

Rekonstruktion von 3D-Koordinaten aus Bildern

Reinhard OLDENBURG, Heidelberg

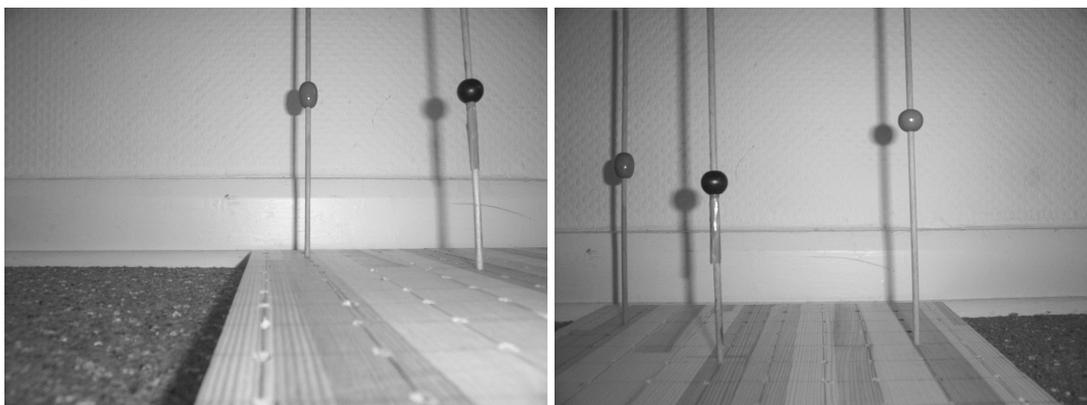
In diesem Aufsatz wird in einem verhältnismäßig einfachen Fall gezeigt, wie aus zwei Photographien einer Szene die 3D-Koordinaten von Punkten und damit die wahren Längen aller Strecken berechnet werden können.

Das Problem

Unendlich viele Aufgaben der analytischen Geometrie beginnen damit, dass die Koordinaten von Punkten im Raum gegeben sind. Von diesen Startpunkten aus werden dann Abstände, Winkel, Punkte, Kugeln und vieles mehr berechnet. Aber wie kommt man zu den 3D-Koordinaten?

Eine auch in der Praxis wichtige Methode besteht darin, Bilder des interessierenden Objektes aus verschiedenen Perspektiven aufzunehmen und daraus die Koordinaten zu berechnen.

Zwei Beispiele für solche Bildpaare:



In dieser Modellsituation hat die zweite Kugel von links im linken Bild die Pixelkoordinaten $(1101,198)$ und im rechten Bild $(378,416)$. Die Kamera wurde um 18cm nach rechts verschoben (ohne Drehung).



Die linke, obere Ecke des Hochhauses hat links die Pixelkoordinaten $(554,195)$ und rechts $(472,241)$. Die Kamera wurde um 3m senkrecht zu ihrer waagrechten optischen Achse verschoben.

Die wesentlichen Schritte zur Lösung werden im folgenden beschrieben.

Schritt 1: Das Lochkameramodell

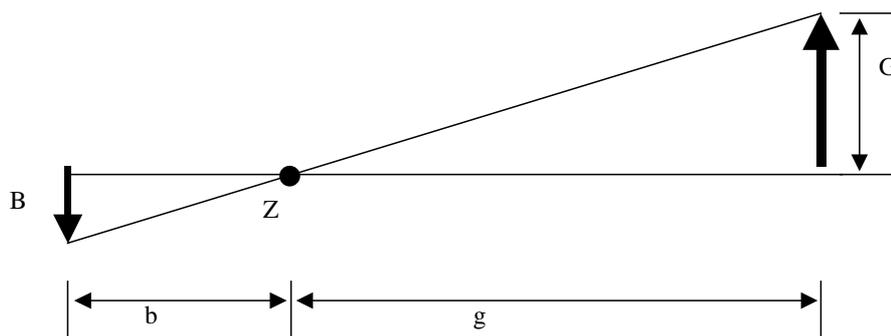
In der Computergrafik werden Bilder oft so berechnet, als würden sie mit einer Lochkamera aufgenommen. Für sie sind die Abbildungsgleichungen nämlich besonders einfach.

Lochkamera und Linsenkamera

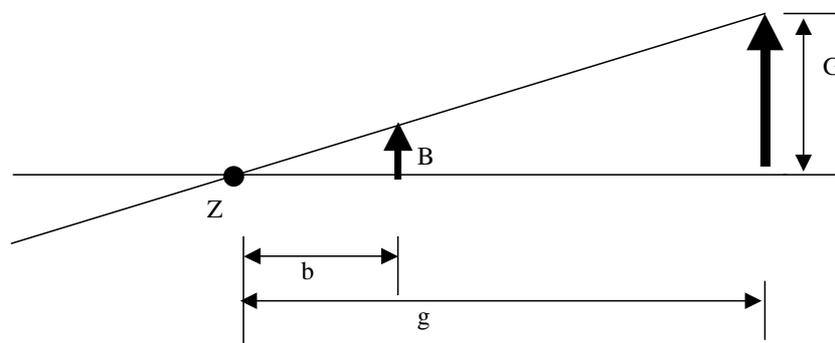
Bei der Lochkamera gilt zwischen Gegenstandsgröße G , der Bildgröße B , der Gegenstandsweite g und der Bildweite b die Gleichung $\frac{G}{g} = \frac{B}{b}$ (Strahlensatz, vgl. auch die Skizze unten).

Bei der Abbildung mit einer Linse gilt die gleiche Beziehung, denn es gibt den ungebrochenen Mittelpunktstrahl. Das Bild ist aber nur scharf, wenn zusätzlich die Linsengleichung $\frac{1}{f} = \frac{1}{b} + \frac{1}{g}$ mit der Brennweite f der Linse erfüllt ist. Bei allen üblichen Aufnahmen mit einer normalen Kamera ist die Gegenstandsweite g viel größer als f und b (beide liegen im Kameragehäuse!). Der Summand $\frac{1}{g}$ ist also im Vergleich

sehr klein und kann vernachlässigt werden: $\frac{1}{f} \approx \frac{1}{b}$, also $b \approx f$. Man kann also mit dem Lochkameramodell arbeiten, wenn man als Bildweite b die Brennweite verwendet.

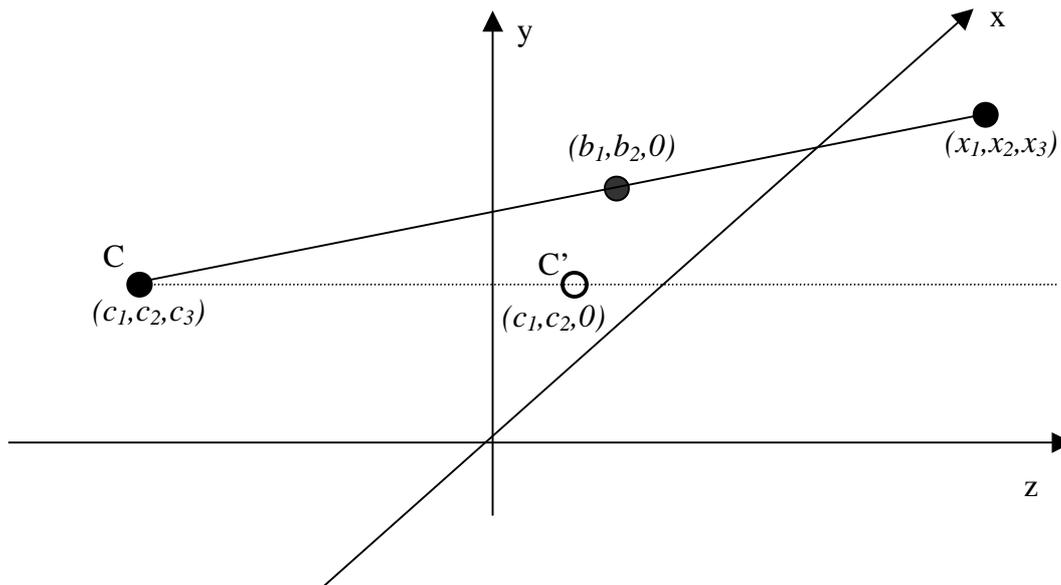


Außerdem ist es für die Mathematik (sprich den Strahlensatz) egal, ob man das Abbildungszentrum Z wie bei der realen Lochkamera zwischen Bild und Gegenstand hat, oder davor:



Schritt 2: Abbildungsgleichung

Das Lochkameramodell erlaubt die Berechnung des Bildes eines Objektes. Besonders einfach geht das, wenn die Bildebene eine Koordinatenebene ist. Wir legen im folgenden die Koordinatenachse z als Blickachse der Kamera nach rechts.



Die Gerade durch den Objektpunkt $X = (x_1, x_2, x_3)$ und das Projektionszentrum $C = (c_1, c_2, c_3)$ kann parametrisiert werden als $\vec{p}(\lambda) = \vec{x} + \lambda \cdot (\vec{c} - \vec{x})$. Der Schnittpunkt $B = (b_1, b_2, 0)$ mit der x-y-Ebene ($z = 0$) erfüllt die Beziehung

$$\begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \lambda \cdot \begin{pmatrix} c_1 - x_1 \\ c_2 - x_2 \\ c_3 - x_3 \end{pmatrix}. \text{ Aus der letzten Komponente bestimmt man}$$

$$\lambda = \frac{-x_3}{c_3 - x_3}. \text{ Also gilt:}$$

$$b_1 = x_1 - \frac{x_3}{c_3 - x_3} \cdot (c_1 - x_1), \quad b_2 = x_2 - \frac{x_3}{c_3 - x_3} \cdot (c_2 - x_2).$$

Alternativ gewinnt man diese Beziehungen aus einer Überlegung mit dem Strahlensatz.

Die Größen b_1 und b_2 sind nicht direkt der Beobachtung zugänglich (sie wären direkt auf dem CCD-Chip zu messen) und müssen daher im nächsten Schritt mit den Pixelkoordinaten in Verbindung gebracht werden.

„Projektionszentrum“ und „Bildebene“ sind mathematische Begriffe. Wie findet man die zugehörigen Koordinaten nun in der Realität? Zentrum und Bildebene liegen innerhalb der Kamera. Solange man diese also nicht aufschraubt, kann man nicht exakt sagen, wo Bildebene und Projektionszentrum liegen. Die Unsicherheit ist aber für alle realen Situationen viel kleiner als die Entfernungen zum Objekt, so dass diese Ungenauigkeit in Kauf genommen werden kann.

Nach den (Näherungs-)Überlegungen im ersten Schritt ist die Brennweite (näherungsweise) gleich dem Abstand des Projektionszentrums zur Bildebene, also $c_3 = -f$. Die Koordinaten c_1 und c_2 sind die x- und y-Koordinaten des Kameramittelpunktes zum Zeitpunkt der Aufnahme.

Die Brennweite muss für die verwendete Kamera bekannt sein. Dabei gibt es zwei Stolperfallen: Bei Digitalkameras wird oft nicht die reale Brennweite angegeben, sondern eine „äquivalente“, die dem Vergleich mit einer Kleinbildkamera dient. Die realen Brennweiten liegen meist unter einem Zentimeter. Auf den Internetseiten der Hersteller sind die realen Brennweiten in der Regel zu finden - wenn auch nicht immer auf den ersten Blick. Bei Zoomkameras muss man noch eine zweite

Stolperfalle umgehen: Man sollte immer den maximalen oder den minimalen Zoom verwenden, da man sonst keine Kontrolle über die Brennweite hat.

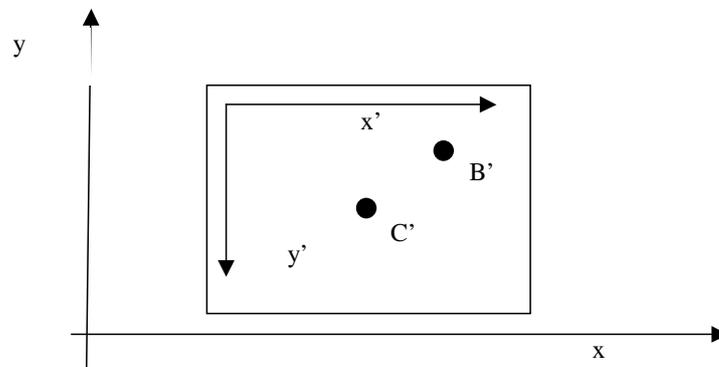
Bei der von mir verwendeten Kodak DX3215 beträgt die Brennweite 30-60mm (equiv), bzw. real 4,5mm bis 9mm. Für alle Aufnahmen wurde die Kamera möglichst weitwinklig ($f = 4,5 \text{ mm}$) eingestellt.

Schritt 3: Digitale Bilder: Pixel und Koordinaten

Bei einer Digitalkamera ist die Bildebene durch den CCD-Chip gegeben. Auf ihm wird die optische Abbildung hergestellt. Positionen auf ihm können einerseits in Pixelkoordinaten, andererseits in Weltkoordinaten (zum Zeitpunkt der Aufnahme) angegeben werden. Jetzt geht es um die Umrechnung zwischen diesen Größen.

Es seien (x', y') die Pixelkoordinaten in einem Bild mit B Pixel Breite und H Pixel Höhe. Das Pixelkoordinatensystem startet nach üblichen Konventionen links oben mit $(0, 0)$. In der Bildmitte liegt der Lotfußpunkt des Projektionszentrums C' . In

Realweltkoordinaten ist er $(c_1, c_2, 0)$, in Pixelkoordinaten ist er $\left(\frac{B}{2}, \frac{H}{2}\right)$.



Die einzelnen Pixel auf dem CCD-Chip besitzen eine Breite s_x und eine Höhe s_y .

Die Realweltkoordinaten eines Bildpunktes $B' = (b_1, b_2, 0)$ berechnen sich dann als

$$b_1 = c_1 + s_x \cdot \left(x' - \frac{B}{2}\right); \quad b_2 = c_2 + s_y \cdot \left(\frac{H}{2} - y'\right).$$

Um die Größe eines Pixels des Chips zu berechnen, muss man sich im Internet umtun. Dabei ist zu beachten, dass die Größenangaben für CCD-Chips immer die Diagonale in Zoll (1 Zoll = 2,54 cm) messen. Und damit es nicht zu einfach wird, gibt es die Konvention, dass der reale Chip immer nur 75% der angegebenen Größe hat (Das erleichtert - so die Argumentation - uns Verbrauchern den Vergleich mit antiken Videoröhren.) Die von mir verwendete Kodak DX3215 besitzt einen CCD-Chip von 1/2.5" Diagonale. Die reale Diagonale ist also

$d = 0,75 \cdot \frac{1}{2,5} \text{ inch} = 0,75 \cdot 0,4 \cdot 2,54 \text{ cm} = 0,762 \text{ cm}$. Der Chip besitzt $B = 1280$ mal $H = 960$ Pixel Auflösung. Breite x und Höhe y des CCD-Chips ergeben sich aus der Lösung der Gleichungen $x^2 + y^2 = d^2$ und $\frac{x}{y} = \frac{B}{H}$. Bei der Kodak DX3215 besitzt der

CCD-Chip demnach die Abmessungen von 0,61 cm Breite und 0,46 cm Höhe.

Chip und Brennweite einer Digitalkamera sind also ausgesprochen klein. Dies ist auch der Grund für die phantastische Tiefenschärfe: Eine Digitalkamera ist eben fast eine Lochkamera.

Schritt 4: Gleichungen lösen

Durch Gleichsetzen können die unzugänglichen Größen b_1 , b_2 aus den Gleichungssystemen der beiden vorhergehenden Schritte eliminiert werden. Es bleibt ein System aus zwei Gleichungen in folgenden Variablen:

x, y, z gesucht
 x', y' bekannte Pixelkoordinaten
 c_1, c_2 bekannt aus Position der Kamera
 c_3 Brennweite der Kamera
 s_x, s_y Pixelmaßstab

Jeder Punkt in einem Bild liefert also zwei Gleichungen für seine drei Unbekannten x, y, z . Zwei Bilder aus verschiedenen Perspektiven liefern aber schon vier Gleichungen - das System wird überbestimmt.

Überbestimmte Gleichungssysteme treten häufig auf, wenn Größen aus Messungen bestimmt werden. Dann gibt es in der Regel keine echte Lösung mehr, die alle Gleichungen erfüllt, aber man kann einen optimalen Ersatz suchen. Das geht so: Aus einem System von m Gleichungen $g_i(x) = 0, i = 1, \dots, m, x \in R^n$ bildet man die

Funktion $f(x) = \sum_{i=1}^m g_i(x)^2$ und sucht ihr Minimum. Wenn die Minimalstelle x eine

Lösung des Systems gibt, minimiert sie die Funktion zu 0. Wenn man umgekehrt ein x findet, das den minimalen Wert 0 ergibt, hat man tatsächlich eine Lösung des gesamten Gleichungssystems. Bei überbestimmten, unlösbaren Systemen kann f nicht den Wert 0 annehmen, aber durch das Minimieren von f macht man die Abweichung von dem, was die Gleichungen vorschreiben, so klein wie möglich. Mehr Information findet sich in (Oldenburg 2005).

Das Verfahren im Pseudocode

Alle Schritte zu einem Programm in Pseudocode destilliert zeigen den Gang der Auswertung:

Eingaben:

- Zwei Bilder B_L, B_R aufgenommen mit Kameraachse in z-Richtung. Die Kamera darf zwischen beiden Bildern nur parallel verschoben, nicht gedreht worden sein.
- Die x-y-Weltkoordinaten der Kameraposition für die beiden Bilder (c_{L1}, c_{L2}) und (c_{R1}, c_{R2})
- Die Pixelkoordinaten (x'_L, y'_L) und (x'_R, y'_R) ein und desselben Objektpunktes P in beiden Bildern.
- Auflösung der Kamera: B Pixel Breite, H Pixel Höhe
- Die reale Größe eines Pixels der Kamera auf dem CCD-Chip s_x, s_y
- Brennweite der Kamera f

Berechnung: Berechne im Least-Square-Sinne eine optimale Näherungslösung (x_1, x_2, x_3) des folgenden Gleichungssystems

$$x_1 + \frac{x_3}{f + x_3} \cdot (c_{L1} - x_1) = c_{L1} + s_x \cdot \left(x'_L - \frac{B}{2} \right)$$

$$x_2 + \frac{x_3}{f + x_3} \cdot (c_{L2} - x_2) = c_{L2} + s_y \cdot \left(\frac{H}{2} - y'_L \right)$$

$$x_1 + \frac{x_3}{f + x_3} \cdot (c_{R1} - x_1) = c_{R1} + s_x \cdot \left(x'_R - \frac{B}{2} \right)$$

$$x_2 + \frac{x_3}{f + x_3} \cdot (c_{R2} - x_2) = c_{R2} + s_y \cdot \left(\frac{H}{2} - y'_R \right)$$

Ausgabe: Weltkoordinaten (x_1, x_2, x_3) des Punktes P

Die Berechnung der optimalen Näherungslösung wird im Beitrag zur numerischen Optimierung (in diesem Band) erläutert.

Die Rekonstruktion der Koordinaten der linken Hochhausecke aus dem einleitenden Bilderpaar ergibt, dass das Hochhaus fast 34,6 m entfernt ist, und 9,6 m über den Standpunkt der Kamera in den Himmel ragt. Das sind klare Fakten, die man mit anderen Methoden natürlich auch prüfen kann.

Die MuPAD-Skripte, die das Verfahren umsetzen, sind über die Adresse www.ph-heidelberg.de/wp/filler/istron9 erhältlich. Auch Derive kann verwendet werden, wenn man die nötigen numerischen Optimierungsprozeduren selbst schreibt.

Reale Anwendungen

Die Bestimmung von 3D-Koordinaten ist in vielen Anwendungsfeldern wichtig und entsprechend gibt es viele Techniken. Eine wissenschaftliche Arbeit, die einen ähnlichen Ansatz wie dieser Aufsatz verwendet ist (Rocchini 2001). Deren 3D-Scanner bewegt eine Digitalkamera mit Schrittmotoren.

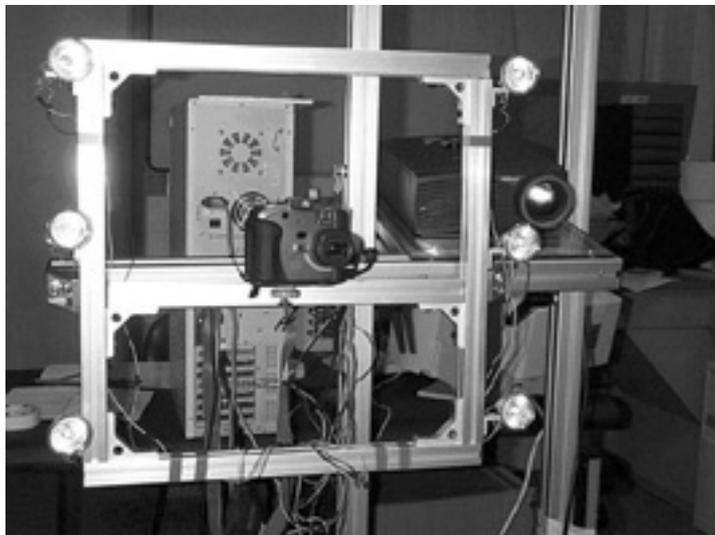


Figure 1: The low cost scanner developed at CNR, which uses structured light and consumer electronic technology.

Für die Vermessung von Oberflächen, z.B. in der Automobilindustrie, werden laser-gestützte Systeme eingesetzt oder Arme mit mehreren Gelenken, die mit Sensoren ausgestattet sind.

Eine interessante astronomische Anwendung ist die STEREO-Mission der NASA (<http://stp.gsfc.nasa.gov/missions/stereo/stereo.htm>): Von einem Paar von Sonden aus, die gleichzeitig in verschiedene Richtungen von der Erde aus starten werden, sollen ab Februar 2006 Bilder der Sonne zur Erde gefunkt werden. Aus den Bildpaaren können z.B. die genauen Abmessungen und damit das Volumen von Sonnenausbrüchen berechnet werden.

Es gibt eine ganze Reihe von Verallgemeinerungen der oben untersuchten Fragestellung: Die Kameraparameter könnten unbekannt sein, die Kameraverschiebung könnte unbekannt sein und u.U. auch eine Drehung beinhalten. Die Theorie zu solchen Problemstellungen bietet beispielsweise das Buch (Ma et al. 2004).

Natürlich müssen professionelle Programme zur 3D-Rekonstruktion (z.B. www.photomodeler.com) dies alles beherrschen, zusammen mit einer automati-

sehen Erkennung der Punktpaare. Der Benutzer eines solchen Programms steuert mit seinen Mausklicks ein gigantisches mathematisches Werk, dessen Grundsteine in diesem Aufsatz beleuchtet wurden.

Literatur

Y. Ma, S. Soatto, J. Kosecka und S. Sastry [2004]: An Invitation to 3D-Vision - From Images to Geometric Models, Springer.

R. Oldenburg [2005]: Numerische Optimierung - ein schneller Weg zu komplexer Modellbildung. In diesem Band.

C. Rocchini, R. Scopigno [2001]: A Low-cost Optical 3D Scanner, ERCIM News No.44.

O. Schreer [2005]: Stereoanalyse und Bildsynthese, Springer.

Adresse des Autors:

Prof. Dr. Reinhard Oldenburg, Pädagogische Hochschule Heidelberg, Im Neuenheimer Feld 561

E-mail: oldenburg@ph-heidelberg.de