

Möglichkeiten und Grenzen aktiven Sehens mit passiven Sensoren

Dietrich W. R. Paulus *and* Joachim Denzler

Lehrstuhl für Mustererkennung (Informatik 5)
Universität Erlangen-Nürnberg
Martensstraße 3, D-91058 Erlangen, F.R. of Germany

Erscheint im

9. Fachgespräch für autonome mobile Systeme
München, Oktober 1993

Möglichkeiten und Grenzen aktiven Sehens mit passiven Sensoren

Dietrich Paulus und Joachim Denzler

Universität Erlangen–Nürnberg
Lehrstuhl für Mustererkennung (Informatik 5)
Martensstr. 3, D–91058 Erlangen
Tel.: +49–9131–85-7894, FAX: +49–9131–303811
email: {paulus,denzler}@informatik.uni-erlangen.de

Dieser Beitrag beschreibt einen Versuchsaufbau aus stationären und beweglichen Kameras zur Ermittlung von 3D Information für die Steuerung eines Roboters. Schwerpunkt bildet die Modellierung der Zoom-Kameras mit dem Zweck des fovealisierten Sehens. Aktives Sehen wird durch motorische Veränderung der Brennweite erreicht.

1 Einleitung

Im Bereich der Bildverarbeitung und –analyse ermöglicht heute die billigere und leistungsfähigere Hardware einen früher unerreichten Durchsatz an Daten. Anstelle der Verarbeitung statischer Einzelbilder oder der Verarbeitung von zu einem anderen Zeitpunkt aufgenommenen Bildfolgen tritt nun die Echtzeit–Bildverarbeitung mit einem kontinuierlich eintreffenden Datenstrom. Durch die harten Obergrenzen für die Reaktionszeit und die maximale Verarbeitungsdauer ergeben sich völlig neue Anforderungen an die Algorithmen und Systeme, und der Begriff des “aktiven Sehens” tritt in den Vordergrund. Hier ergeben sich neue Berührungspunkte von maschinellem Sehen und Robotik (“vision guided robots, automated vehicles, ...”).

Wie in allen Computersystemen ist zur Lösung eines Problems eine Analyse und Problemspezifikation erforderlich. In der Bildanalyse erfolgt hierbei eine Modellierung der Aufnahmebedingungen und der erwarteten Objekte im Sichtbereich. In diesem Beitrag beschäftigen wir uns im speziellen mit der Modellierung der Kameraeigenschaften und der Aufnahme.

2 Begriffe und Definitionen

Mehr als ein Jahrzehnt prägte das sogenannte “Marr Paradigma”, unter dem die statische Analyse von Bildern, inverse Bilderzeugung und Suche nach einer symbolischen Beschreibung des gesamten Bildes zusammengefaßt werden, die Struktur der meisten Bildverarbeitungssysteme [8].

Echtzeitbildverarbeitung, im speziellen Echtzeitbildfolgenverarbeitung, verlangt heute ein Abrücken von Marrs Paradigma. ([1]: “... add a new dimension - time - which

requires new understanding, techniques and paradigms”). Dieses neue Paradigma wird vielfach als “aktives Sehen” bezeichnet ([1], [3]).

In der Literatur findet man zwei hauptsächliche Verwendungen des Begriffs “aktives Sehen”. Beiden ist das Prinzip gemeinsam, daß nicht das gesamte Bild, sondern nur ein interessanter Ausschnitt bearbeitet und ebenfalls nicht auf jedes Bild einer Bildfolge dieselbe Folge von Operationen ausgeführt werden muß. Dies kann durch aktiven Einsatz der Aufnahmegeräte (Hardware) oder der Algorithmen (Software) geschehen. In [2] wird aktives Sehen als ein Studium der Modellierungs- und Kontrollstrategie für die Wahrnehmung verstanden. Deutlich wird der Unterschied zum früheren Marr-Paradigma in dem Gegensatz von “Sehen und Schauen”. Das motivierende Vorbild dieses neuen Paradigma ist der Mensch selbst. So bewegt der Mensch seinen Kopf, um Szenen zu erkunden, und das Objekt des größten Interesses wird automatisch in die Fovea (siehe Abbildung 1) geholt. Dieses Verhalten wird nun auf das Rechnersehen übertragen. Festzuhalten ist, daß keine aktiven Sensoren benötigt werden, sondern vielmehr die Parameter des Sensors adaptiv verändert werden müssen. Gesteuert wird diese Veränderung innerhalb eines geschlossenen Regelkreises aus Sensorik und Aktorik durch den Zustand und das Ziel der Handlung. Nach [2] sind dazu Modelle der Sensoren sowie Fehlerabschätzungen der angewendeten Verfahren nötig.

Brooks verwendet den Begriff “aktives Sehen” im Zusammenhang mit Prozessen, die aktiv interessante Gesichtspunkte in Bildern über die Zeit hinweg verfolgen. Nach [3] bedeutet aktives Sehen eine Strategie für kontinuierliche “Beobachtungen” im Gegensatz zu Schnappschüssen von Szenen.

[4] stellt das passive Sehen dem aktiven Sehen in folgender Weise gegenüber:

passives Sehen	aktives Sehen
Verwendung aller Module	Verwendung einiger geeigneter Module
Bearbeitung des gesamten Bildes	Bearbeitung geeigneter Ausschnitte
Verwendung maximaler Auflösung	Verwendung ausreichender Auflösung

In [11] bedeutet aktives Sehen sowohl die Kontrolle der Kameraparameter (Orientierung, Fokus, Zoom, Blende) als Antwort auf die Anforderungen der Aufgabe und der externen Stimuli als auch die selektive Bearbeitung in Raum, Zeit und Auflösung.

Ein häufig erwähntes Prinzip für das aktive Sehen ist die an biologische Systeme angelehnte *Fovealisierung*. Hierbei wird der zentrale Blickbereich an vergleichsweise vielen Punkten abgetastet, wogegen ein Überblicksbereich in geringerer Auflösung betrachtet wird. Die biologische und eine mögliche technische Lösung sind in Abbildung 1 dargestellt. Ein CCD-Chip mit logarithmischer Auflösung wird im Rahmen eines ESPRIT-Projekts zur Zeit untersucht.

In unserem Beitrag werden die bisher genannten Begriffe folgendermaßen definiert:

Aktives Sehen umfaßt *Aufmerksamkeit* und *selektives Sehen* (bzgl. Raum, Auflösung und Zeit), unabhängig davon, ob dies in der Kamera oder der darauffolgenden Verarbeitung erfolgt [11].

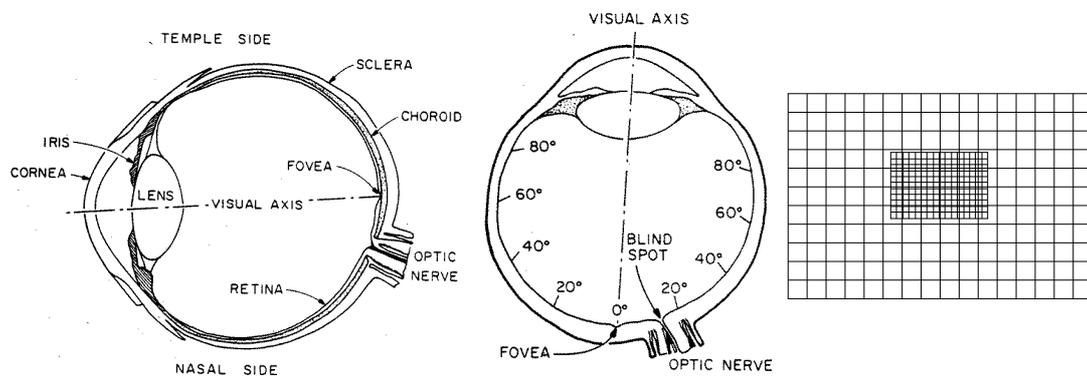


Abbildung 1: Fovea im menschlichen Auge (links, Mitte) und mögliche technische Realisierung (rechts).

Aktive Verfahren liefern Bilder, die durch dynamischen und adaptiven Einsatz der Aufnahmesensoren zum Zwecke der Szenenerkundung entstehen [7].

Passive Sensoren sind Aufnahmeeinrichtungen, die keine aktive Veränderung der Szene oder Beleuchtung vornehmen.

3 Versuchsaufbau

Im Versuchsaufbau werden verschiedene Kameras mit unterschiedlichen Abbildungseigenschaften (Nadelör, telezentrisches Objektiv, Zoom, Makro) verwendet. Eine Möglichkeit des aktiven Sehens besteht in der *Ausnutzung* dieser Unterschiede. Erforderlich ist ein Kamera- und Aufnahmemodell, das diese Unterschiede umfaßt und außerdem der diskreten Abtastung durch die CCD-Kameras Rechnung trägt. Passive Sensoren, die eine Szene nicht verändern, werden dabei den aktiven Sensoren vorgezogen, um gegenseitige Störungen der Sensoren auszuschließen. Neben den oben erwähnten Kameras stehen weitere Geräte zur Verfügung, die aktives Sehen ermöglichen: ein Roboter (5 Achsen) auf einem Linearschlitten mit aufmontierter Kamera, ein rechnergesteuertes Stereogestell, ein handgesteuertes Stereogestell, ein Drehtisch für Objekte und zwei motorisch steuerbare Zoom-Kameras (Stereo). Die Anordnung der Kameras ist aus Abbildung 2 zu ersehen.

Experimentell soll fovealisiertes Sehen für starre Objekte mit Zoom-Kameras simuliert werden. Mit dem Roboter wird die Kamera an definierte Positionen bewegt, um mehrere Ansichten eines Objekts zu erhalten. Mit mehreren Kameras (unterschiedlichen Typs) werden bewegte Objekte aufgenommen. Dazu ist eine Datenfusion der verschiedenen Informationsquellen erforderlich.

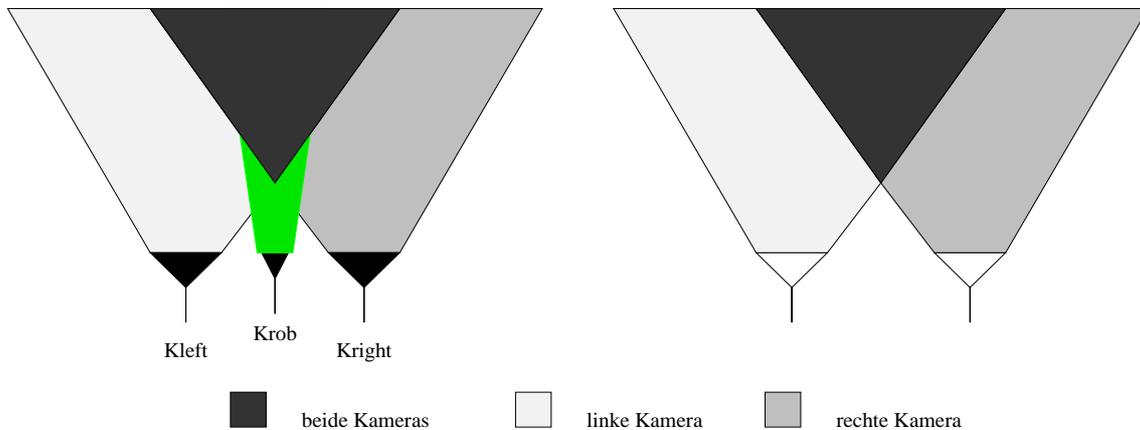


Abbildung 2: Kameraanordnung im Versuchsaufbau mit parallelen Achsen.

4 Möglichkeiten und Grenzen

Es soll nun beschrieben werden, welche *prinzipiellen* Möglichkeiten und Grenzen der Erkennung und Lokalisierung von 3D-Objekten und deren Bewegung sich aus dem Versuchsaufbau ergeben.

Grundlegende Modelle vieler Bildverarbeitungsalgorithmen sind die Lochkamera (z.B. in [10]), die schwache perspektivische Projektion (z.B. in [12]) oder die skalierte parallele Projektion (z.B. in [6]).

Wir verwenden im folgenden das Lochkameramodell; Unterschiede ergeben sich in der Literatur in der Wahl des Koordinatenursprungs, der im Brennpunkt oder in der Bildebene liegen kann. Wegen des veränderlichen Brennpunkts in Zoomkameras wird hier der Nullpunkt in der unbewegten Bildebene gewählt. Die Abbildung von Weltkoordinaten in ein Kamerakoordinatensystem (Rotation und Verschiebung) wird als Identität angenommen, da sie keinen Einfluß auf die folgenden Untersuchungen hat.

Die grundlegenden Beziehungen zwischen den Größen im Lochkameramodell (Brennweite f_1 , Bildgröße a_1 , Entfernung Z , Objektgröße Y) mit Nullpunkt in der Bildebene sind in (1) zusammengefaßt.

$$\frac{a_1}{f_1} = \frac{Y}{f_1 - Z} \quad (1)$$

Mittels Zoom-Objektiven können fovealisierte Abbildungseigenschaften (Abbildung 1 rechts) für unbewegte Objekte simuliert werden, indem ein Objekt mit einer großen und einer kleinen Brennweite aufgenommen wird. Damit werden die Probleme der Datenfusion umgangen, die bei unterschiedlichen optischen Achsen entstehen würden.

Bewegung und Tiefe eines Objekts lassen sich mit einer fovealisierten Kamera errechnen. In Abbildung 3 ist ein Objekt der Größe Y in den Abständen Z und $Z - d$ von der Kamera dargestellt (zur besseren Übersichtlichkeit einmal gespiegelt). Wird

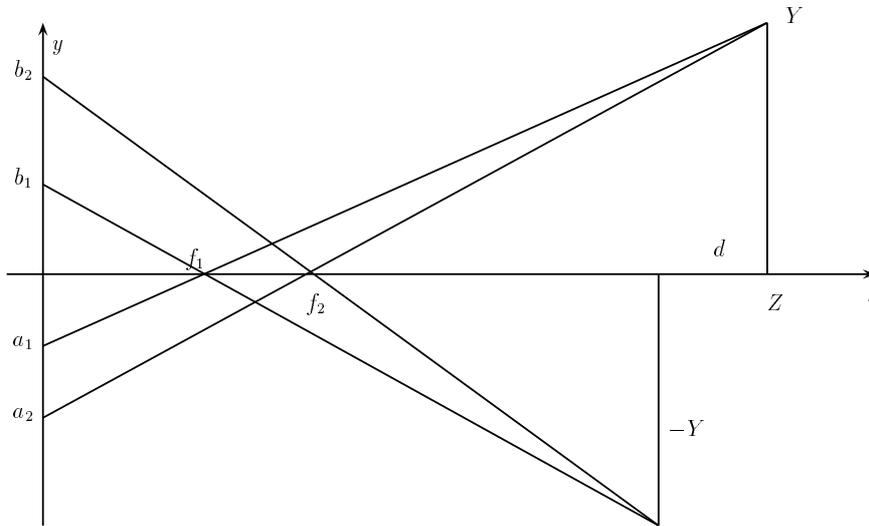


Abbildung 3: Bewegung und Tiefe mit einer Zoom Kamera.

angenommen, daß es sich um zwei Aufnahmen desselben Objekts zu zwei unterschiedlichen Zeitpunkten handelt, so ergibt sich ein Modell der Bewegung. Tabelle 1 faßt die Gleichungen zusammen. Ist die Korrespondenz der Punkte a_1, a_2, b_1, b_2 bekannt, so lassen sich bei bekannten Brennweiten f_1, f_2 aus den vier Gleichungen die drei Unbekannten Größen für Entfernung Z , Größe Y und Verschiebung d bestimmen. Zur Ermittlung der Entfernung eines unbekannten Objekts genügen (wie auch bei Stereoverfahren) zwei Bilder (a_1, a_2) , wobei zwei Gleichungen für zwei Unbekannte zu lösen sind. Vorausgesetzt ist die Kenntnis der Korrespondenz (a_1, a_2, b_1, b_2) und der Brennweiten (f_1, f_2) .

$\frac{a_1}{f_1} = \frac{Y}{f_1 - Z}$	$\frac{b_1}{f_1} = \frac{-Y}{f_1 - Z - d}$
$\frac{a_2}{f_2} = \frac{Y}{f_2 - Z}$	$\frac{b_2}{f_2} = \frac{-Y}{f_2 - Z - d}$

Tabelle 1: Gleichungen für Bewegung und Tiefe (nach Abbildung 3).

Das Verhältnis zweier Bildgrößen a_1 und a_2 eines Objekts im Abstand Z ergibt sich durch (2). Es ist unabhängig von der Größe des Objekts.

$$\frac{a_1}{a_2} = \frac{f_1(f_2 - Z)}{f_2(f_1 - Z)} \quad (2)$$

In realen Kameras muß nun noch die Diskretisierung durch den CCD-Chip in das Modell mit einbezogen werden. Nehmen wir typische Werte für Kleinbildkameras (Teleobjektiv 135mm, Weitwinkel 35mm, Bildgröße $36 \times 24 \text{ mm}^2$) bei einer Bildauflösung von 512^2 Bildpunkten an, so ergibt sich (3). Im Gegensatz zu (2) ist hier die Größe des Objekts von Bedeutung. Idealisierend wird angenommen, daß benachbarte CCD-Elemente voneinander unabhängig sind und daß die optische Achse die Bildebene im Zentrum schneidet.

$$i_1(Z, Y) = \left\lceil \frac{f_1 Y \cdot 256}{(f_1 - Z) \cdot 12} + 0.5 \right\rceil \quad i_2(Z, Y) = \left\lceil \frac{f_2 Y \cdot 256}{(f_2 - Z) \cdot 12} + 0.5 \right\rceil \quad (3)$$

Um aus zwei Bildpunkten a_1 und a_2 Rückschlüsse auf das Objekt ziehen zu können, müssen sie unterscheidbar sein. Die Pixeldifferenz δ zweier Bilder ist in (4) als

$$\delta(Z, Y) = i_1(Z, Y) - i_2(Z, Y) \quad (4)$$

angegeben und für ein Objekt fester Größe in Abbildung 4 graphisch dargestellt. Für ein Objekt der Größe 10 cm ergibt sich bei einer Entfernung von 10 m 30 Pixel Differenz. Diese Aussagen sind allerdings stark einzuschränken, da vor allem bei kleiner Brennweite die Fläche eines Pixels einer großen Fläche auf dem Objekt entspricht und die Zuordnung von a_1 zu a_2 damit nicht eindeutig möglich ist. Untersuchungen zur Fehleranfälligkeit bei der Trennung zweier benachbarter Kanten finden sich in [1].

Zur Berechnung räumlicher Information aus den Bildern ist es erforderlich, die internen Parameter der Kamera zu kennen (Brennweite, Schnitt der optischen Achse mit der Bildebene). Diese lassen sich durch Kamerakalibrierung ermitteln.

Basierend auf dem Lochkameramodell wird in [10] eine Stereokamera kalibriert. In Experimenten zeigt sich eine komplexe Abhängigkeit der Kameraparameter von der gewählten Blende. Die radialen Verzerrungen ändern sich ebenfalls mit der Brennweite und Blende. Beim aktiven Sehen werden Brennweite, Fokus und Blende gezielt variiert und somit laufend die Kameraparameter geändert. In der Regel ist eine Kalibrierung ein aufwendiges Verfahren. Es ist wünschenswert, Kameraparameter einfach schätzen zu können. Eine Möglichkeit für die Ermittlung des optischen Zentrums zeigt folgendes Lemma.

Lemma Für das Lochkameramodell einer Zoom-Kamera gilt (Abbildung 5):

Die Bilder (a_i, b_i) eines Objektpunkts (X, Y, Z) betrachtet mit Brennweiten f_i liegen auf einer Geraden durch den Nullpunkt, d.h. dem Schnittpunkt der optischen Achse mit der Bildebene.

Beweis Wir zeigen: Die Gerade $b = ma + c$, die durch zwei Punkte (a_1, b_1) und (a_2, b_2) , hat $c = 0$.

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = \frac{f_i}{Z - f_i} \begin{pmatrix} X \\ Y \end{pmatrix}$$

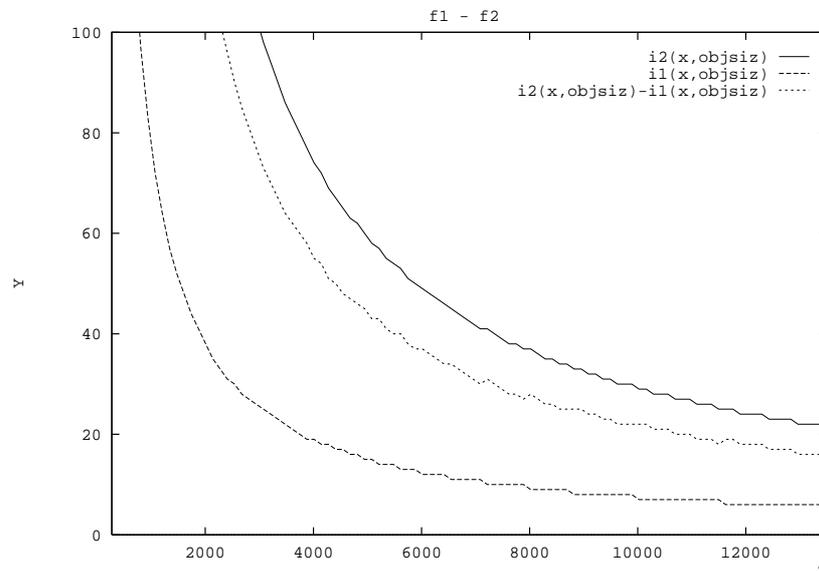


Abbildung 4: Abstand der Bildpunkte eines Objekts der Größe 10 cm auf einem Kleinbild bei zwei Brennweiten (35mm und 135mm).

Zwei Gleichungen für die Gerade durch zwei Punkte:

$$b_i = ma_i + c; \quad c = b_1 - \frac{b_2 - b_1}{a_2 - a_1} a_1$$

Umformung und Einsetzen der Lochkameragleichungen (Tabelle 1) ergibt:

$$c = \frac{Y f_1}{z - f_1} - \frac{\frac{f_2}{z - f_2} Y - \frac{f_1}{z - f_1} Y}{\frac{f_2}{z - f_2} X - \frac{f_1}{z - f_1} X} \frac{f_1}{z - f_1} X = \frac{Y f_1}{z - f_1} - \frac{Y}{X} \frac{f_1 X}{z - f_1} = 0$$

q.e.d.

Corollar Alle Geraden durch korrespondierende Bildpunkte unter zwei Brennweiten schneiden sich in einem Punkt in der Bildebene – dem optischen Zentrum.

Werden also die Bilder eines Punktpaars bei der Veränderung der Brennweite eines Zoomobjektivs verfolgt, so ergibt sich das optische Zentrum aus dem Schnittpunkt der beiden Geraden durch diese Punkte. Wegen der Diskretisierungs- und Segmentierungsfehler werden in der Praxis Ausgleichsgeraden durch mehrere Punkte verwendet.

5 Aktives Sehen mit Zoom-Kameras

In diesem Kapitel beschreiben wir erste Ergebnisse zur Polyedervermessung mit aktivem Sehen. Mit einer der beiden Überblickkameras (Abbildung 2) wurde ein ru-

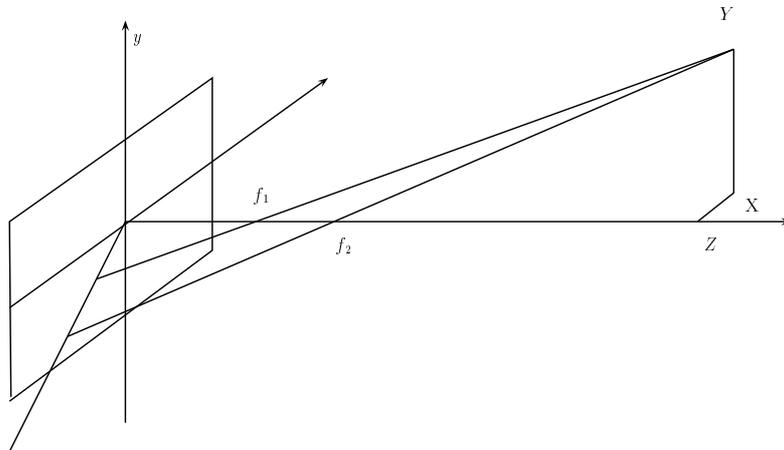


Abbildung 5: Modell der Zoom Kamera und Erläuterung des Lemmas.

hendes polyedrisches Objekt durch motorische gleichmäßige Veränderung des Zoom-Objektivs mit 30 verschiedenen Brennweiten zwischen 9.5 mm und 152 mm aufgenommen (Abbildung 6).

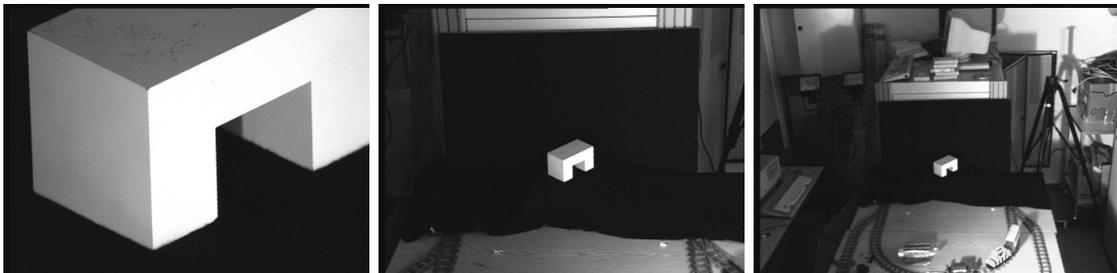


Abbildung 6: Polyeder Bilder (640×480) 01,20 und 29.

In realen Bildern ist die theoretisch ausreichende Differenz von einem Pixel zur Unterscheidung eines Objekts (4) nicht erreichbar, da viele Objektpunkte zur Beleuchtung eines Pixels beitragen. An den Objektgrenzen kommt es zu Streuungen; es ist daher nicht stabil festzustellen, welche Punkte im hochaufgelösten Bild mit Punkten im niedrig aufgelösten Bild korrespondieren. Anstelle von Pixelkorrespondenzen ist es sinnvoll, Merkmale zuzuordnen, die stabil erkannt werden können und zudem nur mäßige Unterschiede in der Brennweite zu verwenden.

Im folgenden werden Vertices, die als Schnittpunkte von Geraden entstehen, als Merkmale verwendet. Da die Geraden aus einer Linienapproximation durch zahlreiche Bildpunkte entstehen, ist dieses Merkmal hinreichend stabil. Die Kantenbilder (Abbildung 7) wurden auf die Kanten mit senkrechter Orientierung eingeschränkt. Eine Linienverfolgung mit Hysterese-Schwellwertverfahren [5], anschließender Gera-

denapproximation und Vertexdetektion liefert die Ergebnisse in Abbildung 8. Bilder, Kanten, Linien und Vertices sind Objekte einer objektorientierten Umgebung für die Bildverarbeitung; die verwendeten Programme stammen aus ANIMALS [9].

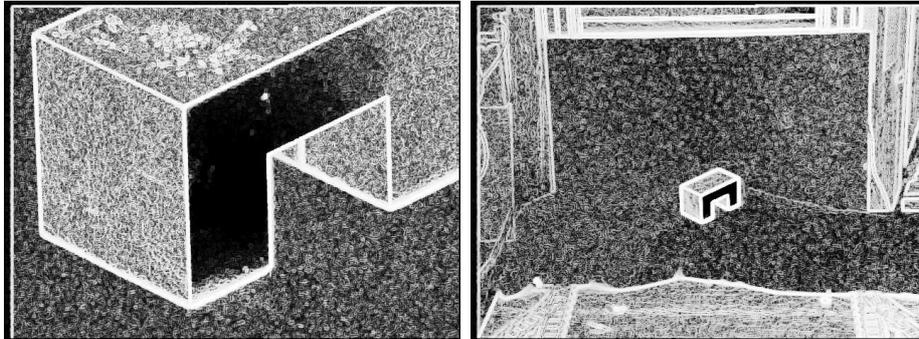


Abbildung 7: Kantenstärkebilder 01 und 20.

In Abbildung 9 ist die Verfolgung von einzelnen Endpunkten sowie Ausgleichsgeraden durch die gefundenen Punkte dargestellt. Der Schätzwert für den Schnittpunkt dieser Geraden ergibt nach dem Lemma als Wert für das optische Zentrum die Koordinaten (319.3, 297.8).

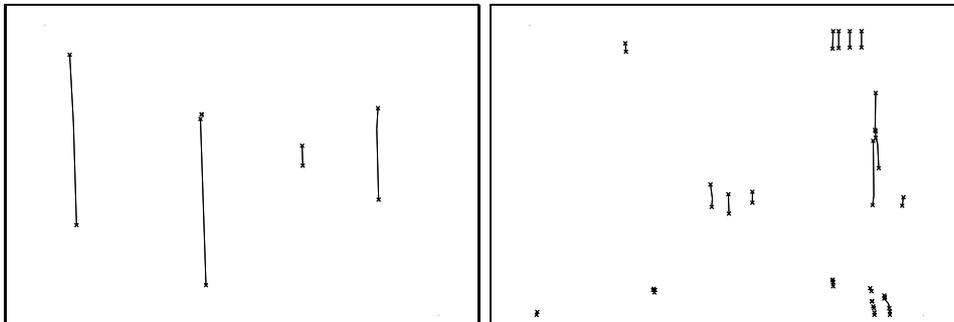


Abbildung 8: Fast senkrechte Geraden in den Bildern 01 und 20.

In den Bildern 21–29 sind die Kanten des Polyeders wegen seiner geringen Größe im Bild nicht mehr ohne Vorwissen stabil zu detektieren. Abbildung 10 links zeigt die X-Koordinate aller Linienenden, die im Bereich 250–400 liegen für alle Bilder der Folge. Für eine Linie ist eine Approximation als Parabel angegeben, die für die Bilder 1–20 eine befriedigende Näherung der Punkte liefert. Für einzelne Einstellungen der Kamera, die vom Kameramotor an den Rechner gemeldet werden, wird die Brennweite durch Kalibrierung ermittelt. Für Zwischenwerte der Einstellung ergibt sich die Brennweite durch Interpolation. Abbildung 10 rechts zeigt die Verfolgung einer Linie in den Bildern 1–20.

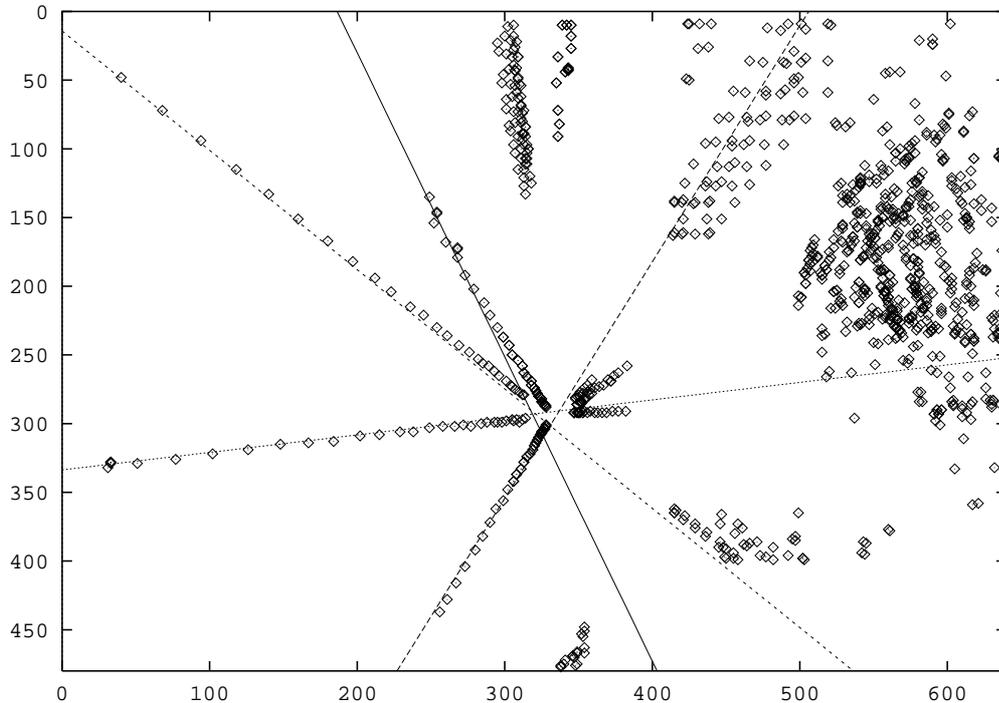


Abbildung 9: Plot der Endpunkte der Linien.

Nun ist es möglich, mittels bekannter Bildpunkte (a_i, b_i) , die mit bekannten Brennweiten f_i erzeugt wurden, die vollständigen 3D-Koordinaten des Objektpunktes zu errechnen. In Abbildung 11 ist die ermittelte Entfernung eines Objektpunktes in ca. 3000 mm Entfernung gegenüber den beiden Brennweiten f_1 und f_2 aufgetragen. Dabei ist zu erkennen, daß bereits kleine Fehler in der Bestimmung der f_i zu großen Fehlern in der berechneten Entfernung z führen. Dasselbe gilt für Fehler in der Bestimmung der (a_i, b_i) . Dieses Ergebnis ist mit Erkenntnissen aus dem Stereosehen vergleichbar, in dem bei einer sehr kleinen Stereobasis bereits geringe Fehler in der Disparitätsberechnung zu großen Fehlern in der Tiefenberechnung führen [10]. Dies führt im Zusammenhang mit am Anfang des Abschnitts besprochenen Schwierigkeiten bei großen Brennweitenunterschieden zu einer “Tradeoff”-Situation. In unseren Versuchen konnten die Brennweiten bisher nicht genügend genau bestimmt werden. Auch unter der Annahme der linearen Brennweitenveränderung zwischen zwei Bildern (Abbildung 10, Bilder 1–10), mittels welcher aus drei Aufnahmen (a_i, b_i) eines Objektpunktes bei bekannter Anfangsbrennweite dessen 3D-Koordinaten berechnet werden können, genügt die Genauigkeit der Bildpunktbestimmung nicht, um bei einer Entfernung von drei Metern ein Objekt zu vermessen.

Als erstes Ergebnis unserer Untersuchungen läßt sich festhalten, daß zur Bestimmung von Tiefe durch Punktkorrespondenzen sowohl an die Kamerakalibrierung als

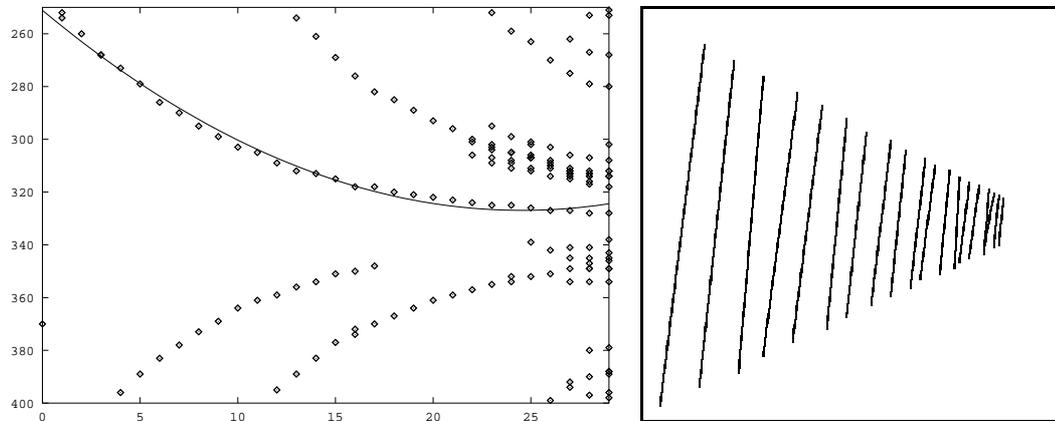


Abbildung 10: Automatische Linienverfolgung in Bild 1 bis 29. Links: X-Werte der Endpunkte der Linien. Rechts: eine der Linien selbst.

auch an die Bestimmung der 2D Bildkoordinaten höchste Genauigkeitsanforderungen gestellt werden müssen, die in diesem Stadium unserer Arbeit noch nicht erreicht sind. Die stabil detektierte Folge der Endpunkten (Abbildung 10 rechts) verspricht allerdings zum einen gute Merkmale in Subpixelgenauigkeit. Zum anderen soll im Zuge eines Kreislaufs von Aktorik und Sensorik auf exakte Tiefenbestimmung verzichtet werden.

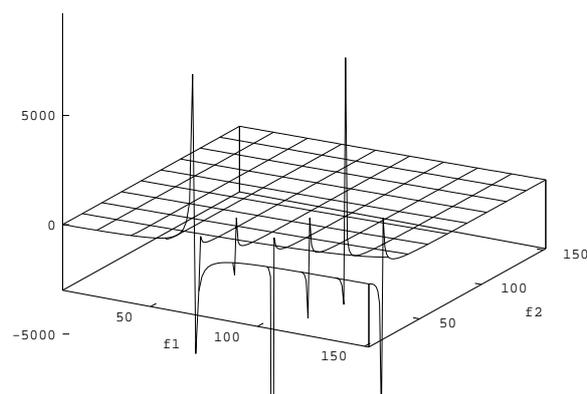


Abbildung 11: Entfernung z in Abhängigkeit der Brennweiten f_1 und f_2 (jeweils in mm). Reale Entfernung: 3000 mm.

6 Zusammenfassung und Ausblick

In unserem Beitrag haben wir dargestellt, daß mittels Zoomkameras die 3D–Vermessung eines Objekts möglich ist. Wir konnten zeigen, daß ebenfalls eine Berechnung des optischen Zentrum durch aktives Sehen durchgeführt werden kann. In ersten praktischen Untersuchungen stellte sich heraus, daß sowohl an die Berechnung der Brennweiten als auch der Bildpunkte höchste Genauigkeitsanforderungen gestellt werden müssen.

Die Brennweitereinstellungen der Kamera können mit einer hohen Genauigkeit wiederholt angefahren werden. Somit genügt es, diese Brennweiten einmal zu bestimmen und bei der folgenden Berechnung der 3D–Koordinaten als bekannt vorauszusetzen (siehe Tabelle 1). In unsere weiteren Untersuchungen werden wir Subpixelgenauigkeit und schritthaltende Kalibrierung mit einbeziehen. Das Prinzip des aktives Sehen werden wir durch die Einbeziehung zusätzlicher Kameras (z.B. Roboterkamera) zur aktiven Vermessung verfolgen.

Literatur

- [1] R. Bajcsy: *Active Perception*, *Proceedings of the IEEE*, Bd. 76, Nr. 8, 1988, S. 996–1005.
- [2] R. Bajcsy, M. Campos: *Active and Exploratory Perception*, *Computer Vision, Graphics and Image Processing*, Bd. 56, Nr. 1, 1992, S. 31–40.
- [3] A. Blake, A. Yuille (Hrsg.): *Active Vision*, MIT Press, Cambridge, Mass., 1992.
- [4] C. Brown: *Issues in Selective Perception*, in *Proceedings of International Conference on Pattern Recognition*, Den Haag, 1992, S. 21–30.
- [5] H. Brüning: *Konzeption und Realisierung einer flexiblen Bildsegmentierung*, Dissertation, Technische Fakultät, Universität Erlangen–Nürnberg, Erlangen, 1990.
- [6] D. W. Jacobs: *Recognizing 3D Objects using 2D Images*, PhD thesis, Department of Electrical Engineering and Computer Science, MIT, Massachusetts, 1992.
- [7] E. Krotkov: *Active Computer Vision by Cooperative Focus and Stereo*, Springer, Berlin, 1989.
- [8] D. Marr: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman, San Francisco, 1982.
- [9] D. W. R. Paulus: *Objektorientierte und wissensbasierte Bildverarbeitung*, Vieweg, Braunschweig, 1992.
- [10] S. Posch: *Automatische Bestimmung von Tiefeninformation aus Grauwert-Stereobildern*, Deutscher Universitäts Verlag, Wiesbaden, 1990.
- [11] M. J. Swain, M. Stricker: *Promising Directions in Active Vision*, Tec. Rep. CS 91-27, University of Chicago, November 1991.
- [12] S. Vinther, R. Cipolla: *Toward 3D Object Model Acquisition and Recognition using 3D Affine Invariants*, in *4th British Machine Vision Conference*, erscheint 1993.