

3D-Exploration mit einer sich bewegenden Kamera*

Bernd Jähne¹ und Helmut Herrmann²

¹Interdisziplinäres Zentrum für wissenschaftliches Rechnen (IWR)
Universität Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg
Telefon: 06221/54 8827, Telefax: 54 8850
Bernd.Jaehne@iwr.uni-heidelberg.de, <http://klimt.iwr.uni-heidelberg.de>

²Aeon Verlag & Studio Walter H. Dorn, Fraunhoferstr 51B, D-63454 Hanau
Telefon: 06181/23525, Telefax: 257954
Helmut.Herrmann@aeon.de, <http://www.aeon.de>

Zusammenfassung

Fortschritte in der Bewegungsbestimmung erlauben nun auch den Einsatz der Structure-from-Motion-Technik zur genauen 3D-Rekonstruktion. Das allgemeine Prinzip wird hier in der Sprache der Differentialgeometrie formuliert und mit zwei einfachen Kameraanordnungen, einer planar und einer auf einem Kegel rotierenden Kamera, demonstriert.

1 Einleitung

In der Literatur wird eine Vielzahl von Verfahren zur 3D-Exploration beschrieben. Während sich in der industriellen Praxis Stereo-, Lichtschnitt- und codierte Beleuchtungstechniken zur genauen Messung durchgesetzt haben [*Klette et al.*, 1996; *Jiang und Bunke*, 1997], spielen andere Verfahren wie Shape-from-Shading und Structure-from-Motion [*Horn*, 1986], von denen bekannt ist, daß das menschliche visuelle System teilweise auf diesem Wege seine 3D-Information bezieht, kaum eine Rolle. Obwohl deren Prinzipien wohl verstanden sind, sind diese Verfahren in der Praxis bisher wenig robust und zu ungenau.

In den letzten Jahren hat allerdings die Bewegungsanalyse große Fortschritte gemacht, insbesondere der Ansatz, die Bewegung nicht nur aus zwei aufeinanderfolgenden Bildern, sondern aus einer Analyse der Grauwertstrukturen in Orts/Zeit-Bildern zu ermitteln [*Jähne*, 1993]. Dadurch konnte die Genauigkeit der Bewegungsbestimmung erheblich gesteigert werden [*Haußecker and Jähne*, 1997; *Jähne*, 1997c]. Somit erscheint es angebracht zu untersuchen, welche Möglichkeiten sich durch Orts/Zeit-Bilder für die 3D-Exploration ergeben.

2 Bewegung als Orientierung im Orts/Zeit-Raum

In diesem Abschnitt betrachten wir den Zusammenhang zwischen Bewegung und Orientierung im Orts/Zeit-Raum als Basis für die 3D-Rekonstruktion mit einer sich

*Vorabdruck, Tagungsband des Symposiums Bildverarbeitung'97 - Forschen, Entwickeln, Anwenden, 26.-28. November 1997, R.-J. Ahlers (Hrsg.), Technische Akademie Esslingen

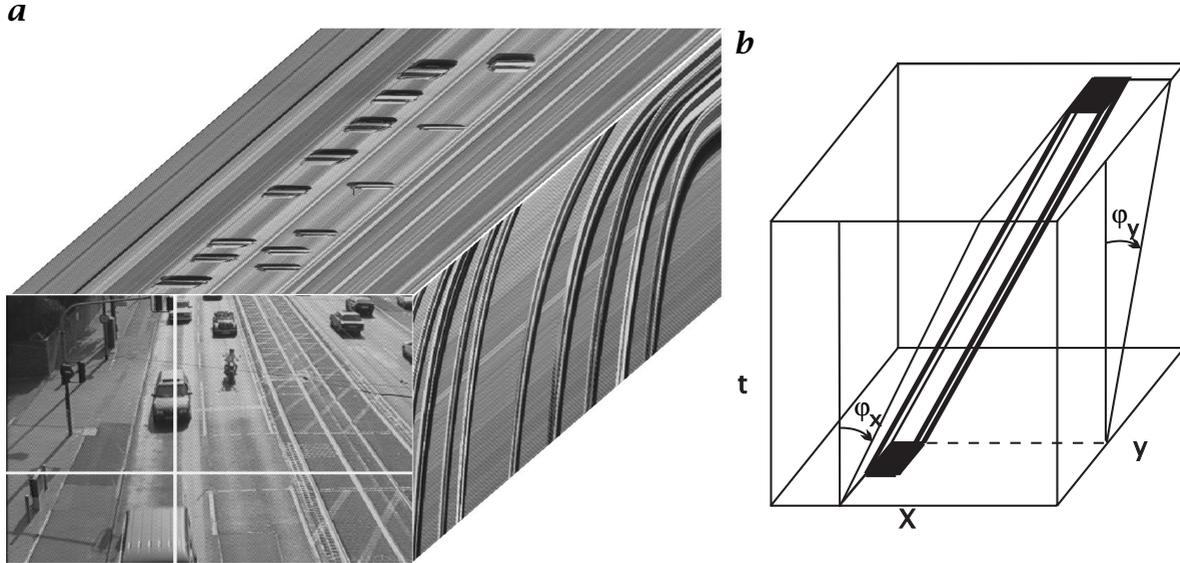


Abbildung 1: Orts/Zeit-Bilder: **a** Verkehrsszene auf der Hanauer Landstraße in Frankfurt/Main. Die Zeitachse kommt aus der Tiefe auf den Betrachter zu. Auf der rechten Seite des Quaders ist eine yt -Scheibe für die mit einer vertikalen weißen Linie im xy -Bild markierte x -Koordinate zu sehen, während die Oberseite eine mit der horizontalen weißen Linie markierte xt -Scheibe zeigt. **b** Schematische Darstellung der Beziehung zwischen Orientierung und Geschwindigkeit, vgl. (2). (Aus Jähne [1997a].)

bewegenden Kamera. Der Orts/Zeit-Raum ist ein dreidimensionaler Raum, der über einer Zeitkoordinate und zwei Raumkoordinaten aufgespannt ist. Ein dreidimensionales Orts/Zeit-Bild kann man sich als Stapel aufeinanderfolgender Bilder denken und diesen als Bildquader wie in Abb. 1 darstellen.

Zur Analyse der Bewegung in Orts/Zeit-Bildern betrachten wir zunächst ein einfaches Beispiel mit nur einer Raumkoordinate (horizontal) und einer Zeitkoordinate (vertikal). Ein ruhendes eindimensionales Objekt weist vertikal orientierte Grauwertstrukturen auf. Bewegt sich dagegen das Objekt, verschiebt es sich von Bild zu Bild und erscheint als geneigte Grauwertstruktur. Dabei ist die Geschwindigkeit u mit der Orientierung im Orts/Zeit-Bild direkt verknüpft. Im einfachen Fall eines 2D-Orts/Zeit-Bildes ist sie gegeben durch

$$u = -\tan \varphi, \quad (1)$$

wobei φ der Winkel ist zwischen der t -Achse und der Richtung, in der die Grauwerte konstant sind. Das Minuszeichen in (1) kommt daher, daß Winkel gegen den Uhrzeigersinn positiv gezählt werden. Die Erweiterung auf zwei Raumdimensionen ist einfach, wie in Abb. 1b gezeigt. Hier ergibt sich die Geschwindigkeit zu

$$\mathbf{u} = - \begin{bmatrix} \tan \varphi_x \\ \tan \varphi_y \end{bmatrix}. \quad (2)$$

Die Winkel φ_x und φ_y sind als die Winkel zwischen der t -Achse und den Projektionen des Vektors in Richtung der konstanten Grauwerte auf eine beliebige yt - bzw. xt -Ebene definiert.

Die Orientierung der Grauwertstrukturen im Orts/Zeit-Raum durch die Bewegung ergibt eine ausgezeichnete Richtung, in der sich die Grauwerte nicht ändern. Diese Richtung kann durch einen Einheitsvektor \bar{e} ausgedrückt werden, der über (2) in direkter Beziehung zur Geschwindigkeit steht:

$$\bar{e} = \frac{1}{\sqrt{u_x^2 + u_y^2 + 1}} \begin{bmatrix} u_x \\ u_y \\ 1 \end{bmatrix}. \quad (3)$$

3 Tiefe aus Orientierung im Orts/Zeit-Raum

Bolles et al. [1987] haben erstmals die Zusammenhänge zwischen linearer Bewegung und Tiefengewinnung aufgezeigt und später auf beliebige Bewegungen verallgemeinert [*Baker und Bolles*, 1989]. Dabei sind sie von den Begriffsbildungen der Stereometrie ausgegangen. Auch wir gehen hier von einer Stereokameraanordnung aus, beschreiben dann aber die grundlegenden Zusammenhänge mit Hilfe der Differentialgeometrie im Raum gekrümmter Kurven.

Stereokameraaufnahmen sind wohl die bekannteste und weitverbreitetste Methode zur Gewinnung von Tiefeninformation aus Bildern. Das Grundprinzip besteht darin, daß sich ein Objekt durch die unterschiedlichen Betrachtungswinkel in den beiden Bildern an verschiedenen Positionen befindet, wobei die Positionsdifferenz, die Parallaxe, eine Funktion der Tiefe des Objekts ist. Wir betrachten hier eine Kameraanordnung mit sich schneidenden optischen Achsen und berechnen die Abhängigkeit der Parallaxe von der Position des Objekts in z -Richtung (Abb. 2a).

Die Kameras seien symmetrisch so ausgerichtet, daß sich ihre optischen Achsen im Objektraum im Abstand Z_0 von der stereoskopischen Basis b , der Verbindungslinie zwischen den Kameras, schneiden. Dann ergibt sich die Parallaxe p eines Punktes nahe der optischen Achse an der Stelle $Z_0 - \Delta Z$ zu [*Jähne*, 1997b]

$$p = \frac{d_i \cos^2(\varphi)}{Z_0(Z_0 - \Delta Z \cos^2 \varphi)} b \Delta Z. \quad (4)$$

Dabei ist d_i die Bildweite, die bei großen Gegenstandsweiten etwa gleich der Brennweite ist, und φ der halbe Winkel zwischen den Kameraachsen. Gleichung (4) besagt, daß die Parallaxe im wesentlichen proportional zur Stereobasis und der Tiefe bezogen auf den Schnittpunkt der optischen Achsen der beiden Kameras ist.

Hat man es mit ruhenden oder sich nur langsam ändernden Objekten zu tun, kann man auf die zweite Kamera verzichten und stattdessen die erste Kamera auf die Position der zweiten bewegen und so beide Stereobilder mit der gleichen Kamera aufnehmen. Geht man noch einen Schritt weiter und nimmt während der Bewegung der Kamera auf einer vorgegebenen Bahn kontinuierlich Bilder auf, ergibt sich eine Bildsequenz, welche mehr Information enthält als lediglich ein einziges Stereobildpaar. Durch die sich in der Bildsequenz kontinuierlich ändernden Objektpositionen scheinen sich die Objekte zu bewegen, und es entstehen die bereits im ersten Abschnitt angesprochenen Grauwertorientierungen. Daraus ergibt sich, daß

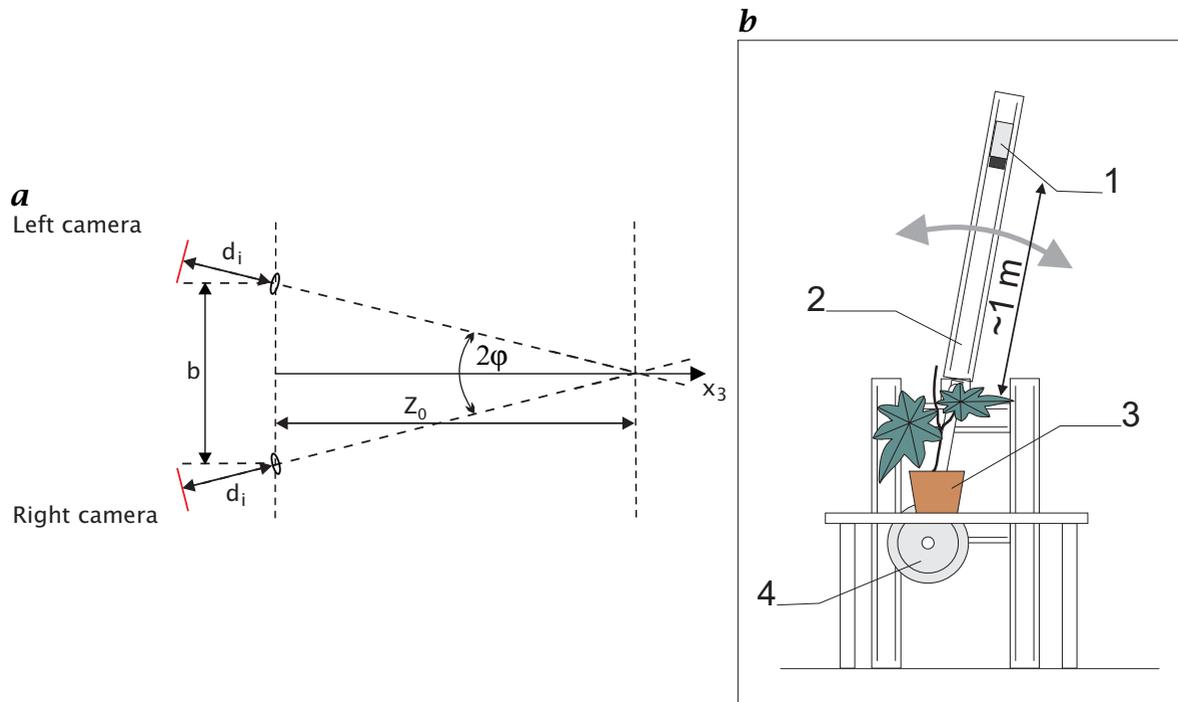


Abbildung 2: Von der Stereokamera zur planar rotierenden Kamera: **a** Stereokameraaufbau mit geneigten Achsen. **b** Rotierende Kamera: (1) Kamera, (2) Schwingarm, (3) beobachtetes Blatt in der Nähe der Drehachse, (4) Gegengewicht.

die Geschwindigkeit der Objekte in einer solchen Bildsequenz die Tiefeninformation enthält.

Um die Beziehung zwischen der Geschwindigkeit \mathbf{U} der Kamera, der in der Bildebene gemessenen Geschwindigkeit $\mathbf{u} = [u, v]^T$ und der Tiefeninformation zu erhalten, kann man die Stereobasis als einen zeitlich veränderlichen Vektor im Raum auffassen, physikalisch gesprochen, als eine Bahnkurve, deren zeitliche Ableitung die Geschwindigkeit $\mathbf{U} = [U, V, W]^T$ der Kamera ist. Die Parallaxe entspricht dann der in der Bildebene gemessenen Geschwindigkeit. Die Gleichungen vereinfachen sich dadurch, daß die Geschwindigkeiten klein sein sollen (höchstens eine Verschiebung von 1 Pixel/Bild, um eine kontinuierliche Bildfolge ohne zeitliches Aliasing zu erhalten). Deswegen entfallen die Korrekturterme in (4) mit $\cos^2(\varphi)$. Unter all diesen Voraussetzungen ergibt sich für eine Kamera mit einer beliebigen Geschwindigkeit aus (4)

$$\mathbf{u} = \frac{d_i}{Z_0(Z_0 - \Delta Z)} \mathbf{U} \Delta Z. \quad (5)$$

Dabei sind alle 3D-Koordinaten (Z und Z_0) und Geschwindigkeiten in einem an die Bahnkurve der Kamera angepaßten Koordinatensystem, das man in der Differentialgeometrie als *begleitende Dreibein* einer Kurve im Raum bezeichnet, angegeben. Der Ursprung dieses Koordinatensystems befindet sich im momentanen Punkt \mathbf{X}_0 der Bahnkurve. Die x -Achse liegt in Richtung des Geschwindigkeitsvektors $\mathbf{U} = d\mathbf{X}/dt$, die y -Achse in Richtung der *Binormalen* $\mathbf{U} \times \mathbf{A}$, $\mathbf{A} = d\mathbf{U}/dt$, und die z -Achse in Rich-

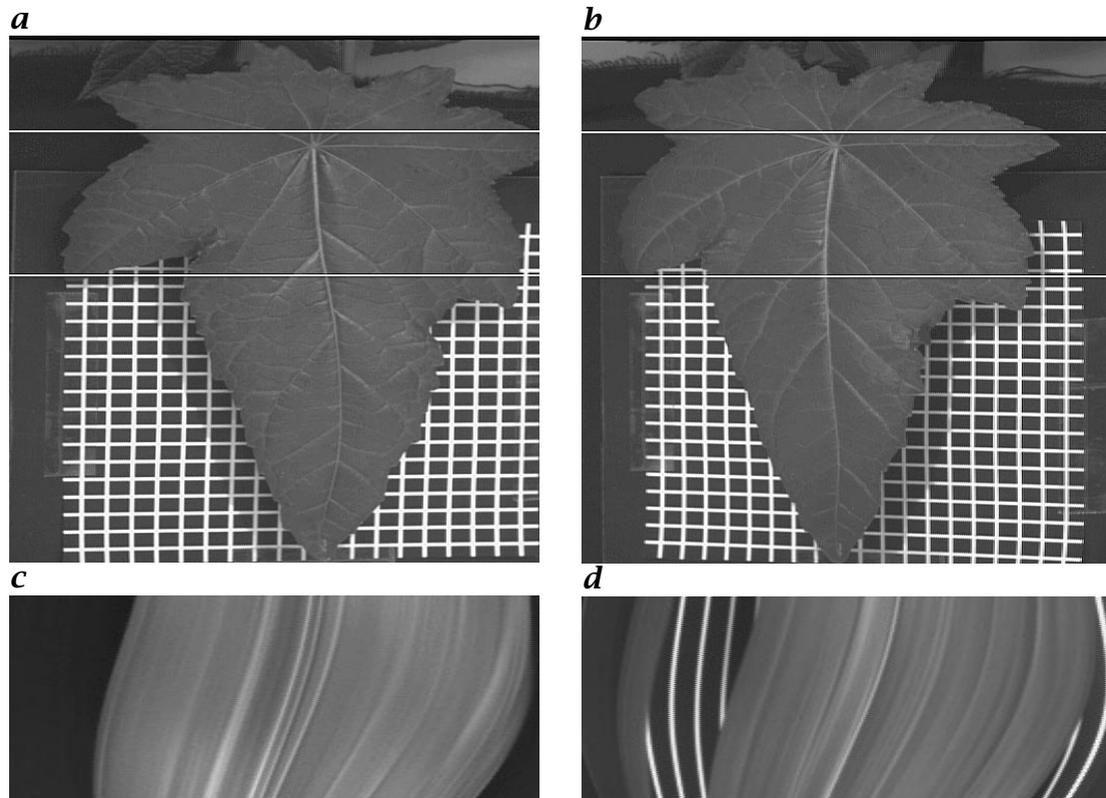


Abbildung 3: Bildsequenz eines Pflanzenblatts, aufgenommen mit der planar rotierenden Kamera: **a** und **b** erstes und letztes Bild der Sequenz, **c** und **d** Orts/Zeit-Schnitte der in **a** und **b** markierten Zeilen.

tung der *Hauptnormalen* der Bahnkurve $(U \times A) \times U$. Der Zusammenhang zwischen der Bewegung der Kamera und der Geschwindigkeit in der Bildebene stellt sich in Gleichung (5) so einfach dar, weil er im richtigen Koordinatensystem beschrieben wurde. Für die Bestimmung der 3D-Koordinaten eines Objektes müssen diese Koordinaten selbstverständlich noch in Weltkoordinaten transformiert werden.

In den beiden folgenden Abschnitten werden zwei einfache sich bewegende Kamerasysteme behandelt, deren Bewegung durch eine entsprechende mechanische Anordnung eine besonders einfache 3D-Rekonstruktion erlaubt.

4 Realisierung I: Planar rotierende Kamera

Dieser Kameratyp ist die direkte Erweiterung eines Stereokameraaufbaus mit geeigneten Achsen in ein System mit einer sich bewegenden Kamera, eine in der Ebene rotierende Kamera. Eine Skizze dieser Anordnung mit Anwendung auf die 3D-Vermessung von Blattwachstum in der Botanik [Jähne *et al.*, 1996] ist in Abb. 2b dargestellt. Die Bewegung dieser Kamera entspricht einer ebenen Kreisbewegung. Wenn wir das Zentrum des Weltkoordinatensystems in den Schnittpunkt der optischen Achse der Kamera mit der Drehachse legen, so kann die Bahnkurve der Kamera bei

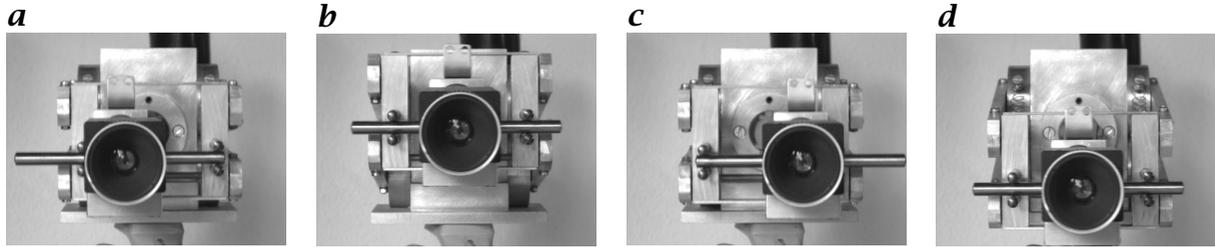


Abbildung 4: Rotierende Kamera zur 3D-Rekonstruktion. Die Kamera bewegt sich kreisförmig auf einem Kegel, ohne selbst zu rotieren.

gleichmäßiger Bewegung beschrieben werden als

$$X(t) = R \begin{bmatrix} \cos(\omega t) \\ \sin(\omega t) \end{bmatrix}. \quad (6)$$

Bei dieser einfachen Anordnung fällt der Referenzpunkt für die Tiefenmessung mit der Drehachse und damit dem Zentrum des Weltkoordinatensystems zusammen. Die Bewegung verläuft nur in Richtung der X -Achse. Daher läßt sich die Wirkungsweise dieses Systems durch x t -Schnittbilder in Bildsequenzen zeigen (Abb. 3). Man sieht sehr deutlich, daß sich das über dem Eichgitter gelegene Blatt durch eine andere Orientierung im Orts/Zeit-Bild und damit eine andere Geschwindigkeit darstellt als das Gitter. Nach Gleichung (5) ist die gemessene Geschwindigkeit in einem Punkt des Bildes direkt proportional zur Tiefe.

Man erkennt auch, daß sich Okklusionen, die in konventiellen Stereo- und Lichtschnittverfahren sehr schwer zu behandeln sind, bei diesem Verfahren als *Ereignisse in der Zeit* darstellen. Zu einem gegebenen Zeitpunkt verschwindet oder erscheint eine Struktur plötzlich. Man beachte dazu z. B. die unter dem Blatt verschwindenden bzw. auftauchenden Linien des weißen Gitters in Abb. 3d.

5 Realisierung II: Auf einem Kegel rotierende Kamera

Die im vorherigen Abschnitt beschriebene planar rotierende Kamera hat noch den Nachteil, daß die Bewegung nur in eine Richtung erfolgt. Liegt eine Kante in Richtung dieser Bewegung, so läßt sich deren Geschwindigkeit und damit die Tiefe nicht bestimmen. Dies entspricht dem klassischen Blendenproblem bei der Stereokameraanordnung. Dieses Problem kann mit einer auf einem Kegel rotierenden Kamera vermieden werden. Durch eine komplizierte mechanische Anordnung bewegt sich die optische Achse auf einem Kegel, ohne dabei selbst zu rotieren. Vier Phasen der Kamerabewegung sind in Abb. 4 dargestellt.

Ohne in weitere Details zu gehen, ist offensichtlich, daß die Bewegung im Bild kleine Kreisbahnen sein werden, deren Radien in erster Näherung proportional zur Tiefe sind. Damit ändert sich die Richtung der Bewegung periodisch mit der Drehbewegung, und es läßt sich auch an Kanten immer eine Parallaxe bestimmen.

Ein Beispielbild und seine Auswertung ist in Abb. 5 gezeigt. Die Berechnung der Bewegung wurde mit dem Tensorverfahren durchgeführt [*Haußecker und Jähne*,

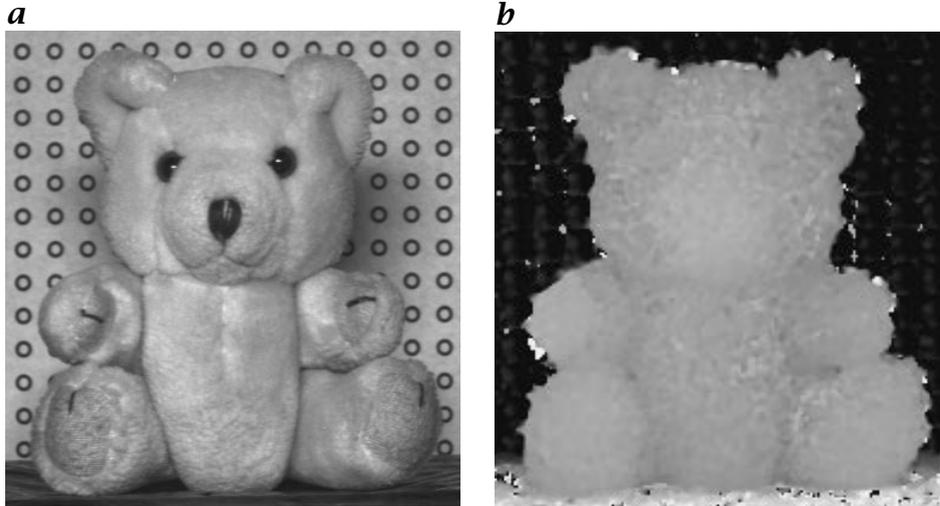


Abbildung 5: Beispiel für die 3D-Rekonstruktion mit der auf einem Kegel rotierenden Kamera (Abb. 4): **a** Originalbild (der unvermeidliche "Computer-Vision-Teddy"), **b** Tiefenkarte (näher ist heller) ohne Nachverarbeitung.

1997], mit dem sich dichte Verschiebungsvektorfelder bestimmen lassen. Der Fehler der berechneten Verschiebung liegt deutlich unter $1/10$ Pixel pro Bild, so daß sich selbst bei einem kleinen Öffnungswinkel des Kegels (entsprechend einer kleinen Stereobasis) genaue und dichte Tiefenkarten bestimmen lassen.

6 Schlußfolgerungen und Ausblick

In diesem Beitrag wurden sowohl das generelle Prinzip als auch erste konkrete Realisierungen gezeigt, wie eine einzelne sich bewegende Kamera zur 3D-Exploration eingesetzt werden kann. Im letzten Abschnitt wurde dargestellt, daß die hohe Genauigkeit moderner Verfahren zur Bewegungsbestimmung diesen Ansatz als interessante Alternative erscheinen läßt. Das langfristige Ziel dieser Arbeiten ist es, dieses Verfahren auch bei wesentlich allgemeineren Kamerabewegungen einzusetzen.

Danksagung

Die teilweise finanzielle Unterstützung dieser Arbeit im Rahmen der interdisziplinären Forschergruppe "Bildfolgenanalyse zum Studium dynamischer Prozesse" von der Deutschen Forschungsgemeinschaft (Jae 395/6-1) wird dankbar anerkannt. Die Autoren danken auch W. Haußecker, TePro Eppingen, für die Konstruktion der auf einem Kegel rotierenden Kamera.

Literatur

- [1] Baker, H. H., and R. C. Bolles, 1989. Generalizing epipolar-plane image analysis on the spatiotemporal surface, *Int. J. Comp. Vision*, 3, 33-49.
- [2] Bolles, R. C., H. H. Baker, and D. H. Marimont, 1987. Epipolar-plane image analysis: an approach to determining structure from motion, *Int. J. Comp. Vision*, 1, 7-55.
- [3] Haußecker, H., and B. Jähne, 1997. A tensor approach for precise computation of dense displacement vector fields, Proc. *Mustererkennung 1997*, Braunschweig, 15.-17. September 1997, E. Paulus und F. M. Wahl (Hrsg.), Informatik Aktuell, Springer, Berlin, 199-208.
- [4] Horn, B. K., 1986. *Robot vision*. MIT Press, Cambridge, MA.
- [5] Jähne, B., 1993. *Spatio-Temporal Image Processing, Theory and Scientific Applications*, Lecture Notes in Computer Science, Vol. 751, Springer, Berlin.
- [6] Jähne, B., *Digitale Bildverarbeitung*, 1997a. 4. vollständig überarbeitete Auflage, Springer, Berlin.
- [7] Jähne, B., 1997b. *Practical Handbook on Digital Image Processing for Scientific Applications*, CRC-Press, Boca Raton, FL, USA.
- [8] Jähne, B., 1997c. Performance characteristics of low-level motion estimators in spatiotemporal images, Proc. *DAGM-Workshop Performance Characteristics and Quality of Computer Vision Algorithms*, Braunschweig, 18. September 18 1997, W. Foerstner (ed.), Institute of Photogrammetry, Univ. Bonn.
- [9] Jähne, B., H. Haußecker, F. Hering, G. Balschbach, J. Klinke, M. Lell, D. Schmundt, M. Schultz, U. Schurr, M. Stitt, and U. Platt, 1996. The role of active vision in exploring growth, transport, and exchange processes, Proc. *Aktives Sehen in technischen und biologischen Systemen*, Workshop der GI-Fachgruppe 1.0.4. Bildverstehen Hamburg, 3.-4. Dezember 1996, B. Mertsching (Hrsg), Proc. in Artificial Intelligence 4, infix, Sankt Augustin, 194-202.
- [10] Jiang, X., und H. Bunke, 1997. *Dreidimensionales Computersehen, Gewinnung und Analyse von Tiefenbildern*, Springer, Berlin.
- [11] Klette, R., A. Koschan und K. Schlüns, 1996. *Computer Vision, räumliche Information aus digitalen Bildern*, Vieweg, Braunschweig.
- [12] Scharr, H., S. Körkel, and B. Jähne, 1997. Numerische Isotropieoptimierung von FIR-Filtern mittels Querglättung, Proc. *Mustererkennung 1997*, Braunschweig, 15.-17. September 1997, E. Paulus und F. M. Wahl (Hrsg.), Informatik Aktuell, Springer, Berlin, 367-374.