Integration virtueller Objekte in natürliche Bildsequenzen

von

Dipl.-Math. Kai Cordes

Prof. Dr.-Ing. Jörn Ostermann

Institut für Informationsverarbeitung (TNT),

Leibniz Universität Hannover

Inhaltsangabe

Für die Integration virtueller Objekte in eine mit einer Videokamera aufgenommenen Bildsequenz werden exakte Informationen über die Bewegungsparameter der verwendeten Kamera benötigt. Nur so können nachträglich mit einem entsprechenden 3D-Bearbeitungsprogramm entworfene, virtuelle Objekte überzeugend in die Szene eingefügt werden. Es wird gezeigt, dass die Informationen über die Bewegung der Kamera vollautomatisch aus den rein photometrischen Eigenschaften der Bilder der Sequenz gewonnen werden können. Der Algorithmus gliedert sich in drei voneinander unabhängige Teile, die den momentanen Stand der Forschung repräsentieren und an denen auch kontinuierlich weiter Forschung betrieben wird. Dabei ist die Integration von künstlich erstellten Objekten oft ein erstes, leicht handhabbares Maß für die Genauigkeit, die mit den entsprechenden Algorithmen erzielt werden konnte. Zu große Fehler bei der Berechnung machen sich unter anderem durch ein Zittern oder nicht konsistentes Bewegen der Objekte bemerkbar.

Die gelungene, vollautomatische Schätzung einer „Freihand"-Szene wird anhand des sich ergebenden Kamerapfades, durch 3D-Punkte, die die Szene wiedergeben und anhand der gelungenen Integration eines Objektes, das in die Szene eingefügt wurde, belegt

Inhaltsverzeichnis

1 Einleitung

Die Integration virtueller Objekte in reale Bildsequenzen gewinnt insbesondere für die TV- und Filmproduktion zunehmend an Bedeutung. So bestehen heutzutage Kinofilme zu einem großen Anteil aus virtuellen Objekten, die zu einer real aufgenommenen Sequenz hinzugefügt werden. Aufgrund der weit fortgeschrittenen Möglichkeiten detailliert Objekte realitätsgetreu darzustellen, werden zunehmend auch real wirkende Teile durch am Computer generierte Modelle ersetzt, um die Aufnahmekonditionen zu vereinfachen. Durch 3D-Modelling-Programme werden die Objekte aus Primitiven wie Kugeln, Quadern, Nurbs, etc. erstellt, mit einer Oberflächenstruktur (*Textur*) versehen und zu der real aufgenommenen Bildsequenz hinzugefügt (*Compositing*). Für jedes Bild einer Sequenz wird eine virtuelle Kamera positioniert, um dessen Bewegung bei der abschließenden Berechnung des Ergebnisses (*Rendering*) zu berücksichtigen.

Damit sich bei einer realen Bildsequenz virtuelle Objekte überzeugend und nahtlos in die Aufnahmen integrieren lassen, müssen alle Bewegungsparameter der für die Aufnahme benutzten Kamera bekannt sein. Diese Informationen können während der Aufnahme durch ein *Motion-Control*-System gespeichert oder, um den Einsatz dieser teuren und unflexiblen Systeme zu umgehen, anschließend durch den Einsatz eines Algorithmus geschätzt werden (*Matchmoving*-Software). Als Eingangsmaterial für die Schätzung stehen dem Algorithmus die Bildsequenz und die Bauart der verwendeten Kamera, durch so genannte innere Kameraparameter definiert, zur Verfügung.

Kommerzielle Programme zur Kameraschätzung und zur 3D-Modellierung sind z. B. Boujou (www.-2d3.com) oder MatchMover (www.realviz.com) bzw. 3D Studio Max (www.discreet.com/products/-3dsmax/) oder Softimage 3D (www.softimage.com). Frei verfügbar sind zur Kameraschätzung der Voodoocameratracker(www.digilab.uni-hannover.de) und zur 3D-Modellierung Blender (www.blender.org).

Die in diesem Artikel vorgestellten Algorithmen sind Teile der Implementation des Voodoo camera tracker.

2 Algorithmus zum Schätzen der Kameraparameter

Der hier vorgestellte Algorithmus zum automatischen Schätzen der Kamerabewegung aus einer Bildfolge gliedert sich in drei Teile: Merkmalsdetektion, Korrespondenzanalyse und die robuste Schätzung der Kameraparameter. Als Grundlage für die Berechnungen dienen die Merkmalspunkte, die in aufeinander folgenden Bildern der Sequenz als korrespondierend erkannt werden müssen. Mit Hilfe der Korrespondenzen werden die Kameraparameter bestimmt. Da sich auch fehlerhafte Korrespondenzen ergeben, muss vorher eine Unterteilung in nutzbare (*inlier*) und nicht nutzbare (*outlier*) Korrespondenzen stattfinden.

2.1 Detektion von Merkmalspunkten

In den Kamerabildern werden diejenigen Punkte als Merkmalspunkte ausgewählt, die aufgrund ihrer Texturumgebung die größte Wahrscheinlichkeit dafür bieten, dass die entsprechenden Punkte in den folgenden Bildern wieder gefunden werden können. Diese Eigenschaft ist dann gegeben, wenn an der Position des Merkmalspunktes hohe Bildsignalgradienten in zwei senkrechte Richtungen vorliegen [1].

Durch die Auswertung von Gradienten wird das Verfahren invariant gegenüber globalen Beleuchtungsänderungen von Bild zu Bild. Da als Argument für das Auftreten eines Merkmalspunktes die Eigenwerte einer Gradientenmatrix genutzt wird, ergibt sich eine Invarianz gegenüber Drehungen des Bildinhaltes.

In Bild 1 ist ein Ergebnis des Harris-Detektors zu sehen. Die weißen Kreuze markieren die ausgewählten Merkmalspunkte, die sich ausschließlich an Ecken und Kanten befinden. In homogenen Bereichen des Bildes finden sich keine Merkmalspunkte, da es unwahrscheinlich ist, diese im nächsten Bild der Sequenz wieder zu finden.

Bild 1 Detektierte Merkmalspunkte mit dem Harris Detector.

Aktuellere Ansätze zur Merkmalspunktdetektion [2] nutzen Auflösungspyramiden, um die Punkte auch in unterschiedlichen Skalen und Oktaven des Eingangsbildes zu detektieren. Dann können diese auch in den Mittelpunkten homogener Regionen liegen.

2.2 Korrespondenzanalyse

Um die Korrespondenz zwischen Merkmalspunkten in aufeinander folgenden Bildern zu bestimmen, wird eine symmetrisch um den Punkt liegende Nachbarschaftsumgebung verwendet. Nun wird zu einer Merkmalspunktnachbarschaft im Bild *i* die ähnlichste Merkmalspunktnachbarschaft in Bild *i+1* bestimmt, um eine Korrespondenz zu erhalten. Die Aneinanderkettung von Korrespondenzen ergibt eine *Trajektorie* (siehe Bild 2). Im Falle einer mit einer handelsüblichen Videokamera aufgezeichneten Bildsequenz mit 24-30 Vollbildern pro Sekunde lässt sich der Suchbereich für eine mögliche Korrespondenz jeweils auf einen Teilausschnitt des Bildes reduzieren, da sich der Bildinhalt auch bei schnellen Kameraschwenks nicht zu sehr ändert. Als Kriterium für die Ähnlichkeit der Regionen wird die Summe der Differenzen der Luminanzwerte verwendet [3]. Neuere Methoden verwenden Modelle, die auch invariant gegenüber Skalierungsänderungen (durch Änderung der Brennweite, *Zoom*) sind [4]. Doch auch diese Ansätze sind nur eine Näherung für die perspektivische Abbildung, die durch eine 3D-Kamerabewegung in der Bildebene hervorgerufen wird. Auch fließen hier keine komplexen Modelle für eine Änderung der Beleuchtung oder für das Reflektionsverhalten der Objekte ein.

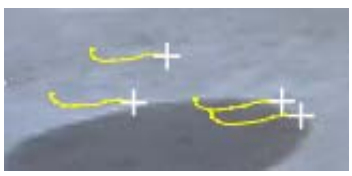

Bild 2 Trajektorien als Aneinanderkettung von Merkmalspunktkorrespondenzen aus mehreren Bildern.

2.3 Schätzung der Kameraparameter

Kameramodell:

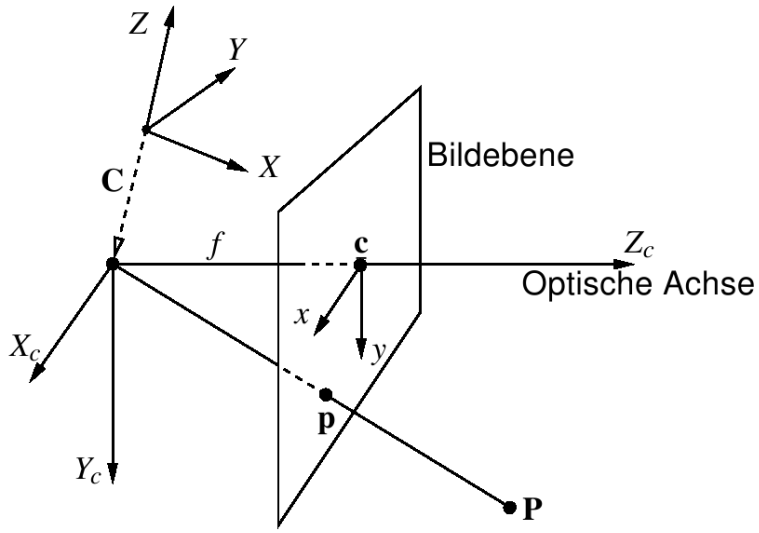

Bild 3 Lochkameramodell zur Abbildung eines 3D-Punktes *P* auf den Punkt *p* in der Bildebene einer Kamera mit der 3D-Position C, Kammerkonstante c und der Brennweite f.

Um die Abbildung eines 3D-Raumpunktes auf die Kamerabildebene zu beschreiben, wird das Lochkameramodell verwendet (siehe Bild 3). Die Abbildung wird durch die Zentralprojektion beschrieben. Da im realen Fall digitale Linsenkameras genutzt werden, müssen die Auswirkungen der Linse (im Wesentlichen radiale Verzerrungen der abgebildeten Bildpunkte) und der diskretisierten Bildebene berücksichtigt werden. Die Abbildungsgleichungen der Zentralprojektion bleiben erhalten.

2.3.1 Eliminierung von Fehlzuordnungen

Durch zufällige Ähnlichkeiten der Bildsignale sowie durch zeitveränderliche, bewegte Objekte in der größtenteils als starr angenommenen Objektgeometrie der Szene entstehen Fehlzuordnungen, die nicht zum Schätzen der Kameraparameter verwendet werden können. Diese werden durch das *RANSAC* (*Random Sample Consensus*)-Verfahren [5] mittels eines Abbildungsmodells erkannt, das den mathematischen Zusammenhang korrespondierender Merkmalspunkte in zwei Kamerabildern beschreibt. Bei einer Kamerabewegung mit Rotation und Translation ist dies die Abbildung mit der so genannten *Fundamentalmatrix* (F-Matrix) [6]. Zur Berechnung einer F-Matrix werden sieben Korrespondenzen benötigt. Das *RANSAC*-Verfahren wählt zufällig sieben Korrespondenzen aus der Menge der vorliegenden Korrespondenzen aus, berechnet die zugehörige F-Matrix und anschließend den Fehler, den alle anderen Korrespondenzen mit diesem Bewegungsmodell hervorrufen. Der Fehler einer Korrespondenz ergibt sich aus der Epipolargeometrie [6]. Liegt der Fehler oberhalb einer vorher festgelegten Schwelle, gilt die Korrespondenz als fehlerhaft (*outlier*). Dieser Vorgang wird iterativ wiederholt. Nach ausreichend vielen Iterationen wird die F-Matrix, die von den meisten Korrespondenzen unterstützt wurde als Beschreibung für die Bewegung der Kamera angenommen. Alle nicht unterstützten Korrespondenzen werden aus der Menge der Korrespondenzen gelöscht.

Beim RANSAC-Verfahren werden allerdings während der Berechnung Informationen verschenkt, da bei jeder Berechnung einer F-Matrix aus zufälligen Kandidaten die Ergebnisse vorhergehender Iterationen nicht genutzt werden. Möglichkeiten, dieses Wissen auszunutzen werden in [7] gezeigt.

2.3.2 Maximum-Likelihood-Schätzung

Hat zu Beginn der Bildsequenz eine ausreichend große Translation der Kamera stattgefunden, können initiale Kameraparameter geschätzt und 3D-Objektpunkte durch Triangulation bestimmt werden. Diese werden zusammen mit den weiteren Bildern der Sequenz genutzt, um eine Kostenfunktion im Sinne einer Maximum-Likelihood-Schätzung zu minimieren [8]. Beobachtungsgrößen der Schätzung sind die Koordinaten korrespondierender Merkmalspunkte in den Bildern. Dabei wird angenommen, dass sich die Position der Punkte wie eine zweidimensionale, weiße, gaußverteilte Zufallsvariable mit dem

Mittelwert an der wahren Position des Merkmalspunktes verhält. Die Minimierung der Kostenfunktion wird bei Betrachtung mehrerer Kamerabilder als *Bündelausgleich* bezeichnet. Ein geschätzter Kamerapfad und die zugehörigen 3D-Punkte sind in Bild 4 dargestellt.

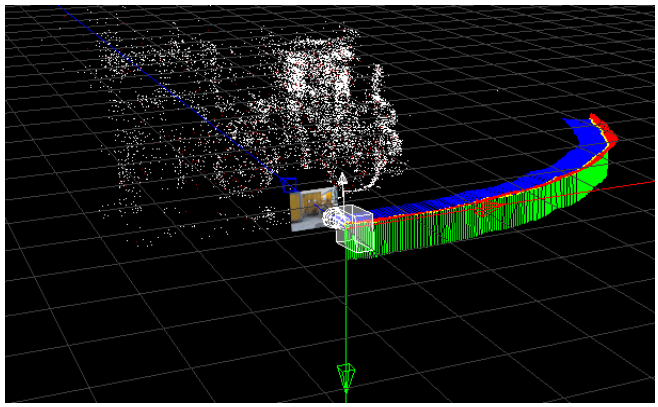

Bild 4 Geschätzter Kamerapfad mit den Rotationsachsen der Kamera und geschätzte 3D-Punkte (weiß) der Szene.

3 Einfügen virtueller Objekte

Die von Algorithmus geschätzten Daten, die die Kamerabewegung dokumentieren, werden in ein Format exportiert, das für 3D-Modelling-Programme als Import-Datei lesbar ist. Diese enthält neben dem errechneten Kamerapfad auch die geschätzten 3D-Punkte der Szene. Diese können dann genutzt werden, um das virtuelle 3D-Modell einzupassen und dessen Schatten entsprechend der 3D-Geometrie der Szene zu konstruieren. Einzelbilder der Originalsequenz und der zugehörigen Sequenz mit 3D-Modell sind in Bild 5 zu sehen. Weitere Beispiele stehen unter www.digilab.uni-hannover.de bereit.

Bild 5 Originalsequenz (oben) und Sequenz mit eingefügtem, virtuellem Objekt (unten); entworfen mit dem Voodoo camera tracker und Blender.

4 Zusammenfassung

In diesem Artikel wurde eine Technik zur Integration virtueller Objekte in reale Sequenzen vorgestellt. In einem ersten Schritt wird die Kamerabewegung bei der Aufnahme einer Videosequenz aus eben dieser Sequenz geschätzt. Durch das Verwenden eines robusten Verfahrens, das Fehlerzuordnungen detektieren kann, wird eine zuverlässige Schätzung des Kamerapfades sowie der 3D-Geometrie der Szene möglich. Der Kamerapfad und die 3D-Geometrie wird in ein 3D-Modellierungsprogramm eingelesen. Die Geometrie ermöglicht dann die Platzierung virtueller Objekte mit korrektem Bezug zur Videosequenz. Mithilfe des Kamerapfads wird dann das virtuelle Objekt gefilmt und anschließend mit der realen Sequenz gemischt. Die Funktionalität des Algorithmus wurde anhand von einigen Bildern einer Testsequenz, in die ein künstlich generiertes Objekt eingefügt wurde, dargestellt.

5 Referenzen

[1] C. Harris, M. Stephens: “A Combined Corner and Edge Detector”. In “4th Alvey Vision Conference”, S. 147-151. 1988

[2] D. Lowe: “Object Recognition from Local Scale-Invariant Features”. In Proceedings of the International Conference on Computer Vision. S. 1150-1157. Corfu, Greece. September 1999

[3] B. Lucas, T. Kanade: “An iterative image registration technique with an application to stereo vision”. In 7th International Conference on Artificial Intelligence”. S. 674-679. 1981

[4] D. Lowe: “Distinctive Image Features from Scale-Invariant Keypoints”. International Journal of Computer Vision. Volume 60 (2004)

[5] R. M. A. Fischler, C. Bolles: “Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography”. Communications of the ACM, Bd. 24(6):381-395. 1981

[6] R. I. Hartley, A. Zisserman: “Multiple View Geometry”. Cambridge University Press. 2000

[7] O. Urfalioglu: “Robust Estimation with Non Linear Particle Swarm Optimization“. Proceedings of Mirage 2005 (Computer Vision / Computer Graphics Collaboration Techniques and Application), INRIA Rocquencout, Frankreich. März 2005

[8] T. Thormählen: „Zuverlässige Schätzung der Kamerabewegung aus einer Bildfolge“. Dissertation. Universität Hannover 2006.

Verfasser-Portraits

Dipl.-Math. Kai Cordes absolvierte 2003 das Studium der Mathematik mit der Studienrichtung Informatik an der Leibniz Universität Hannover mit Abschluss Diplom. Seit Ende des Studiums ist er wissenschaftlicher Mitarbeiter am Institut für Informationsverarbeitung (TNT) der Leibniz Universität Hannover. Seine Forschungsschwerpunkte liegen im Bereich der automatischen Schätzung von Kameraparametern aus monokularen Bildfolgen.

Prof. Dr.-Ing. Jörn Ostermann schloss das Studium der Elektrotechnik 1988 ab und promovierte 1994 an der Universität Hannover. Von 1993 bis 1994 arbeitete er bei AT&T Bell Labs und anschließend bis 2003 bei AT&T Labs-Research. Seit 2003 leitet er das Institut für Informationsverarbeitung an der Leibniz Universität Hannover. Er ist ein Fellow des IEEE, ein ehemaliger Distinguished Lecturer der IEEE CAS Society, Empfänger eines ISO Awards und des AT&T Standards Recognition Awards. Er ist Coautor eines Lehrbuchs zum Thema Video Communications. Er veröffentlichte mehr als 100 wissenschaftliche Beiträge und 22 Patente. Seine Interessen und Arbeitsschwerpunkte liegen auf dem Gebiet der Audio- und Videosignalverarbeitung sowie der Kommunikation über IP-Netzwerke.