Schätzung der Pose einer Kamera mithilfe von Landmarken

Estimating the Pose of a Camera Using Several Landmarks in a Room

H. Janocha und F. Königstein, Universität des Saarlandes

Manuskripteingang: 19. November 2002; zur Veröffentlichung angenommen: 29. Juli 2003.

Für industrielle Vermessungsaufgaben wird zunehmend Bildverarbeitung eingesetzt. Beispielsweise kann hiermit auch die Poseabweichung zwischen Ist- und Sollwert bei Roboter-Endeffektoren verbessert werden. In dieser Arbeit wird ein System vorgestellt, welches die Pose einer Kamera und zusätzlich die 3D-Koordinaten von bestimmten Objektpunkten schätzt. Als Eingabedaten verwendet das System mehrere Bilder, welche Landmarken bei Betrachtung von verschiedenen Kameraposen enthalten. Die Schätzung der Pose wird mittels nichtlinearer Optimierung durchgeführt. Nach lokaler Linearisierung wird eine Kalman-Schätzung in Kovarianzform durchgeführt. Simulations- und Messergebnisse belegen die Leistungsfähigkeit des Verfahrens.

Image processing is increasingly used for industrial measuring tasks. In this paper we will introduce a system which can estimate the pose of a camera as well as the 3D coordinates of particular object points. For input data the system uses several images containing landmarks taken from different camera poses. A nonlinear optimization routine is applied. Following local linearization a Kalman estimation in covariance form is performed. The result of the simulation and measurements prove the system's efficiency.

Schlagwörter: Pose-Schätzung, Photogrammetrie, CCD-Kamera, Landmarke

Keywords: Pose estimation, photogrammetry, CCD camera, landmark

1 Einleitung

Bildverarbeitung wird zunehmend für industrielle Vermessungsaufgaben eingesetzt. Zum einen können mit geeigneter Software Werkstücke vermessen werden, zum anderen ist auch eine Bestimmung der Pose (Position und Orientierung) der Kamera möglich. Eine interessante Anwendung besteht in der Verbesserung der erreichbaren Genauigkeit eines Industrieroboters in der Position und Orientierung des Endeffektors. Wenn am Endeffektor außer einem Werkzeug noch eine Kamera montiert wird, kann aus der Pose der Kamera auf die Pose des Werkstücks geschlossen werden. Für die Posebestimmung der Kamera sind insbesondere zwei Verfahren von Bedeutung: die Verfolgung von Kamerabewegungen durch Auswertung des optischen Flusses im Bild [5] sowie die Posebestimmung durch Auswertung der Positionskoordinaten bestimmter Landmarken im Bild. Letztere Methode wird in diesem Beitrag diskutiert. Dabei werden nicht nur einige Aspekte der zugrunde liegenden Theorie dargestellt, sondern auch praxisbezogene Fragestellungen angesprochen, z. B. die Umrechnung von einem zur Kamera lokalen Koordinatensystem in ein anderes Koordinatensystem, welches lokal zum Endeffektor ist. Weiter zeigen Simulations- und Messergebnisse die Funktionsfähigkeit der entwickelten Software auf. Es werden Hinweise gegeben, an welchen Stellen Verbesserungen möglich sind, d. h. die Messunsicherheit der Poseschätzungen verringert werden kann.

2 Schätzung von Pose und Objektpunktkoordinaten

Die Pose einer Digitalkamera soll anhand von Landmarken geschätzt werden, d. h. mithilfe von Objekten, denen jeweils ein definierter Objektpunkt zugeordnet wird. Z. B.

stellt eine Kreisfläche eine Landmarke dar, deren Mittelpunkt als der Objektpunkt definiert wird. Diese Landmarken müssen im Kamerabild sichtbar sein, und die Korrespondenzen zwischen den 3D-Landmarken und denen, die im Bild erscheinen, müssen feststellbar sein. Es soll möglich sein, dem Schätzverfahren als A-priori-Information lediglich grobe Werte für die Objektpunktkoordinaten und für die Pose der Kamera bei Aufnahme jedes neuen Bildes vorzugeben.

Die Abbildungsgleichungen, mit denen aus vorgegebenen Objektpunkten und aus einer Kamerapose die Bildpunktkoordinaten berechnet werden können, sind nichtlinear, und zwar sowohl bezüglich der Kamerapose als auch der Objektpunktkoordinaten. Nach Linearisierung der Gleichung wird eine Kalman-Schätzung der Kamerapose und der 3D-Objektpunktkoordinaten vorgenommen. Die Qualität der Schätzung hängt in besonderem Maße von der Wahl des Linearisierungspunkts der Kamerapose ab. Um diesen und auch die Linearisierungspunkte der Objektpunktkoordinaten geeignet zu bestimmen, wird wie folgt verfahren:

1. Zunächst werden die A-priori-Schätzungen der Objektpunktkoordinaten als exakt betrachtet und die Kamerapose wird berechnet. Dabei findet das Levenberg-Marquardt-Verfahren zur nichtlinearen Ausgleichsrechnung Anwendung. Die hierbei gefundene Pose ist i. Allg. deutlich besser als die a priori vorgegebene Pose und somit als Linearisierungspunkt für die folgenden Rechnungen geeignet.
2. Nun werden in einem iterativen Verfahren wiederholt sowohl neue Linearisierungspunkte der Pose als auch der Objektpunktkoordinaten berechnet. Dabei wird mithilfe der aktuellen linearisierten Abbildungsgleichungen jeweils eine neue Kalman-Schätzung vorgenommen. Es soll eine Verlustvariable minimiert werden. Falls diese aufgrund der durch die Linearisierung entstandenen Fehler größer wird, muss durch Einführung künstlicher Abbildungsgleichungen (s. u.) eine zu starke Abweichung der Schätzung vom aktuellen Linearisierungspunkt verhindert werden.
   In jedem Iterationsschritt wird eine untergeordnete Iteration ausgeführt, um sicherzustellen, dass die neuen Objektpunkte bestimmte zwischen ihnen geltende Abstandsvorgaben einhalten, die a priori bekannt sind.

Die in Punkt 1 beschriebene Grob-Schätzung der Pose kann mit relativ wenig Rechenaufwand erfolgen, weil hier nur sechs unbekannte Parameter vorliegen, nämlich drei Koordinaten für die Position der Kamera und drei reelle Werte für deren Orientierung. In Punkt 2 ist von a priori bekannten Abständen zwischen Objektpunkten die Rede. Wie später noch erklärt wird, ist es sinnvoll, wenn man die 3D-Abstände zumindest zwischen wenigen Objektpunkten mit anderen Methoden als der Bildverarbeitung relativ genau vermisst und diese Information dem auf Bildverarbeitung basierenden Schätzalgorithmus zur Verfügung stellt.

Im Folgenden wird das in Punkt 2 beschriebene Iterationsverfahren näher betrachtet. Es wird eine Verlustvariable definiert und gezeigt, dass diese mit jedem Iterationsschritt, ggf. nach Anpassung der Schrittweite, verkleinert wird, solange noch kein lokales Extremum der Verlustfunktion erreicht ist. Die 3D-Koordinaten *aller* Objektpunkte seien im Vektor x zusammengefasst. Die sechs reellen Parameter, durch welche die Kamerapose beschrieben werden kann, seien in p enthalten. Die Bildpunktkoordinaten, welche im Vektor y_{mess} stehen, errechnen sich durch die nichtlineare Funktion $F(x, p)$ und durch additive mittelwertfreie Störungen n, welche die Kovarianzmatrix P_{nn} beschreibt:

$$y_{mess} = F(x, p) + n \tag{1}$$

Dem Schätzalgorithmus müssen erwartungstreue A-priori-Schätzungen $x^{(0)}$ und $p^{(0)}$ sowie zugehörige Kovarianzmatrizen $P_{xx}^{(0)}$ und $P_{pp}^{(0)}$ übergeben werden. $x^{(0)}$ sei unkorreliert zu $p^{(0)}$. Die Linearisierungspunkte im n-ten Iterationsschritt seien $x^{(n)}$ und $p^{(n)}$. Weiter sei

$$A := \frac{\partial}{\partial x} F(x, p)\bigg|_{x=x^{(n)}} \quad \text{und} \quad B := \frac{\partial}{\partial p} F(x, p)\bigg|_{p=p^{(n)}} . \tag{2}$$

Eine optimale Schätzung $x^{(n+1)}$, $p^{(n+1)}$ wird durch Minimierung folgender Verlustfunktion $Q(x, p)$ erreicht:

$$\begin{aligned} Q(x, p) = {} & (y_{mess} - F(x, p))^T\, P_{nn}^{-1}\, (y_{mess} - F(x, p)) + \\ & (x - x_0)^T P_{xx}^{-1} (x - x_0) + (p - p_0)^T P_{pp}^{-1} (p - p_0) . \end{aligned} \tag{3}$$

Ausgehend von der nach Gl. (2) linearisierten Abbildungsgleichung werden x und p mit einem Kalman-Filter in der Kovarianz-Form (statt in der Informations-Form) geschätzt. Wenn F linear wäre, erhielte man die optimalen Vektoren x und p. Wenn sich Nichtlinearitäten dagegen zu stark auswirken, kann Q sogar ansteigen. In diesem Fall muss der Iterationsschritt wiederholt werden und eine geringere Änderung der aktuellen Vektoren x und p gewährleistet werden. Weil das Kalman-Filter in der Kovarianz-Form benutzt wird, kann $Q(x, p)$ nicht mit der Levenberg-Marquardt-Methode minimiert werden. Stattdessen werden zusätzlich zu der Messgleichung (1) die „künstlichen“ Gleichungen (4) und (5) als Grundlage für den Iterationsschritt hinzugezogen, durch welche die Annahme ausgedrückt wird, dass der aktuelle Linearisierungspunkt $x^{(n)}$, $p^{(n)}$ das Ergebnis einer zusätzlichen unabhängigen Messung mit den mittelwertfreien Messfehlern δ_x, δ_p und Kovarianzen $P_{\delta x \delta x}$ und $P_{\delta p \delta p}$ sei. Zur Ermittlung der nächsten Iterationsvektoren $x^{(n+1)}$, $p^{(n+1)}$ müssen x und p in Gl. (4) und (5) durch diese ersetzt werden:

$$x^{(n)} = x + \delta_x \tag{4}$$

$$p^{(n)} = p + \delta_p . \tag{5}$$

Bei Näherung an das Minimum von $Q(x, p)$ müssen $P_{\delta x \delta x}$ und $P_{\delta_p \delta_p}$ gegen unendlich gehen, damit dann Gl. (4) und (5) für die Schätzung unwirksam werden. Solange $Q(x, p)$ aber noch vom Minimum entfernt ist, werden $P_{\delta x \delta x}$ und

$P_{\delta_p\delta_p}$ ähnlich wie beim Levenberg-Marquardt-Verfahren adaptiert: Wenn mit einem Iterationsschritt eine Verkleinerung von Q erreicht wurde, werden $P_{\delta x\delta x}$ und $P_{\delta_p\delta_p}$ mit einem bestimmten Faktor vergrößert, um die Schrittweite im nächsten Iterationsschritt zu vergrößern. Im anderen Fall werden die errechneten Vektoren $x^{(n+1)}$, $p^{(n+1)}$ verworfen und $P_{\delta x\delta x}$, $P_{\delta_p\delta_p}$ werden mit einem Faktor verkleinert, sodass eine kleinere Schrittweite erreicht wird; der Iterationsschritt $n \to n+1$ wird dann wiederholt.

Im Folgenden wird gezeigt, dass bei genügend kleinen Kovarianzen $P_{\delta x\delta x}$, $P_{\delta_p\delta_p}$ ein Iterationsschritt erfolgt, dessen zugehörige Änderungen Δx und Δp mit dem Gradienten von $Q(x, p)$ Skalarprodukte mit negativen Werten haben, sodass $Q(x, p)$ bei kleiner Schrittweite Δx, Δp auch verkleinert werden kann.

Voraussetzung für die weiteren Rechnungen ist folgende Wahl für die künstlichen Fehlerkovarianzen:

$$P_{\delta x\delta x} = \alpha\ P_{xx}^{(0)} \text{ und } P_{\delta p\delta p} = \alpha\ P_{pp}^{(0)},\quad \alpha > 0\,, \tag{6}$$

wobei α ein positiv reeller Faktor ist. Entsprechend dem Kalman-Filter wird nun ausgehend von der erwartungstreuen Schätzung $x^{(0)}$, $p^{(0)}$, deren A-priori-Kovarianzen $P_{xx}^{(0)}$ und $P_{pp}^{(0)}$ sind, durch Auswertung der zusätzlichen Informationen in Gl. (1), (4) und (5) eine neue erwartungstreue Schätzung $\tilde{x}^{(n+1)}$, $\tilde{p}^{(n+1)}$ derart gewählt, dass die A-posteriori-Varianzen, also die Hauptdiagonalelemente von $\tilde{P}_{xx}^{(n+1)}$ und $\tilde{P}_{pp}^{(n+1)}$, minimal werden. Später muss aus $\tilde{x}^{(n+1)}$, $\tilde{p}^{(n+1)}$ noch eine Schätzung $x^{(n+1)}$, $p^{(n+1)}$ berechnet werden, indem eventuell vorgegebene Abstände zwischen bestimmten Landmarken berücksichtigt werden. Für $\tilde{x}^{(n+1)}$, $\tilde{p}^{(n+1)}$ ergibt eine Rechnung:

$$\begin{pmatrix}\tilde{x}^{(n+1)}\\ \tilde{p}^{(n+1)}\end{pmatrix} = \begin{pmatrix}x^{(n)}\\ p^{(n)}\end{pmatrix} + \frac{\alpha}{1+\alpha}\begin{pmatrix}x^{(0)}-x^{(n)}\\ p^{(0)}-p^{(n)}\end{pmatrix} + \begin{pmatrix}K_x\\ K_p\end{pmatrix}$$
$$\cdot\left[y_{mess} - F\left(x^{(n)}, p^{(n)}\right) - \frac{\alpha}{1+\alpha}\,(A|\ B)\begin{pmatrix}x^{(0)}-x^{(n)}\\ p^{(0)}-p^{(n)}\end{pmatrix}\right]\,. \tag{7}$$

Dabei wurden die Abkürzungen

$$\begin{pmatrix}K_x\\ K_p\end{pmatrix} = \begin{pmatrix}P_{xx}^{(0)}A^T\\ P_{pp}^{(0)}B^T\end{pmatrix}\left(AP_{xx}^{(0)}A^T + BP_{pp}^{(0)}B^T + \frac{1+\alpha}{\alpha}P_{nn}\right)^{-1} \tag{8}$$

verwendet.

Im Folgenden wird gezeigt, dass mit der Schätzung nach Gl. (7) die Verlustvariable $Q(\tilde{x}^{(n+1)}, \tilde{p}^{(n+1)})$ (siehe Gl. (3)) gegenüber $Q(x^{(n)}, p^{(n)})$ zumindest dann verkleinert werden kann, wenn $\alpha > 0$ genügend klein gewählt wird und wenn $(x^{(n)}, p^{(n)})$ noch kein Extrempunkt von Q ist.

Unter der Voraussetzung, dass P_{nn} regulär ist, gilt nach Gl. (8):

$$\lim_{\alpha\to 0}\begin{pmatrix}K_x\\ K_p\end{pmatrix} = \frac{\alpha}{1+\alpha}\begin{pmatrix}P_{xx}^{(0)}A^T\\ P_{pp}^{(0)}B^T\end{pmatrix}\ P_{nn}^{-1}\,. \tag{9}$$

Auch Gl. (7) kann für genügend kleine $\alpha > 0$ vereinfacht werden. Der letzte Term $(A|\ B)\begin{pmatrix}x^{(0)}-x^{(n)}\\ p^{(0)}-p^{(n)}\end{pmatrix}$ hat die Vorfaktoren $\begin{pmatrix}K_x\\ K_p\end{pmatrix}\cdot\frac{\alpha}{1+\alpha}$, ist also wegen Gl. (9) proportional zu $\left(\frac{\alpha}{1+\alpha}\right)^2$ und braucht daher für kleine α nicht berücksichtigt zu werden. Somit wird aus Gl. (7):

$$\begin{pmatrix}\tilde{x}^{(n+1)}\\ \tilde{p}^{(n+1)}\end{pmatrix} \approx \begin{pmatrix}x^{(n)}\\ p^{(n)}\end{pmatrix} + \frac{\alpha}{1+\alpha}$$
$$\cdot\left(\begin{pmatrix}x^{(0)}-x^{(n)}\\ p^{(0)}-p^{(n)}\end{pmatrix} + \begin{pmatrix}P_{xx}^{(0)}A^T\\ P_{pp}^{(0)}B^T\end{pmatrix}\ P_{nn}^{-1}\left[y_{mess} - F\left(x^{(n)}, p^{(n)}\right)\right]\right)\,. \tag{10}$$

Für kleine α wird Q unter der Bedingung kleiner, dass die Änderung

$$\begin{pmatrix}\Delta x\\ \Delta p\end{pmatrix} = \begin{pmatrix}\tilde{x}^{(n+1)}\\ \tilde{p}^{(n+1)}\end{pmatrix} - \begin{pmatrix}x^{(n)}\\ p^{(n)}\end{pmatrix} \tag{11}$$

mit dem Gradienten grad Q ein negatives Skalarprodukt hat. Dieser Gradient errechnet sich mit Gl. (3) und (2) zu

$$\frac{1}{2}\text{grad } Q\left(x^{(n)}, p^{(n)}\right) = \frac{1}{2}\begin{pmatrix}\partial Q/\partial x\\ \cdots\cdots\\ \partial Q/\partial p\end{pmatrix}$$
$$= \begin{pmatrix}\left(P_{xx}^{(0)}\right)^{-1}\left(x^{(n)}-x^{(0)}\right)\\ \left(P_{pp}^{(0)}\right)^{-1}\left(p^{(n)}-p^{(0)}\right)\end{pmatrix} \tag{12}$$
$$-\begin{pmatrix}A^T\\ B^T\end{pmatrix}\ P_{nn}^{-1}\left[y_{mess} - F(x^{(n)}, p^{(n)})\right]\,.$$

Somit gilt mit Gl. (10)

$$\begin{pmatrix}\Delta x\\ \Delta p\end{pmatrix} = -\frac{1}{2}\cdot\frac{\alpha}{1+\alpha}\begin{pmatrix}P_{xx}^{(0)} & 0\\ 0 & P_{pp}^{(0)}\end{pmatrix}\ \text{grad } Q\left(x^{(n)}, p^{(n)}\right)\,. \tag{13}$$

Das Skalarprodukt

$$\left(\Delta x^T \middle|\ \Delta p^T\right)\ \text{grad } Q\left(x^{(n)}, p^{(n)}\right)$$
$$= -\frac{1}{2}\cdot\frac{\alpha}{1+\alpha}\left(\text{grad } Q\left(x^{(n)}, p^{(n)}\right)\right)^T\begin{pmatrix}P_{xx}^{(0)} & 0\\ 0 & P_{pp}^{(0)}\end{pmatrix}$$
$$\cdot\,\text{grad } Q\left(x^{(n)}, p^{(n)}\right)$$

ist unter der Voraussetzung, dass $P_{xx}^{(0)}$ und $P_{pp}^{(0)}$ regulär sind und grad $Q(x^{(n)}, p^{(n)})$ nicht verschwindet, negativ, weil dann $P_{xx}^{(0)}$ und $P_{pp}^{(0)}$ positiv definit sind. Somit wird bei genügend kleinem $\alpha > 0$ mit dem Iterationsschritt eine Verkleinerung von Q erreicht.

Nach der Bestimmung von $\tilde{x}^{(n+1)}$ gemäß Gl. (7) wird daraus $x^{(n+1)}$ berechnet, indem die zuvor unberücksichtigten, exakt vorgegebenen Abstände zwischen bestimmten Objektpunkten mit einbezogen werden. Die Pose $p^{(n+1)}$ dagegen wird gleich $\tilde{p}^{(n+1)}$ gesetzt.

3 Simulationsergebnisse

Es werden 10 Objektpunkte angenommen, welche gleichmäßig (zufällig) in einem Würfelvolumen von $(800\,\text{mm})^3$ verteilt sind, dessen Mittelpunkt mit dem Nullpunkt des 3D-Koordinatensystems übereinstimmt. Diese Punkte sollen in vielen Bildern aus jeweils verschiedenen Perspektiven betrachtet werden. Sowohl die jeweilige Kameraposition als auch -orientierung werden zufällig bestimmt. Es sollen jeweils alle Objektpunkte im Bild sichtbar sein, wobei ein maximaler Blickwinkel von ±50 Grad zur optischen Achse zugelassen wird (das Sichtfeld hat also den Winkel 100 Grad). Daher wird zunächst eine beliebige, zufällige optische Achse (Blickrichtung) erzeugt und dann die Kameraposition so gewählt, dass die Kamera den Abstand 1 m vom Koordinatennullpunkt hat und die optische Achse durch den Nullpunkt verläuft. Schließlich werden auf alle drei Kamerakoordinaten jeweils zufällige Werte zwischen −0,25 m und 0,25 m addiert. Dies bewirkt, dass die Blickrichtung grob in die Mitte des Volumens der 10 Objektpunkte zeigt.

Im Rechenprogramm wird ein globales Koordinatensystem zugrunde gelegt, auf das sich die Objektpunktkoordinaten und die Pose der Kamera beziehen. Die Koordinaten der Objektpunkte brauchen zu Beginn nur auf ±100 mm in jeder Koordinatenrichtung bekannt zu sein. Dem Schätzalgorithmus werden also A-priori-Koordinaten zur Verfügung gestellt, welche jeweils zufällig um maximal ±100 mm von den zuvor exakt definierten abweichen. Weiter erhält der Schätzalgorithmus entsprechende Standardabweichungen der Objektpunktkoordinaten, hier also

$$\sigma = {}^{1}\!\big/\!\sqrt{3} \cdot 100\ \text{mm} = 57{,}7\ \text{mm} \tag{14}$$

(Gleichverteilung zwischen −100 mm und 100 mm). In der Praxis kann man z. B. eine grobe Messung mit einem Meterstab vornehmen.

Nachdem die Abbildungen der Objektpunkte auf die Bildebene für die verschiedenen Perspektiven der virtuellen Kamera berechnet und diese Bilder ausgewertet wurden, konnte durch die Schätzungen zwar eine Verbesserung der Objektpunktkoordinaten erreicht werden. Es waren aber immer noch Abweichungen in der Größenordnung von mehr als 100 mm vorhanden (die A-priori-Unsicherheit war ±100 mm). Dies hat folgenden Grund: Die durch die Kamera erhaltenen Bilder ändern sich nicht, wenn man sowohl Kamera als auch alle Objektpunkte mit einem konstanten Translationsvektor verschiebt. Ebenso bleiben die Bilder bei einer Drehung aller Punkte und der Kamera um eine gemeinsame Achse gleich. Somit sind in den Bildern also keine Informationen über diese Translation bzw. Rotation enthalten. Diese sechs Freiheitsgrade machten sich in der zuvor genannten Simulation durch eine nur geringe Verbesserung der Objektpunktkoordinaten bemerkbar.

Abhilfe kann man wie folgt schaffen: Man definiert, dass die a priori angegebene Pose der Kamera vor Auswertung des ersten Bildes exakt sei. Dazu kann man z. B. den Nullpunkt des globalen Koordinatensystems so festlegen, dass er mit dem optischen Zentrum der Kamera (in ihrer ersten Lage) übereinstimmt und die Blickrichtung sowie die x- und y-Richtung der Bildebene mit jeweils einer Koordinatenachse zusammenfällt. Ausgehend von dieser Definition ist es in der Praxis problemlos möglich, die Objektpunktkoordinaten zu Beginn zumindest mit einer Unsicherheit von ±100 mm mit einem Meterstab zu vermessen. Durch die Verknüpfung der ersten Kamerapose mit dem Koordinatensystem existieren nun nicht mehr die Freiheitsgrade, die oben durch eine mögliche Rotation und Translation der Kamera und der Objektpunkte vorhanden waren.

Allerdings erhält man so immer noch keine zufriedenstellenden Schätzungen. Dies wird im Folgenden begründet und es wird eine Möglichkeit zur Abhilfe aufgezeigt. Im Gegensatz zu den zuvor genannten Bedingungen für die Simulation wird nun die Aufgabe betrachtet, aus sechs Objektpunkten, deren 3D-Koordinaten völlig unbekannt sind, und zwei Bildern dieser Punkte, welche mit zwei unterschiedlichen und unbekannten Kameraposen erzeugt wurden, sowohl diese Kameraposen als auch die Objektpunktkoordinaten zu berechnen. Dabei wird zunächst idealisierend angenommen, dass die Bildkoordinaten exakt vorliegen, also z. B. Fehler wegen der Pixelrasterung nicht vorhanden seien. Es wird ein zur ersten Kamerapose lokales Koordinatensystem definiert. Somit müssen bezüglich dieses Koordinatensystems die zweite Kamerapose sowie die Objektpunktkoordinaten berechnet werden. Das Problem führt auf 12 nichtlineare Gleichungen mit 12 Unbekannten.

Es zeigt sich jedoch, dass eine der Gleichungen von den übrigen abhängig ist. Dies kann man sich leicht erklären: Wie gesagt, sei die erste Kamera im Ursprung des Koordinatensystems. Wenn man nun die x-, y- und z-Koordinaten der zweiten Kameraposition und zugleich aller Objektpunkte mit einem konstanten Faktor multipliziert und die optische Achse sowie den Rechtsvektor der Kamera gleich lässt, so bleiben beide aufgenommenen Bilder unverändert. Somit ist das Problem bis auf einen Freiheitsgrad lösbar. Man erzwingt eine eindeutige Lösung, indem man eine zusätzliche einschränkende Bedingung angibt, z. B. indem man dem Schätzverfahren den *exakten* 3D-Abstand zwischen zwei Objektpunkten als A-priori-Information übergibt.

Daher wird im Programm noch die Möglichkeit vorgesehen, Abstände zwischen ausgewählten Objektpunkten vorgeben zu können. Der Einfachheit halber werden diese Abstände als exakt betrachtet. Tatsächlich kann man solche Abstände mit einem Meterstab auch deutlich genauer messen als z. B. die 3D-Koordinaten der Punkte. Es wurden nun Simulationen unter den zu Beginn des Kapitels 2 genannten Voraussetzungen durchgeführt. Bei der Kamera wurde eine Auflösung von 700 × 700 Pixel angenommen; die Mittelpunkte der Landmarken im Bild seien mit einer maximalen Unsicherheit von 1 Pixel bestimmbar. Bei Verwendung eines Verfahrens zur subpixel-genauen Bestimmung der Mittelpunkte erhält man entsprechend bessere Schätzwerte für die Kamerapose und die 3D-Koordinaten der Landmarken.

Zusätzlich wurden acht Abstände zwischen jeweils zwei der 10 Objektpunkte vorgegeben. Nach der Auswertung von zwei Bildern lagen die Standardabweichungen der geschätzten Objektpunktkoordinaten im Mittel bei 1,2 mm, während sie bei Vorgabe nur eines Abstands im Mittel bei 1,5 mm lagen. Gibt man gar keinen Abstand vor, so liegt die Standardabweichung bei 18 mm (zum Vergleich: die A-priori-Standardabweichung ist $\sigma = 57{,}7$ mm). Nach der Auswertung von 20 Bildern erhält man für 8, 1 und 0 vorgegebene Abstände im Mittel folgende Standardabweichungen für die Objektpunktkoordinaten: 0,75 mm, 0,77 mm, 18 mm. Wenn also kein Abstand vorgegeben wird, verbessert sich die Schätzung auch durch Auswertung vieler aus verschiedenen Perspektiven aufgenommener Bilder nicht.

Unsicherheiten in den geschätzten Posen rühren einerseits von Schätzfehlern der Objektpunktkoordinaten und andererseits von Fehlern bei der Bestimmung der Bildpunktkoordinaten her, welche z. B. aufgrund der Pixelrasterung der Kamera entstehen. Im Laufe der Zeit, d. h. nach Auswertung vieler Bilder, werden die Objektkoordinaten immer besser geschätzt. Allerdings wird nach ca. 5 bis 10 Bildern die Verbesserung nach jedem neuen Bild immer geringer. In der Praxis treten ab dem zehnten Bild kaum noch Verbesserungen auf. Um festzustellen, welcher Anteil der Posefehler durch die Fehler der Objektpunktkoordinaten bedingt ist, wurden auch Simulationen mit exakt bekannten Objektpunktkoordinaten durchgeführt und die Posefehler berechnet. Nach einer Mittelung der Positionsfehlervarianzen der Kamera über die Schätzungen bei verschiedenen Posen der Kamera erhielt man in x-, y- und z-Richtung die Standardabweichungen 0,82 mm, 0,95 mm und 0,95 mm. Zum Vergleich: Bei einer Simulation mit unsicheren Objektpunktkoordinaten erhält man

- bei einem vorgegebenen Abstand zwischen Objektpunkten Unsicherheiten der Kameraposition von 1,67 mm, 1,50 mm und 1,51 mm in x-, y- und z-Richtung bzw.
- bei acht vorgegebenen Abständen Unsicherheiten von 1,63 mm, 1,46 mm und 1,48 mm in x-, y- und z-Richtung.

Bei den hier angegebenen Standardabweichungen handelt es sich um Mittelwerte über die Positionsfehler der Kamera nach der Auswertung der Bilder mit den Nummern 8 bis 20. Der Unterschied zwischen dem Fall mit einem bzw. acht vorgegebenen Abständen fällt gering aus. Deutlichere Unterschiede stellt man nur in den Posefehlern und den Fehlern der Objektpunktkoordinaten fest, wenn noch wenige Bilder ausgewertet sind, d. h. die Objektpunkte noch unzureichend bekannt sind.

Aus dem Vergleich der Angaben zur Positionsunsicherheit der Kamera im Fall exakt bekannter Objektpunktkoordinaten mit den Unsicherheiten bei unsicheren Koordinaten ist ein Faktor in der Größenordnung von 2 ersichtlich. Somit kann man sagen, dass die Unsicherheit der geschätzten Kamerapose etwa zu gleichen Teilen durch die Unsicherheit der Objektpunktkoordinaten und die Unsicherheit der ermittelten Bildpunktkoordinaten bedingt wird. Allerdings wurden hier nur die Bilder mit den Nummern 8 bis 20 in die Mittelung der Positionsfehler der Kamera einbezogen, d. h. zuvor konnte die Unsicherheit der Objektpunktkoordinaten (a priori: $\sigma = 57{,}7$ mm) schon deutlich reduziert werden.

Es wurden nun weitere Simulationen unter den gleichen Bedingungen durchgeführt, wobei aber die Bildkoordinaten der Objektpunkte mit einem maximalen Fehler von 0,5 Pixel (statt zuvor 1,0 Pixel) versehen wurden. Erwartungsgemäß ergab sich eine Halbierung sämtlicher Standardabweichungen, d. h. sowohl die Kameraposition als auch die Orientierungswinkel der Kamera und die Objektpunktkoordinaten wurden mit halber Unsicherheit geschätzt. Das bedeutet für die Praxis: Die Objektpunktkoordinaten müssen im Bild möglichst genau geschätzt werden. Eine Subpixel-Auflösung erreicht man z. B. dann, wenn man den jeweiligen Objektpunkt als den Mittelpunkt eines ausgefüllten Kreises definiert, welcher in der Bildebene als Ellipse erscheint und eine Fläche von 100 Pixel besitzt. Der Schwerpunkt der Ellipse ist der gesuchte Punkt in der Bildebene.

4 Messergebnisse

Zur experimentellen Verifikation des Schätzverfahrens wurden an den Wänden eines Raums 18 Landmarken angebracht. Im Bild 1 erkennt man sie als kleine dunkle Kreise auf hellem Hintergrund. Diese Landmarken bzw. Teilmengen davon wurden mit einer CCD-Kamera aus 35 verschiedenen Perspektiven aufgenommen. Die Kamera hatte eine Auflösung von 768×576 Pixel. Die durch das Objektiv bedingten Verzeichnungen sind im Bild 1 deutlich durch gekrümmte Kurven als Bilder von Strecken im 3D-Raum erkennbar. Durch eine Vorab-Kalibrierung wurden die Verzeichnungsparameter ermittelt [6], sodass der Einfluss der Verzeichnung rechnerisch eliminiert werden kann.

Bezüglich der nach dieser Korrektur verbleibenden Restfehler liefert das Kalibrierprogramm folgende Werte: Die

Bild 1: Anordnung der Landmarken im Raum.

Standardabweichung der korrigierten Bildkoordinaten eines Punktes zu den Koordinaten des nach dem Lochkameramodell berechneten Punktes ist etwa 0,2 Pixel in x- und y-Richtung. Tatsächlich sind aber höhere Werte anzusetzen, denn eine für die Kalibrierung notwendige Vermessung eines schachbrettartigen Musters, welches als Kalibriernormal dient, bedingt Messabweichungen. Das mit einem Laserdrucker erzeugte Muster im DIN-A4-Format ist zudem nicht exakt rechteckig, wie mit einem Lineal festgestellt wurde. Eine fehlerbehaftete Längen- und Breitenabmessung der Rechtecke des Schachbrettmusters hat aber direkten Einfluss auf die vom Kalibrierprogramm errechnete Brennweite der Kamera.

Weil die Objektpunktkoordinaten vorab nur grob bekannt sein müssen, wurden sie mit einer maximalen Abweichung von 50 mm gemessen, und diese Information wurde an das C++-Programm zur Poseberechnung übergeben. Wie bereits erläutert, muss mindestens *ein* Abstand zwischen einem Paar von Objektpunkten (möglichst) exakt bekannt sein und dem Programm a priori vorgegeben werden. Es wurden die Fälle ,,ein bekannter Abstand“ und ,,fünf bekannte Abstände“ untersucht. Alle Abstände wurden vorab mit einem Meterstab gemessen. Die relativen Fehler dieser Messungen haben direkten Einfluss auf die relativen Fehler sämtlicher vom Programm errechneten Längen. Zusätzlich wurden weitere Abstände zwischen ausgewählten Objektpunkten vermessen, welche aber nicht a priori verwendet wurden, sondern als Kontrolle über die Abweichungen der vom Programm ermittelten Objektpunktkoordinaten dienten.

Die Verifikation der vom Programm gelieferten Posen wird dadurch erschwert, dass man mit einem anderen Messverfahren als Bildverarbeitung die absolute Position des optischen Zentrums der Kamera nicht messen kann, weil dieses innerhalb der Kamera (oder im Objektiv) ist. Es können jedoch Positionsänderungen relativ zu einem bestimmten Koordinatensystem gemessen werden. Dessen Bezug zu dem Koordinatensystem, bezüglich dessen das Programm die Posen berechnet, ist jedoch nicht direkt ersichtlich, auch dann nicht, wenn das Programm sein Koordinatensystem mit der als exakt definierten ersten Pose der Kamera verknüpft. Denn die optische Achse (Blickrichtung) der Kamera ist bei Kenntnis der Orientierung des Kameragehäuses nur näherungsweise angebbar.

Um die rechnerische Beziehung beider Koordinatensysteme herzustellen, kann man die Kamera z. B. auf einem x-y-Tisch befestigen. Bei dem verwendeten Tisch werden maximale Abweichungen in der Größenordnung von 50 μm erreicht. Wenn man mit der Kamera bei konstanter Orientierung drei Punkte anfährt, die nicht auf einer Geraden liegen, und jeweils die Posen bezüglich des vom Poseprogramm verwendeten Koordinatensystems berechnen lässt, so kann man beide Koordinatensysteme rechnerisch bereits in Beziehung setzen. Diese Methode eignet sich auch, um von einem zur Kamera lokalen Koordinatensystem in ein Koordinatensystem umzurechnen, das lokal zu einem Werkzeug am Endeffektor eines Roboters ist, wenn die Kamera ebenfalls am Endeffektor montiert ist.

In diesem Versuch werden jedoch nur Längenänderungen verglichen. Z. B. wird die Kamera um 50 mm in einer Richtung verschoben, und es wird die Länge des Verschiebungsvektors berechnet, den das Poseprogramm liefert. Hierzu muss bemerkt werden, dass vor allem zufällige Fehler, bedingt durch die Pixelrasterung und das Messrauschen, erfasst werden, denn die Verschiebung um 50 mm ist klein im Verhältnis zur Größe des Raums. Systematische Fehler, bedingt durch unzureichende Korrektur der Verzeichnung, bleiben hierbei weitgehend unentdeckt.

Bei Verschiebungen der Kamera um bis zu 50 mm ergaben sich Abweichungen der geschätzten Position in der Größenordnung von 1 bis 2 mm. Diese Größenordnung ist plausibel, wenn man annimmt, dass die 2D-Koordinaten der Objektpunkte mit einer maximalen Abweichung von etwa 0,2 Pixel bestimmt werden können und die 3D-Objektpunktkoordinaten ausreichend genau geschätzt seien. Aufgrund der Brennweite des Kameraobjektivs ergibt sich ein Blickfeld mit einem Öffnungswinkel von etwa 50 Grad in der zur Kamera lokalen x-Richtung (Normalobjektiv). Eine plausible Abschätzung für die Abweichungen der geschätzten Kameraposition ergibt sich, indem man den Fehlervektor in der Bildebene rückwärts in die Ebene projiziert, auf der die Landmarken sind. Bei einem Kameraabstand von 4,4 m zur Türebene erhält man Fehlerabstände von 1,1 mm. Von gleicher Größenordnung ist der Fehler, der bei den in Abschnitt 3 genannten Simulationen entstand. Dort wurden maximale Bildpunkt-Fehler von 1 Pixel angenommen, während bei den realen Experimenten Abweichungen von bis zu 0,2 Pixel auftraten. Das ist kein Widerspruch, denn der Faktor 5 (1/0,2) findet sich auch in den Ausdehnungen des Messvolumens wieder: $4\,\mathrm{m}/0{,}8\,\mathrm{m} = 5$.

Dennoch können die Abweichungen verringert werden. Der Fehler in den geschätzten 2D-Koordinaten eines Objektpunkts setzt sich zusammen

- aus dem Fehler aufgrund einer unzureichend genauen Korrektur der Verzeichnung und
- aus dem Fehler in der Bestimmung der 2D-Objektpunktkoordinaten im entzerrten Bild, bedingt durch die Pixelrasterung und durch stochastische Veränderungen der Grauwerte (Messrauschen).

Abgesehen vom Messrauschen sind bei beiden Fehlerquellen Verbesserungen möglich. Die Korrektur der Verzeichnung lässt sich vermutlich optimieren durch Anfertigung eines genaueren Schachbrettmusters als Kalibriernormal (s.o.). Außerdem kann der Algorithmus zur subpixelgenauen Mittelpunktsbestimmung der Landmarken (Ellipsen im Bild) verbessert werden. Bisher werden nur die zur Ellipse gehörenden Pixel detektiert, und daraus wird der Schwerpunkt bestimmt, wobei allen Pixeln das gleiche Gewicht zugeordnet wird. Eine Verbesserung ist möglich durch eine Gewichtung der Randpixel entsprechend ihrem

Grauwert, denn im Bild erscheint nicht ein scharfer, sondern ein fließender Übergang vom Bereich der Ellipse zum Hintergrund.

Eine weitere Verringerung der Messunsicherheit der berechneten Pose ist prinzipiell durch den Einsatz einer Kamera mit einer höheren Auflösung möglich. Allerdings macht dies nur Sinn, wenn die anderen Parameter, die zur Messunsicherheit beitragen, auch entsprechend optimiert werden. Unter anderem ist dann eine Approximation der radialen Verzeichnungsfunktion durch ein Polynom 6. Grades statt wie bisher 4. Grades in Betracht zu ziehen. Denn die durch fehlerhafte Kompensation der Verzeichnung entstehenden Fehler in den Bildkoordinaten sollten deutlich kleiner als 1 Pixel sein, wobei nun auf die kleineren Pixel Bezug genommen wird.

Im Folgenden werden die Unsicherheiten in den geschätzten Objektpunktkoordinaten diskutiert. Dabei besteht jedoch das Problem, sie mit einem anderen Verfahren als Bildverarbeitung zu messen, um einen Vergleich mit den berechneten Koordinaten durchführen zu können. Auf eine hochgenaue Messung der 3D-Koordinaten wurde verzichtet. Die Objektpunktkoordinaten wurden zuerst durch Auswertung bestimmter Kamerabilder, dann noch einmal unabhängig davon mit anderen Bildern bestimmt, und für jede Objektpunktkoordinate wurde die Differenz zwischen beiden geschätzten Werten ermittelt. Auf diese Weise erhält man einen Anhaltspunkt für die Größe zufälliger, aber nicht systematischer Fehler. Sowohl im Fall von einem als auch von fünf a priori vorgegebenen Abständen zwischen Objektpunkten betrugen die Abweichungen in x- und y-Richtung durchschnittlich 1 mm und in z-Richtung 3 mm (das ist die Richtung senkrecht zu der im Bild 1 sichtbaren Tür). Zusätzlich wurden die sich durch die Schätzung ergebenden Abstände zwischen bestimmten Objektpunkten mit gemessenen Abständen verglichen. Die Größenordnung der Fehler von 1 mm konnte dadurch bestätigt werden.

5 Ergebnis und Ausblick

Dieser Beitrag geht auf verschiedene Aspekte der Posebestimmung einer Kamera mithilfe von Landmarken ein. In Kapitel 2 wird das verwendete Iterationsverfahren beschrieben, mit dem man unter Kenntnis der aus dem Bild extrahierten 2D-Koordinaten der Objektpunkte die Pose der Kamera bestimmen und die Schätzungen der 3D-Positionen der Objektpunkte verbessern kann. Die in Kapitel 3 angegebenen Simulationsergebnisse beziehen sich auf 10 Landmarken, die in einem Quadervolumen von $(800\,\text{mm})^3$ verteilt sind, auf einen Kameraabstand von 1 m zum Mittelpunkt des Quaders, auf eine Kamera mit 700×700 Pixel Auflösung und ein Blickfeld, dessen Öffnungswinkel 100 Grad beträgt. Es ergaben sich Schätzfehler von etwa 0,7 mm bei den Objektpunktkoordinaten und von etwa 1,5 mm bei den Koordinaten der Kameraposition. In Kapitel 4 werden Messergebnisse angegeben für die im Raum (siehe Bild 1) angebrachten Landmarken. Dabei hat die Kamera je nach Pose Abstände von 2 bis 4 Metern zu den Landmarken, 768×576 Pixel Auflösung und ein Normalobjektiv (deren Blickfeld hat einen Öffnungswinkel von etwa 50 Grad). Die Fehler der geschätzten Kameraposition betragen etwa 1 bis 2 mm, die Fehler der Objektpunktkoordinaten liegen in der Größenordnung von 1 mm.

Die Schätzung der Kamerapose und der 3D-Objektpunktkoordinaten wird mit dem Kalman-Filter in der Kovarianzform realisiert. Die Linearisierung der optischen Abbildungsgleichungen ist nur dann möglich, wenn die Objektpunktkoordinaten a priori zumindest grob bekannt sind. Wenn sie jedoch a priori gar nicht bekannt sind, können mit der so genannten mengenbasierten Zustandsschätzung bessere Ergebnisse erzielt werden. Dabei werden unbekannte Größen nicht stochastisch beschrieben, sondern durch Mengen, in denen sie mit Sicherheit liegen. Diese Mengen können verkleinert werden, wenn neue Information durch Messgleichungen geliefert wird. In [2] werden ellipsoide Mengen benutzt. Eine Erweiterung dieses Konzepts stellt das so genannte SSI-Filter dar, welches die mengenbasierte Darstellung mit einer stochastischen Beschreibung der Größen kombiniert und damit die Vorteile beider Verfahren vereint [3].

Danksagung

Die hier vorgestellten Arbeiten wurden von der Deutschen Forschungsgemeinschaft finanziert. Die Autoren danken für die ihnen zuteil gewordene Unterstützung.

Literatur

[1] *Faugeras, O.:* Three-Dimensional Computer Vision. Camebridge: The MIT Press 1999.

[2] *Hanebeck, U.:* Lokalisierung eines mobilen Roboters mittels effizienter Auswertung von Sensordaten und mengenbasierter Zustandsschätzung. Fortschrittsberichte VDI, Reihe 8, Nr. 643, (1997).

[3] *Hanebeck, U.:* Zustandsschätzung im Fall simultan auftretender mengenbasierter und stochastischer Unsicherheiten. In: Automatisierungstechnik at 48 (2000), S. 265–272.

[4] *Jähne, B.:* Digitale Bildverarbeitung. Berlin: Springer 2001.

[5] *Janocha, H., Königstein, F.:* Verfahren zur Schätzung der Pose von Industrierobotern mit Bildverarbeitung. 47. Internationales Wissenschaftliches Kolloquium. Ilmenau. 23.–26.9.2002.

[6] Open Source Computer Vision Library, speziell der Link zum Camera Calibration Tool. Erreichbar unter http://www.intel.com/research/mrl/research/opencv.

Prof. Dr.-Ing. habil. Hartmut Janocha, Inhaber des Lehrstuhls für Prozessautomatisierung der Universität des Saarlandes. Adresse: Universität des Saarlandes, Gebäude 13, Lehrstuhl für Prozessautomatisierung, Postfach 151150, D-66041 Saarbrücken, Tel.: 0681-302-2694,
E-Mail: janocha@lpa.uni-saarland.de

Dipl.-Ing. Florian Königstein, wissenschaftlicher Mitarbeiter im Bereich Bildverarbeitung am Lehrstuhl für Prozessautomatisierung.