faktoren ab: Erstens von der Unsicherheit in der relativen Rotation zwischen dem Kamerakoordinatensystem und dem Greiferkoordinatensystem und zweitens von der Präzision, mit der der Roboter die Befehle durchführt. Wegen der reinen Translation bleibt die Normale der Ebene bezüglich des Kamerakoordinatensystems zeitlich konstant. Die ungefähren Werte $(-0.2, 0.2, 0.95)$ der Normalenkomponenten wurden mittels Kalibrierung nach [22] mit Hilfe nur einer Ebene an der ersten Aufnahme ermittelt. Die Bezeichnung „ungefähr“ bezieht sich auf bekannt inhärente Ungenauigkeiten des verwendeten Verfahrens. Aus den Komponenten der Normalen erkennt man, daß es sich um eine – zumindest im Fall des Bildpaars – fehlerempfindliche Bewegungs- und Geometriekonfiguration handelt, weil der Winkel zwischen $\boldsymbol{N}$ und $\boldsymbol{v}$ niedrig ist.

Aus der aufgenommenen Bildfolge wurden Kantenelemente extrahiert, zu geradlinigen Kantensegmenten gruppiert und deren Schnittpunkte ermittelt. Diese Vorverarbeitungsschritte wurden automatisch nach der Aufnahme von dem Bildauswertesystem [9] durchgeführt. Das Grauwertbild und die detektierten Kantensegmente und Verschiebungsraten aus der Aufnahme zum Zeitpunkt t_{15} werden in der Abb. 2 gezeigt. Ziel unserer Untersuchung ist die Ermittlung der Fehlerverstärkung, die beim Auswertungsschritt der Bewegungs- und Strukturermittlung auftritt. Daher ist es wünschenswert, daß der Fehler in den Verschiebungsraten nur auf die Ungenauigkeit in der Ermittlung der Schnittpunkte und nicht auf falsche Zuordnungen zurückzuführen ist. Deshalb wurden die zeitlichen Zuordnungen der Schnittpunkte modellbasiert unter Verwendung der a priori Information über ihre Anordnung auf der Eichplatte ermittelt.

Abbildung 2: Aufnahmen zu Zeitpunkten t_0, t_5, t_{10} und t_{15} (links) und die extrahierten Geradensegmente (mitte) und Verschiebungsvektorfelder (rechts).

Der erste Eindruck aus Abb. 2 ist, daß es sich um eine rein translatorische Bewegung handelt, deren Expansionspunkt ungefähr die Bildpunktkoordinaten (260, 360) besitzt. Der Hauptpunkt hat die Koordinaten $(x_0, y_0) = (262, 267)$, daher besitzt die Translationsgeschwindigkeit eine vernachlässigbare x-Komponente. Unter Berücksichtigung des Skalierungsfaktors ($s_y = 1100$) kommt man auf einen Winkel zwischen dem $\boldsymbol{v}$-Vektor und der z-Achse von $\arctan(93/1100) \approx 5^\circ$ Grad. Die Diskrepanz zu dem eingestellten Wert, der einer reinen Translation in z-Richtung entspricht, ist auf die Ungenauigkeit der Hand-Auge-Kalibrierung zurückzuführen. Diese Tatsache muß man in den folgenden Diagrammen berücksichtigen, wo als tatsächlicher Wert die Translationsgeschwindigkeit in z-Richtung aufgezeichnet ist.

Für den Aktualisierungsschritt werden alle drei Schätzer (EKF, MGSO und IEKF) ausprobiert. Zusätzlich wird eine Standard-Methode zur Schätzung der Bewegung einer Ebene aus einem Bildpaar [16] angewendet, die wir mit „NI“ (nicht-inkrementell) bezeichnen, während die tatsächlichen Werte mit „GT“ (ground-truth) bezeichnet werden.

Aus den Schätzwerten des Azimut- und Polarwinkels der Translationsgeschwindigkeit in

Abb. 3 erkennen wir eine bei allen Filtern auftretende Verzerrung, die eine Wanderung des Expansionspunktes nach links und unten verursacht. Es ist zu bemerken, daß eine Verzerrung im Polarwinkel von 2° Grad eine Verschiebung des Expansionspunkts um ca. $\tan(2°) \cdot 100 \approx 38$ Bildpunkte verursacht. Die Erklärung für diese Verzerrung findet sich in den Verläufen der Schätzwerte für die Komponenten der Winkelgeschwindigkeit. Die größte Verzerrung – der Leser sei hier auf die Skalierung der Ordinate in den Diagrammen der Abb. 3 hingewiesen – mit ca. 0.002 rad/T tritt bei ω_x auf, was aus der analytischen Fehleruntersuchung in [6] zu erwarten ist: nur die Summe $-v_y + \omega_x$ kann robust berechnet werden. Entsprechend, aber niedriger ist der Beitrag von ω_y zur Kompensierung der v_x-Komponente. Allerdings haben wir erhofft, daß sich die Anfälligkeit für eine Verwechselung zwischen Rotation und Translation entlang einer längeren Bildfolge abschwächen würde. Immerhin weist aber jeder rekursive Schätzer ein stabileres Verhalten im Vergleich zum nicht-inkrementellen Verfahren auf. Bemerkenswert ist, daß die Schätzung der Normalen von den obgenannten Effekten nicht beeinträchtigt wird. Der Fehler im geschätzten Polarwinkel liegt zwischen 3° und 5° Grad, und das ist von Bedeutung für die Praxis, weil dieser Winkel die Anfahrrichtung eines Sauggreifers auf eine zu greifende Ebene angibt.

Zur Simulation von mehreren Bewegungs- und Geometriekonfigurationen verwenden wir ein Modell der im Experiment mit realen Bilddaten benutzten Eichplatte und die gleichen Hand-Auge und Roboter-Welt Transformationen sowie auch dieselben internen Kameraparameter. Wenn nicht darauf hingewiesen wird, wird ein gleichverteiltes Meßrauschen von ± 0.5 Bildpunkten angenommen und kein Prozeßrauschen verwendet. Die letzte Annahme wurde gemacht, damit wir vollständig den Einsatz des a priori Wissens über die Glattheit der Bewegung ausnutzen. Die Startwerte sind immer Null für die Geschwindigkeiten, und der Startwert für die Normale liegt in der YZ-Ebene und hat einen Neigungwinkel von 45° Grad. Es hat sich herausgestellt, daß das Hochsetzen auf sehr große Werte der Startkovarianz, wie es bei anderen Ansätzen vorgeschlagen wird, auf Divergenz in allen drei verwendeten Filtern führt. Außerdem besitzt man a priori Wissen über den Umfang der Geschwindigkeiten: Die durch den Abstand skalierte Geschwindigkeit $\boldsymbol{u}$ und die Winkelgeschwindigkeit $\boldsymbol{\omega}$ können nicht beliebig groß werden, weil die Länge der Verschiebungsrate mitwächst. Angesichts der Tatsache, daß bei dem vorhandenen Skalierungsfaktor eine zur Bildebene parallele Winkelgeschwindigkeit von 0.01 [rad/T] eine Verschiebungsrate von 10 Bildpunkten verursacht – entsprechendes gilt für die Translation –, beschränken wir die Startkovarianz für die skalierte Translationsgeschwindigkeit $\boldsymbol{u}$ auf $10^{-4}\,[1/T]^2$ und für die Winkelgeschwindigkeit $\boldsymbol{\omega}$ auf $10^{-5}\,[\mathrm{rad}/T]^2$.

Beim ersten Experiment variieren wir die Richtung der Translationsgeschwindigkeit. Wir stellen die Größe des effektiven Gesichtsfelds auf 6x6 Quadrate der Abb. 2. Die optische Achse bildet mit der Ebenennormalen einen Winkel von 30° Grad. Wir zeichnen dieselben Fehler wie oben für fünf verschiedene Bewegungen m_i mit entsprechenden Translationsgeschwindigkeiten $(0, i, 6)$mm (Abb. 4) auf. Der Vektor der Translationsgeschwindigkeit wandert in der YZ-Ebene von einer zur optischen Achse parallelen Lage zu einer zur Ebenennormalen parallelen Lage. Der Zusammenhang zwischen dem Fehler in Translationsrichtung und dem Fehler in ω_x ist eindeutig zu erkennen: Beide Fehler wachsen mit zunehmender Abweichung der Translationsrichtung von der optischen Achse. Weitere Experimente werden in [4] präsentiert.

4 Schlußüberlegungen

Der Verlauf der Schätzfehler bei der Realweltbildfolge zeigt, daß unter den drei rekursiven Schätzer der IEKF und der MGSO-Filter niedrigere Fehler aufweisen. Wir sollen hier darauf hinweisen, daß hier zur Erstellung eines Vergleichs ein Fall ausgewählt wurde, wobei der EKF nicht divergiert, was mehrmals der Fall ist. Aus den synthetischen Experimenten schließen wir, daß der IEKF ein stabiles Verhalten aufweist, das aber immer noch eine Verzerrung enthält, die von der Kopplung der Translation und der Rotation und von der relativen Lage der Ebenennormalen und der Translationsgeschwindigkeit abhängt. Wegen dieser Verzerrung, die dem Problem inhärent ist, müssen zusätzliche Bildinformationen eingeführt und /oder komplexere Bewegungen modelliert werden. Schon die Tatsache, daß die Dualität der Lösung bei konstanter Richtung der Translationsgeschwindigkeit erhalten bleibt, weist darauf hin, daß bei einem Bewegungsmodell mit variierender Translationsrichtung oder bei Einführung der Stellgrößen des Roboters in den Prädiktionsschritt die vom Winkel zwischen Normalen und Translation abhängige Fehlerempfindlichkeit abgeschwächt werden kann. Unsere Fehlerempfindlichkeitsuntersuchung kann

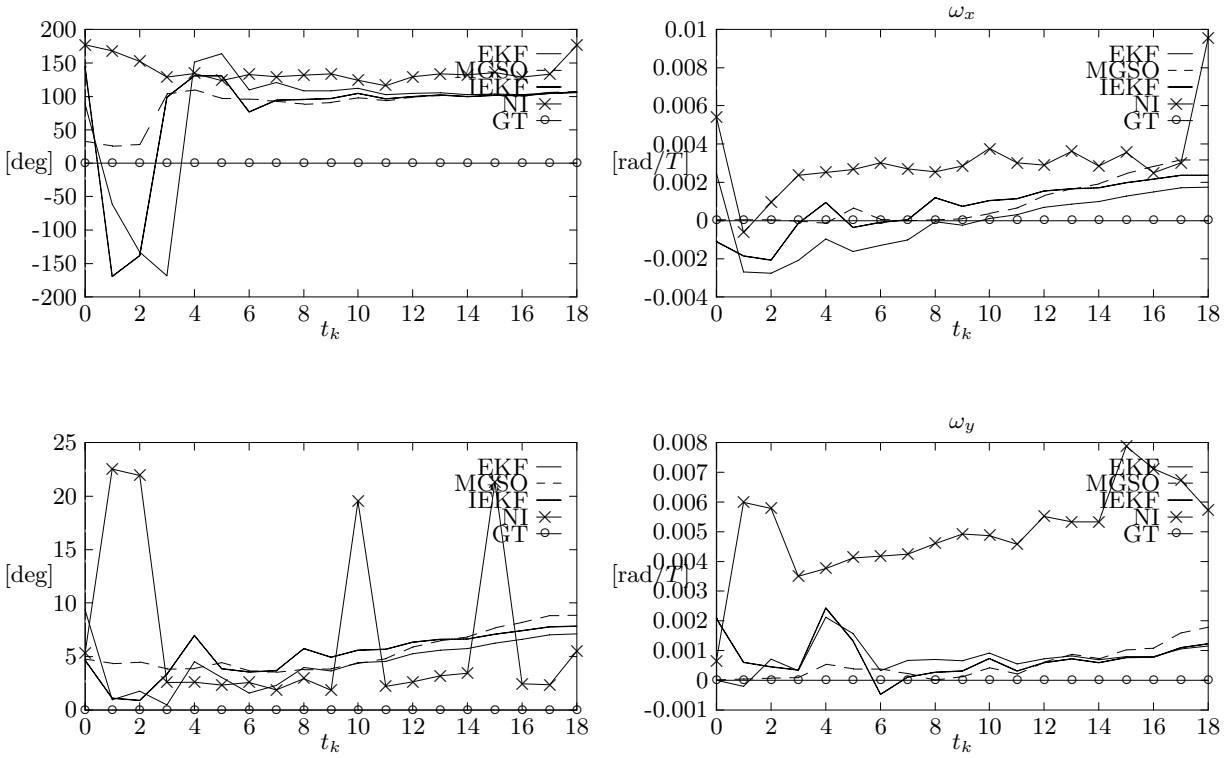

Abbildung 3: Schätzwerte für den Azimutwinkel (links oben), den Polarwinkel (links unten) der Translationsgeschwindigkeit, die x- und y-Komponente (rechts oben bzw. unten) der Winkelgeschwindigkeit.

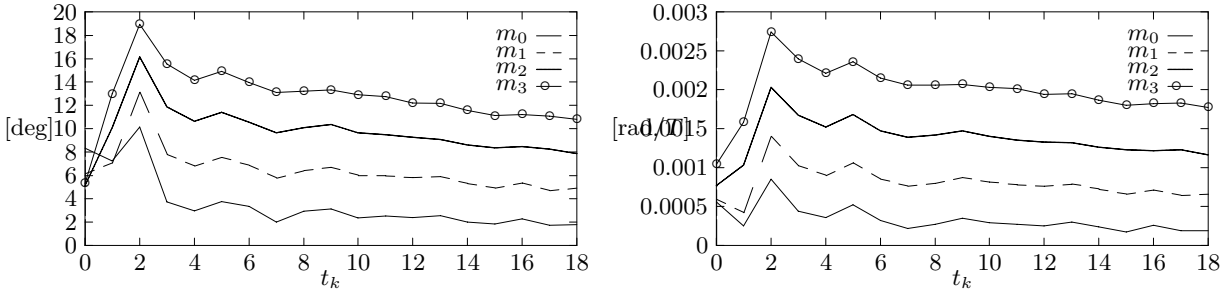

Abbildung 4: Fehlerwinkel in der vom IEKF geschätzten Translationsgeschwindigkeit (links) und absoluter Fehler in ω_x (rechts) bei variierender Translationsrichtung

bei dem Entwurf von aktiven Verfahren zur Berechnung von relativen Bewegungen verwendet werden. Es bleibt als offene Frage nachzuweisen, daß aktives Bewegungssehen (wie z.Bsp. durch Verfolgungsbewegungen des Auges) Instabilitäten und Mehrdeutigkeiten bei der Auswertung von monokularen Bildfolgen aufhebt.

Literatur

[1] H. Ando. Dynamic reconstruction of 3D structure and 3D motion. In *Proc. IEEE Workshop on Visual Motion*, pp. 101–110, Princeton, NJ, Oct. 7-9, 1991.

[2] T. Broida and R. Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13:497–513, 1991.

[3] N. Cui, J. Weng, and P. Cohen. Extended structure and motion analysis from monocular image sequences. In *Proc. Int. Conf. on Computer Vision*, pp. 222–229, Osaka, Japan, Dec. 4-7, 1990.

[4] K. Daniilidis. Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und

relativen Bewegungen aus monokularen Bildfolgen. Dissertation, Fakultät für Informatik, Universität Karlsruhe (TH), Juli 1992.

[5] K. Daniilidis and H.-H. Nagel. Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing*, 8:297–303, 1990.

[6] K. Daniilidis and H.-H. Nagel. The coupling of rotation and translation in motion estimation of planar surfaces. In *IEEE Conf. on Computer Vision and Pattern Recognition 1993*, pp. 188–193, New York, NY, June 15-17, 1993.

[7] E.D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241–261, 1988.

[8] L. Dreschler and H.-H. Nagel. Volumetric Model and 3D Trajectory of a Moving Car Derived from Monocular TV Frame Sequences of a Street Scene. *Computer Graphics and Image Processing*, 20:199–228, 1982.

[9] V. Gengenbach. Automatischer Zugriff eines Roboters auf ungeordnete Werkstücke mit Hilfe einer 3D-Lagebestimmung durch ein Mehrkamerasystem. Diplomarbeit, Fakultät für Informatik der Universität Karlsruhe, Januar 1990.

[10] C.G. Harris and J.M. Pike. 3D positional integration from image sequences. *Image and Vision Computing*, 6:87–90, 1988.

[11] J. Heel. Dynamic motion vision. *Robotics and Autonomous Systems*, 6:297–314, 1990.

[12] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, New York, NY and London, UK, 1970.

[13] D. Koller, K. Daniilidis, T. Thorhallsson, and H.-H. Nagel. Model-based object tracking in traffic scenes. In *Proc. Second European Conference on Computer Vision*, pp. 437–452, Santa Margerita, Italy, May 23-26, G. Sandini (Ed.), Lecture Notes in Computer Science 588, Springer-Verlag, Berlin et al., 1992.

[14] G.A. Korn and T.M. Korn. *Mathematical Handbook for Scientists and Engineers.* McGraw-Hill, New York, 1968.

[15] R.V. Raja Kumar, A. Tirumalai, and R. C. Jain. A non-linear optimization algorithm for the estimation of structure and motion parameters. In *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 136–143, San Diego, CA, June 4-8, 1989.

[16] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proc. Royal Society of London*, B208:385–397, 1980.

[17] D.W. Murray and D.M. Pickup. Recursive updating of planar motion. In *Proc. British Machine Vision Conference*, pp. 169–177, Glasgow, UK, Sept. 24-26, 1991.

[18] J. Oliensis and J.I. Thomas. Incorporating motion error in multi-frame structure from motion. In *Proc. IEEE Workshop on Visual Motion*, pp. 8–13, Princeton, NJ, Oct. 7-9, 1991.

[19] H.S. Sawhney, J. Oliensis, and A.R. Hanson. Description and reconstruction from image trajectories of rotational motion. In *Proc. Int. Conf. on Computer Vision*, pp. 494–498, Osaka, Japan, Dec. 4-7, 1990.

[20] H. Shariat and K.E. Price. Motion estimation with more then two frames. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-12:417–434, 1990.

[21] M. Spetsakis and J. Aloimonos. A multi-frame approach to visual motion perception. *International Journal of Computer Vision*, 6:245–255, 1991.

[22] R. Tsai. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Trans. Robotics and Automation*, 3:323–344, 1987.