

RAEUMLICHE ANALYSE VON BILDSEQUENZEN
MITHILFE KORRESPONDIERENDER KANTEN

Bernd Neumann

Fachbereich Informatik, Universitaet Hamburg
Schlueterstr.70, D-2000 Hamburg 13

ZUSAMMENFASSUNG

Ueber eine Bildsequenz mit bewegten Objekten koennen Aussagen gemacht werden, ohne dass spezielles Vorwissen ueber Form und Bedeutung einzelner Objekte herangezogen wird. Dieser Beitrag geht davon aus, dass die in den Bildern sichtbaren Projektionen von Objektkanten durch Geraden angenaehert sind, und korrespondierende Geraden aufeinanderfolgender Bilder ermittelt werden koennen. Es wird gezeigt, inwieweit man von den Lage- und Richtungsaenderungen einer Geraden auf ihren raeumlichen Verlauf schliessen kann. Ein Analyseverfahren wird vorgestellt, das die Geraden in Objektkandidaten gruppiert und dabei sowohl ihren Bewegungsablauf als auch ihre raeumliche Lage ermittelt.

Beitrag zum 2. DAGM-Symposium
"Angewandte Szenenanalyse und Mustererkennung"
Karlsruhe 10.-12. Oktober 1979

RAEUMLICHE ANALYSE VON BILDSEQUENZEN MITHILFE KORRESPONDIERENDER KANTEN

Bernd Neumann
Fachbereich Informatik, Universitaet Hamburg

1. EINLEITUNG

Bei der Analyse einer Bildfolge, die die zeitliche Entwicklung einer natuerlichen Szene zeigt, spielen bewegte Objekte eine besondere Rolle. Sie koennen z.B. unter bestimmten Umstaenden durch Auswerten von Bildveraenderungen vom unbewegten Hintergrund getrennt und als isolierte Objekte verfolgt werden, ohne dass Vorwissen ueber moegliche Objektformen und -eigenschaften herangezogen werden muss [3,7]. Darueberhinaus lassen sich Aussagen ueber ihre 3D-Gestalt ableiten, wenn die bewegten Objekte starre Koerper darstellen und verschiedene Ansichten praesentieren. Das Vorgehen ist analog zur stereoskopischen Analyse einer statischen Szene - in einem Fall werden verschiedene Ansichten durch Positionsveraenderung des Koerpers, im anderen Fall durch solche des Betrachters erzeugt. Allerdings ergeben sich bei der 3D-Analyse bewegter Szenen noch zusaetzliche Probleme, z.B. dadurch, dass die Trajektorien in der Regel unbekannt oder mehrere Koerper gleichzeitig in Bewegung sind.

Ullman hat in seinem "Structure from Motion Theorem" [6] die prinzipiellen Moeglichkeiten aufgezeigt: Aus 3 verschiedenen Ansichten von 4 nicht-koplanaren Punkten eines starren Koerpers lassen sich die relativen 3D-Positionen dieser Punkte errechnen. Ein anwendungsnaeherer Ansatz wurde von Bonde und Nagel [2] verfolgt. Sie betrachten Szenen der natuerlichen Umwelt, bei denen Objekte ebene Bewegungen ausfuehren, z.B. Fussgaenger und Fahrzeuge im Strassenverkehr. Unter der Voraussetzung, dass markante Punkte einzelner Koerper hinreichend genau verfolgt werden koennen, bietet ihr Verfahren die Moeglichkeit, beliebig viele Ansichten in den 3D-Analyseprozess einzubeziehen. Die genannte Voraussetzung ist nichttrivial, sie ist eine Variante des bisher erst fuer eingeschaenkte Szenenvorlagen befriedigend geloesten "Korrespondenzproblems" [1,4].

Der hier vorgestellte Ansatz weicht von den bisher bekannten Verfahren zur 3D-Analyse entscheidend darin ab, dass nicht die Lageaenderung von Punkten sondern Lage- und Richtungaenderung von geraden Kantenstuecken (im folgenden kurz "Kanten" genannt) ausgewertet wird. Kanten haben den Vorteil, dass die fuer eine 3D-Analyse wesentliche Information aus einer groesseren Bildumgebung abgeleitet wird. Zudem koennen mehr deskriptive Merkmale zur Loesung

des Korrespondenzproblems herangezogen werden; dieses wird in diesem Beitrag jedoch ausgeklammert. Vorgegeben seien (projizierte) Lage und Richtung von Kanten als Funktion des Zeitindex einer Bildsequenz. Wie kann man

- A) die Kanten in unabhaengig bewegte Objekte gruppieren (inkl. Hintergrund),
- B) die raeumliche Bewegung eines Objektes ermitteln,
- C) den raeumlichen Kantenverlauf bestimmen ?

In Abschnitt 2 werden zunaechst kinematische Gleichungen fuer eine einzelne Kante angegeben. Sie setzen die unbekanntes raeumlichen Winkel- und Bewegungskomponenten mit den vorgegebenen Groessen in Beziehung. Mit diesen Bewegungsgleichungen kann geprueft werden, ob sich mehrere Kanten starr bewegen. Es ergibt sich die Moeglichkeit, die drei Aufgaben A, B und C mit einem einzigen Verfahren zu loesen. Dieses wird in Abschnitt 3 vorgestellt. Abschnitt 4 schliesslich enthaelt Angaben ueber die Implementation.

2. KINEMATIK EINER KANTE

Fuer die folgende Analyse wird angenommen, dass (i) Kanten fest mit der Oberflaeche von starren Objekten verbunden sind (also z.B. keine Schatten sind) und (ii) die Bilder von einer festen Kamera in annaeherd orthographischer Projektion aufgenommen werden. Die raeumliche Bewegung eines Objektes kann durch 3 Translationskomponenten t_x , t_y , t_z und 3 Rotationskomponenten φ_x , φ_y , φ_z in einem festen Koordinatensystem beschrieben werden. Die xy -Ebene wird parallel zur Bildebene gewaehlt. Wegen orthographischer Projektion koennen Verschiebungen in z -Richtung nicht entdeckt werden. Der Nullpunkt der z -Achse ist also irrelevant, ebenso die t_z -Komponente, die fortan als Null angenommen wird.

Eine Kante wird als Gerade beschrieben ohne Festlegung von Anfangs- und Endpunkt. Vier Groessen definieren ihren raeumlichen Verlauf (s. Abb. 1):

d, φ	Polarkoordinaten der projizierten Geraden
θ	Winkel zwischen xy -Ebene und Geraden
ψ	Winkel zwischen xy -Ebene und Ursprungsplot auf Gerade

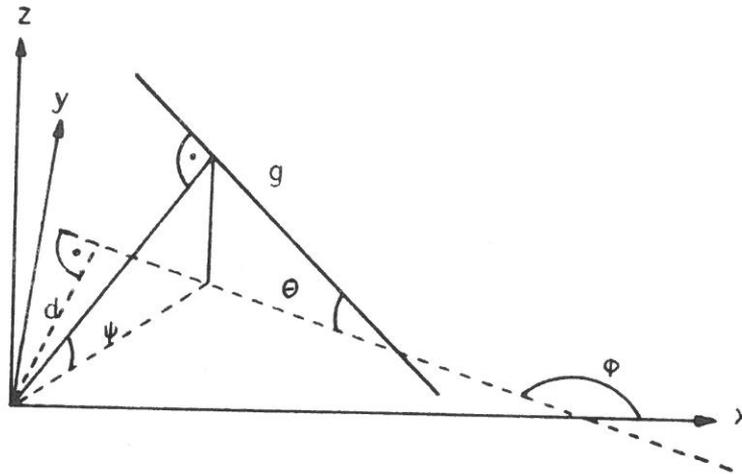


Abb. 1: Eine Gerade g beschrieben durch d , φ , θ und ψ .

d und φ koennen dem Bild unmittelbar entnommen werden, θ und ψ dagegen beschreiben den unbekanntem raeumlichen Verlauf. Seien 1 und 2 Indizes, die die Zeitpunkte vor und nach einer Rotation kennzeichnen. Dann gelten folgende kinematische Gleichungen:

$$\tan \theta_1 =$$

$$\frac{\sin \varphi_1 \cos \rho_x \cot(\varphi_2 - \rho_z) - \cos \varphi_1 \cos \rho_y - \sin \varphi_1 \sin \rho_x \sin \rho_y}{\sin \rho_x \cot(\varphi_2 - \rho_z) + \cos \rho_x \sin \rho_y} \quad (1)$$

$$\sin \theta_2 = \sin \theta_1 \cos \rho_x \cos \rho_y + \cos \theta_1 \sin \varphi_1 \sin \rho_x \cos \rho_y - \cos \theta_1 \cos \varphi_1 \sin \rho_y \quad (2)$$

$$\begin{aligned} \sin \psi_2 = \sin \psi_1 * (\cos \theta_1 \cos \rho_x \cos \rho_y + \sin \theta_1 \sin \varphi_1 \sin \rho_x \cos \rho_y \\ - \sin \theta_1 \cos \varphi_1 \sin \rho_y) \\ + \cos \psi_1 * (\cos \theta_1 \sin \rho_x \cos \rho_y - \sin \varphi_1 \sin \rho_y) \end{aligned} \quad (3)$$

$$1 - \frac{\sin^2 \psi_2}{\cos^2 \theta_2} = \left(\frac{d_2}{d_1}\right)^2 * \left(1 - \frac{\sin^2 \psi_1}{\cos^2 \theta_1}\right) \quad (4)$$

Aus (3) und (4) koennen Bestimmungsgleichungen fuer 1 und 2 analog zu (1) und (2) abgeleitet werden, so dass insgesamt folgende Funktionen berechnet werden koennen:

$$\Theta_1 = \text{THETA1}(\rho_x, \rho_y, \rho_z, \varphi_1, \varphi_2) \quad (5)$$

$$\Theta_2 = \text{THETA2}(\rho_x, \rho_y, \Theta_1, \varphi_1) \quad (6)$$

$$\Psi_1 = \text{PSI1}(\rho_x, \rho_y, \Theta_1, \varphi_1, d_1, d_2) \quad (7)$$

$$\Psi_2 = \text{PSI2}(\rho_x, \rho_y, \Theta_1, \varphi_1, \Psi_1) \quad (8)$$

Θ und Ψ einer einzelnen Kante koennen natuerlich nicht bestimmt werden, auch nicht aus beliebig vielen Ansichten, weil mit jeder neuen Ansicht unbekannte Rotationswinkel verknuepft sind. Erst durch Analyse mehrerer gleichbewegter Kanten ergeben sich genuegend viele Bestimmungsgleichungen.

3. DAS ANALYSEVERFAHREN

Das Verfahren geht von einer Menge uninterpretierter Kanten aus, d.h. weder ihr raeumlicher Verlauf noch ihre Zugehoerigkeit zu einzelnen Objekten ist bekannt. Jede Kante ist entsprechend ihrer 3D-Interpretierbarkeit, d.h. ihrer moeglichen Θ - Ψ -Zuweisungen, markiert, anfangs also durch Marken fuer alle moeglichen Θ - Ψ -Kombinationen. Ausgehend von Bild 1 einer Sequenz, werden nun Bewegungskomponenten hypothetisiert, die die Kanten in die korrespondierenden Kanten des naechsten Bildes transformieren, zunaechst unter der Annahme, dass sich alle betrachteten Kanten gemeinsam als starrer Verband bewegen. Der translatorische Anteil kann als Verschiebung eines heuristisch gewaehlten koerperfesten Punktes abgespalten werden. Die Rotationshypothese erlaubt es, aus den moeglichen Θ - Ψ -Kombinationen diejenigen herauszufiltern, die mit den beobachteten d - und φ -Werten kompatibel sind. Nach dem ersten Schritt bleibt fuer jede Kante theoretisch nur noch ein Θ - Ψ -Paar uebrig (Anwendung von THETA1 und PSI1); eine realistische Implementation wird jedoch Toleranzen beruecksichtigen muessen (s. Abschnitt 4), so dass im allgemeinen eine reduzierte Menge von Θ - Ψ -Kombinationen zu erwarten ist. Durch THETA2 und PSI2 werden die verbleibenden Marken entsprechend der Bewegungshypothese transformiert, so dass sie den moeglichen raeumlichen Verlauf in Bild 2 beschreiben. Daraufhin kann der naechste Schritt analog zum ersten erfolgen.

Kanten, fuer die die Bewegungshypothesen falsch sind, werden nach wenigen Schritten keine moegliche Markierung mehr besitzen und scheiden aus. Sinkt die Zahl der verbleibenden Kanten unter ein Minimum (etwa 2 oder 3), so werden nach den Regeln einer Tiefensuche alternative Bewegungshypothesen erkundet. "Richtige" Hypothesen lassen sich bis zum Ende der Sequenz fortschreiben. Die zugehoerigen Kanten stellen einen Objektkandidaten dar. Nach jedem erfolgreich gefundenen Objektkandidaten wird das

Verfahren fuer die verbleibenden Kanten neu aufgesetzt, bis keine sinnvoll interpretierbaren Kanten uebrig sind.

Das Konzept der Tiefensuche wurde gewaehlt, um einfache Verhaeltnisse schnell analysieren zu koennen. Man kann z.B. gleich am Anfang alle unbewegten Kanten durch Verfolgen einer Nullhypothese abtrennen. Bei gleichmaessig bewegten Objekten kann der Suchraum entsprechend durch Verfolgen einer Hypothese konstanter Bewegung klein gehalten werden. Die Filtertechnik ist dem "relaxation labelling" verwandt, wie es z.B. von Rosenfeld [5] fuer Bereichsinterpretationen vorgeschlagen wurde. Die Problemstruktur ist hier jedoch insofern komplizierter, als die Markierungen durch jeden Filterschritt transformiert werden.

4. IMPLEMENTATION

Um den Rechenaufwand klein zu halten und um das Verfahren an ungenaue Ausgangsdaten anzupassen, wurden die 4 Funktionen Gl. 5-8 diskretisiert und tabellarisch implementiert. Dies sei am Beispiel von THETA1 ausgefuehrt. Jeder diskrete Wert repraesentiert ein Intervall, die 5 Argumente insgesamt also einen Hyperwuerfel im Argumenteraum. Als Ergebnis liefert THETA1 die Menge aller diskreten Thetawerte, in deren Intervalle der Hyperwuerfel abgebildet wird. Werden Winkel in 5-Grad-Schritten diskretisiert, ergeben nur ca. 10% aller Hyperwuerfel Thetamengen mit mehr als 3 Elementen (THETA1 ist relativ "glatt"). Dementsprechend ist in der implementierten Tabelle je Eintragung Platz fuer 3 Elemente. Ausnahmen werden als undefiniert markiert.

Der Speicherbedarf fuer alle 4 Tabellen betraegt (unter Ausnutzung von Symmetrien) ca. 10 MByte; hier muss also peripherer Speicher benutzt werden. Durch eine geeignete Organisation in Seiten und einen angepassten Seitenaustausch-Algorithmus koennen dennoch schnelle mittlere Zugriffszeiten erwartet werden. Genauere Ergebnisse und letztlich auch ein Erfolgsnachweis fuer das vorgeschlagene Verfahren stehen noch aus, da derzeit erst einfache Testdatensaetze analysiert worden sind.

LITERATURVERZEICHNIS

- [1] T.Banard, W.Thompson: "Disparity Estimation Using Feature Point Matching", WCATI, Philadelphia, 1979
- [2] T.Bonde, H.H.Nagel: "Deriving a 3D-Description of a Moving Rigid Object from Monocular TV-Frame Sequences", WCATI, Philadelphia, 1979
- [3] R.Jain, H.H.Nagel: "On a Motion Analysis Process for Image Sequences from Real World Scenes", IEEE Comp. Soc. Conf. PRIP, Chicago, 1978
- [4] R.Kraasch et al.: "Automatische dreidimensionale Beschreibung bewegter Gegenstaende", dieser Band, 1979
- [5] A.Rosenfeld et al.: "Scene Labelling by Relaxation Operations", IEEE Trans. Sys. Man Cyb. SMC-6, 1976
- [6] S.Ullman: "The Interpretation of Visual Motion", MIT Press, Cambridge, Mass., 1979
- [7] M.Yachida et al.: "Automatic Motion Analysis System of Moving Objects from the Records of Natural Processes", IJCPR-78, Kyoto, 1978