

Metrische 3-D-Vermessung einer Straßenszene mit Spezialwissen unter minimaler Interaktion

Erweiterte monokulare Rekonstruktion

MICHAEL TRUMMER¹, HERBERT SÜSSE², JOACHIM DENZLER³, FRANK DITRICH⁴, Jena

Zusammenfassung: Das bisher von der Polizei in Baden-Württemberg eingesetzte Verfahren zur Vermessung von Unfallszenen basiert auf der Berechnung der projektiven Transformation der Straßen- in die Bildebene anhand von Punktreferenzen. Hierzu werden per Hand Markierung auf die Fahrbahn aufgetragen und vermessen, um mittels CAD- und Bildkoordinaten aus mehreren Aufnahmen über Ausgleichsrechnung die projektive Transformation zu bestimmen. Naturgemäß ist dieses 2-D-Verfahren auf die Straßenebene beschränkt und die Ergebnisse weisen Fehler auf, sobald die Straßenoberfläche nicht ideal planar ist. Um nun Abstände zwischen beliebigen Raumpunkten der Unfallszene bestimmen zu können, wird als neuer Ansatz die partielle 3-D-Rekonstruktion der Szene, nämlich der Referenz- sowie der zu vermessenden Raumpunkte gewählt. Dabei kommt eine Erweiterung der monokularen Rekonstruktion eines Vierecks zum Einsatz, welche als a-priori-Wissen die CAD-Koordinaten dieses Vierecks und die internen Parameter der Kamera benötigt und die Interaktion am Einsatzort auf ein Minimum beschränkt.

Summary: For the purpose of measuring a traffic crash scene, currently, the Police Department in Baden-Württemberg is using a method that is based on calculating the projective transformation between the street and image plane with point correspondences. Therefore point markings are drawn on the street by hand and the transformation between known CAD and image coordinates is yielded as least squares solution. Naturally, this 2D method is restricted to the street plane and produces errors for street surfaces being not perfectly plane. Now, for measuring spatial distances between arbitrary 3D points, the new approach is the partial 3D reconstruction of the scene, namely the reference and measurement points. For this, an extension of the monocular reconstruction of a quadrangle is applied, using the CAD coordinates of this quadrangle and the intrinsic camera parameters as special knowledge. In-place user interaction is minimized.

1 Einleitung

Bei der Analyse von Verkehrsunfällen kommen Mittel der digitalen Bildverarbeitung zum Einsatz, insbesondere weil Sachverständige nicht immer vor Ort sein können. Zum Aufklären des Unfallhergangs und der Schuldfrage kann es z. Bsp. nötig sein, aus Bildern der Unfallszene Abstände von Trümmerteilen oder die Länge einer Bremsspur zu messen.

Zu diesem Zweck hat vor zwei Jahren bei der Firma DAKO in enger Zusammenarbeit mit der Polizeidirektion Freiburg die Entwicklung eines entsprechenden Programms begonnen, das Ergebnis ist das 2-D-Vermessungssystem DAKO-MESS. Dieses Programm ermöglicht Vermessungen in der Straßenebene mit dem folgenden Verfahren. Bei der Abbildung in die Bildebene gemäß

1 trummer@minet.uni-jena.de

2 Dr. rer. nat. habil. Herbert Süße, nbs@uni-jena.de

3 Prof. Dr.-Ing. Joachim Denzler, Friedrich-Schiller-Universität Jena, Fakultät für Mathematik und Informatik, Lehrstuhl für Digitale Bildverarbeitung, Ernst-Abbe-Platz 2, 07743 Jena, denzler@informatik.uni-jena.de, www.inf-cv.uni-jena.de

4 ditrich@minet.uni-jena.de

Lochkameramodell wird die Straßenebene einer zweidimensionalen projektiven Transformation unterworfen. Diese hat acht Parameter. Folglich sind vier Punktkorrespondenzen zwischen Straßen- und Bildebene vonnöten, um die Transformation zu berech-

Bildpunkte angewendet und der Euklidische Abstand der transformierten Koordinaten bestimmt.

Grundlage von DAKO-MESS ist also die projektive Transformation einer Ebene (der Straßenebene) in die Bildebene. Wenn nun

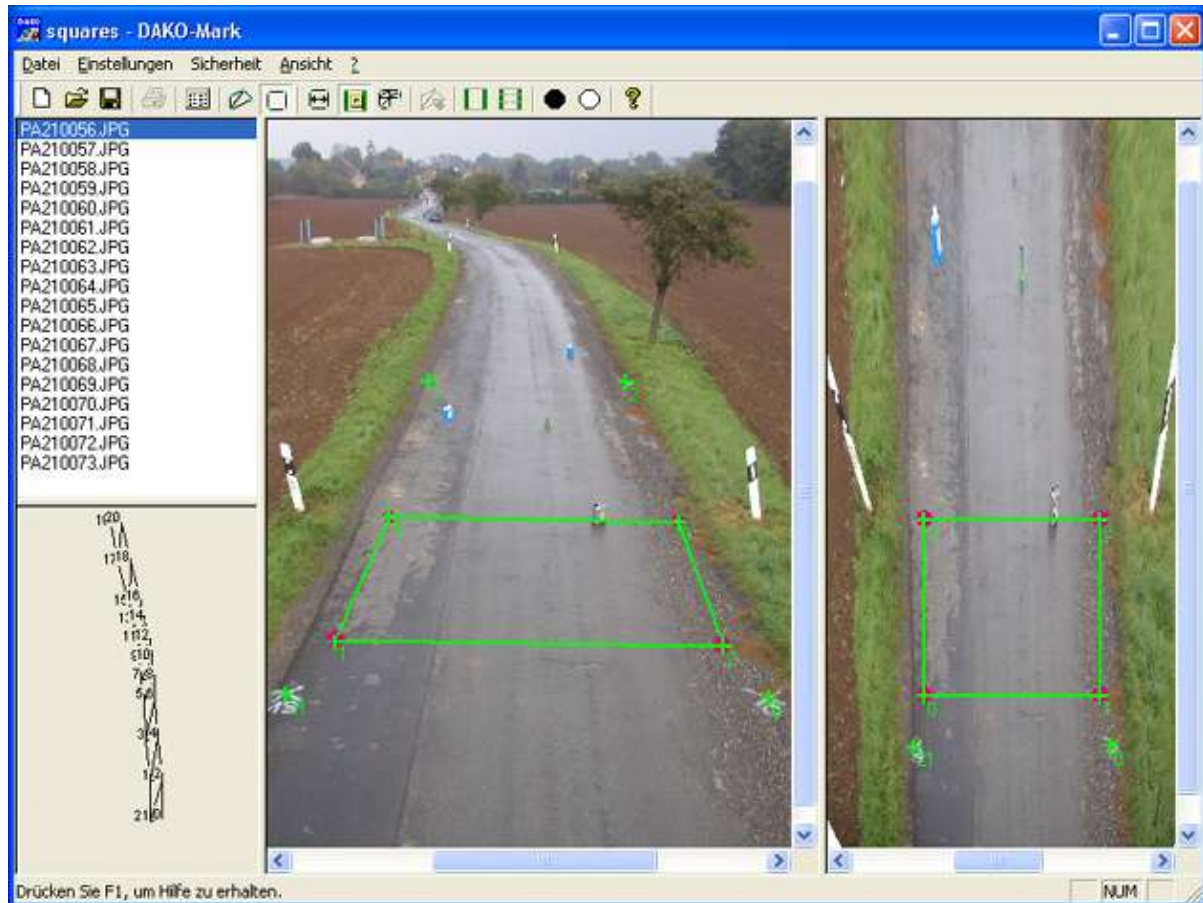


Abb.1: DAKO-MESS

nen (2 Gleichungen je Punktkorrespondenz macht $4 \cdot 2 = 8$ Gleichungen für 8 Unbekannte), s. Abb. 1. Mit mehr als vier Korrespondenzen ist Ausgleichsrechnung zur Fehlerkompensation nötig.

Das Verfahren sieht also vor, daß mindestens vier Punkte per Hand auf der Straße markiert und vermessen werden. Dies liefert die metrischen CAD-Koordinaten der Referenzpunkte. Anhand dieser Koordinaten und der entsprechenden Bildkoordinaten wird die projektive Transformation von der Bild- in die Straßenebene berechnet. Schließlich wird, zur Vermessung zweier Punkte der Straßenebene, die ermittelte Transformation auf die Koordinaten der entsprechenden

aber die Straßenoberfläche nicht eben ist oder Punkte zu vermessen sind, die nicht in der Straßenebene liegen (z. Bsp. Trümmersteine im Straßengraben), dann ist dieses Verfahren aufgrund seiner Konzeption nicht anwendbar.

Deswegen ist eine Methode gesucht, welche die Vermessung von 3-D-Punkten ermöglicht, dazu Spezialwissen verwendet und die notwendige Interaktion möglichst gering hält, um dem Nutzer die Anwendung nicht zu erschweren.

2 Erweiterte monokulare Rekonstruktion

2.1 Ansatz

Zur Vermessung einer Strecke im Raum werden aus mindestens zwei Bildern die Endpunkte im Dreidimensionalen rekonstruiert und deren Abstand berechnet. Das Ziel, mit minimaler Interaktion am Einsatzort möglichst genaue Ergebnisse stabil und robust zu erhalten, wird durch Separierung des Problems erreicht.

2.2 Spezialwissen

Als a-priori-Wissen werden die internen Kameraparameter, einschließlich Verzeichnungsparameter, verwendet sowie die CAD-Koordinaten eines in die Szene eingebrachten planaren Vierecks. Diese Daten für Kamera und Referenzviereck müssen nur einmal ermittelt werden, was im Vorfeld des Einsatzes im Labor erfolgt.

2.3 Interaktion

Nachdem das Referenzviereck in die Szene gelegt ist (je nach Entfernung zur Kamera z. Bsp. ein A4-Blatt oder ein entspr. größeres Viereck), beschränkt sich die Nutzerinteraktion auf das Markieren der zu vermessenden Punkte je Bild. Für die Selektion der abgebildeten Eckpunkte des Referenzvierecks ist eine Automatik vorgesehen (z. Bsp. mit Hilfe von unterschiedlichen farbigen Markierungen). Allgemeiner formuliert, muß vom Nutzer je Aufnahme das Abbild des Raumpunktes markiert werden, der Euklidisch rekonstruiert werden soll. Der Aufwand vor Ort ist minimal, da lediglich das Referenzviereck plaziert werden muß. Unter dem Gesichtspunkt geringen Aufwands ist natürlich auch die Zahl der Referenzpunkte möglichst niedrig zu wählen. Doch bei Dreiecken führt das Verfahren zu nichtlinearen Gleichungssystemen (vgl. Voss et al., 1995), die hier umgangen werden.

Wenn zudem als spezielles Viereck ein Quadrat gewählt wird, so ist es gleichgültig, bei welchem Eckpunkt mit der Auswahl im Bild begonnen wird.

2.4 Vorgehensweise

Zunächst betrachtet man ein einzelnes Bild. Mit dem a-priori-Wissen ist es möglich, aus den Bildkoordinaten der Eckpunkte des planaren Vierecks selbige im Raum Euklidisch zu berechnen. Man erhält die Lage dieser vier Punkte im Raum mit dem Projektionszentrum der Kamera als Koordinatenursprung. Das ist die **monokulare Rekonstruktion**.

Die **Erweiterung** bezieht sich auf die Verwendung mehrerer Bilder/Kameras.

Man hat also je Bild die Lage der Eckpunkte des Referenzvierecks im Raum, aber in unterschiedlichen Koordinatensystemen (nämlich bezogen auf das jeweilige Projektionszentrum). Da nun die monokulare Rekonstruktion Euklidisch ist, kann man die Koordinatensysteme durch Rotation und Translation zur Deckung bringen (bei Abwesenheit von Fehlern). Dies ergibt die externen Parameter einer jeden Kamera in Bezug zur ersten.

Damit sind die vollständigen Abbildungsmatrizen bekannt und Triangulation ist möglich.

2.4.1 Monokulare Rekonstruktion

Es wird mit einem Bild gearbeitet. Seien $\mathbf{y}^{(i)} = (y_1^{(i)} \ y_2^{(i)})^T$ (stets $i=1, \dots, 4$) die bekannten CAD-Koordinaten der Eckpunkte des Referenzvierecks und $\mathbf{x}^{(i)} = (x_1^{(i)} \ x_2^{(i)})^T$ die zugehörigen Bildkoordinaten. Dann bezeichnet $\mathbf{X}^{(i)} = (X_1^{(i)} \ X_2^{(i)} \ X_3^{(i)})^T$ die entsprechenden 3-D-Punkte, $\tilde{\mathbf{y}} / \tilde{\mathbf{x}} / \tilde{\mathbf{X}}$ zeigt jeweils die Schreibweise in homogenen Koordinaten an, $\tilde{\mathbf{y}}^{(i)} = (y_1^{(i)} \ y_2^{(i)} \ 1)^T$ usw. \sim steht für Gleichheit bis auf einen Faktor, Vektoren sind stets Spaltenvektoren.

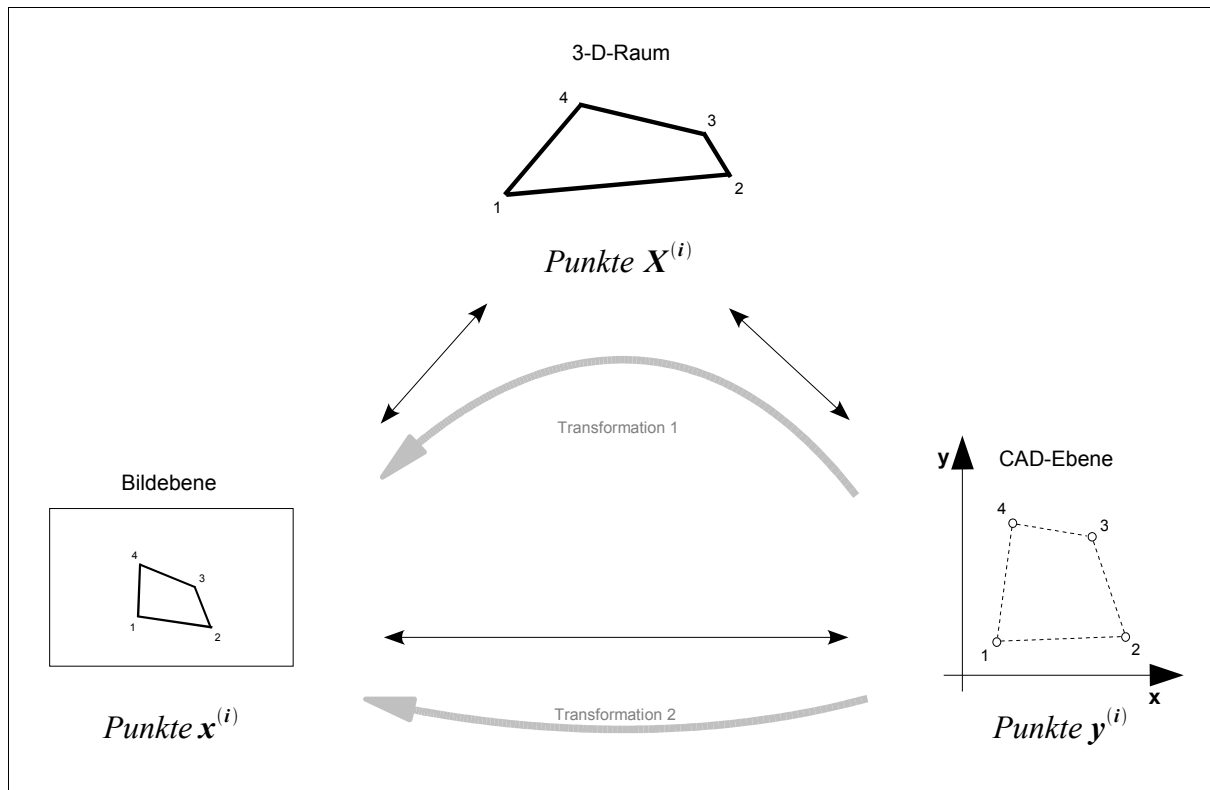


Abb. 2: Betrachtete Transformationen bei der monokularen Rekonstruktion

Die Punkte $X^{(i)}$ liegen in einer Ebene im Raum (die im Raum eingebettete CAD-Ebene), lassen sich also in der Form

$$X^{(i)} = d + y_1^{(i)} e^{(1)} + y_2^{(i)} e^{(2)} \quad (1)$$

schreiben, wobei d der Ursprung der CAD-Ebene im Raum ist und $e^{(1)}$, $e^{(2)}$ die normierten Einheitsvektoren der Achsen.

Die Punkte $x^{(i)}$ sind die (nach Möglichkeit von Verzerrung befreiten) Bildpunkte der $X^{(i)}$, somit

$$\tilde{x}^{(i)} \sim P \tilde{X}^{(i)}$$

mit der Abbildungsmatrix P der Kamera. Da aber der Ursprung des Koordinatensystems in das Kamerazentrum gelegt und die Kameraorientierung gemäß Einheitsmatrix I_3 gesetzt wurde, vereinfacht sich die Abbildungsgleichung zu

$$\tilde{x}^{(i)} \sim K X^{(i)}.$$

Zusammen mit der Beziehung (1) ergibt sich

$$\tilde{x}^{(i)} \sim K d + K y_1^{(i)} e^{(1)} + K y_2^{(i)} e^{(2)} \quad (2)$$

mit der Matrix der internen Kameraparameter

$$K = \begin{pmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Dies ist die erste Transformation gem. Abb. 2.

Andererseits existiert eine zweidimensionale projektive Transformation H von der CAD- in die Bildebene (Transformation 2 gem. Abb. 2), so daß

$$\tilde{x}^{(i)} \sim H \tilde{y}^{(i)}.$$

Sei nun $H = (h_1, h_2, h_3)$, dann ist leicht nachzurechnen, daß

$$\tilde{x}^{(i)} \sim h_1 y_1^{(i)} + h_2 y_2^{(i)} + h_3. \quad (3)$$

Aus den Gleichungen (2) und (3) ergibt sich

jetzt durch Koeffizientenvergleich bzgl. der $y_j^{(i)}$ ($j=1,2$)

$$\begin{pmatrix} K e^{(1)} \\ K e^{(2)} \\ K d \end{pmatrix} \sim \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} .$$

Man erhält also drei lineare Gleichungssysteme, die sehr simpel zu lösen sind, da K bereits in Zeilenstufenform ist (obere Dreiecksmatrix). Wegen der zugrundeliegenden projektiven Beziehungen gilt die Gleichheit aber nur bis auf einen gemeinsamen Faktor, der jedoch durch die Forderungen $|e^{(1)}|=1$ oder $|e^{(2)}|=1$ oder praktisch besser $|e^{(1)}| \cdot |e^{(2)}|=1$ leicht zu eliminieren ist. Folglich ergeben sich die Vektoren $d, e^{(1)}, e^{(2)}$ und damit die Punkte $X^{(i)}$, indem man mit linearen Mitteln eine projektive Transformation bestimmt und drei lineare Gleichungssysteme löst.

2.4.2 Bestimmung einer Euklidischen 3-D-Transformation aus Punktkorrespondenzen

Man betrachtet nun n Bilder ($n > 1$). Mit der monokularen Rekonstruktion erhält man für das Referenzviereck n verschiedene Euklidische 3-D-Rekonstruktionen mit Punkten ${}^{(j)}X^{(i)}$ (Punkt i im Bild $j, j=1, \dots, n$). Es ist bekannt, daß sich aus unterschiedlichen Bildern rekonstruierte Punkte ${}^{(k)}X^{(i)}, {}^{(l)}X^{(i)}$ ($1 \leq k, l \leq n, k \neq l$) nur durch Rotation und Translation unterscheiden. Im Dreidimensionalen braucht man zur Bestimmung dieser Transformation drei nicht kollineare Punkte.

Nun hat man aber vier gestörte Punktkorrespondenzen und gesucht ist somit die beste Lösung für die Euklidische Transformation im Sinne der kleinsten Defektquadrate. Also sind für Punkte aus Bildern k und l die Rotationsmatrix R und der Translationsvektor t gesucht, zu lösen ist das Ausgleichsproblem

$$\sum_{i=1}^4 \| {}^{(k)}X^{(i)} - (R^{(l)}X^{(i)} + t) \|^2 \rightarrow \text{Minimum},$$

R Rotationsmatrix.

Wegen der nichtlinearen Restriktionen an die Orthonormalmatrix R ist dieses Problem schwierig. Beherrschbar wird es, indem die Transformationen als duale Einheitsquaternionen dargestellt werden, vgl. WALKER et al., 1991. Dadurch kann das oben formulierte Ausgleichsproblem auf ein Eigenwertproblem zurückgeführt werden. So wird die Matrix

$$[\tilde{RT}]^{(l,k)} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

auf direktem Weg stabil berechnet und ist zudem die Lösung im Sinne der kleinsten Fehlerquadrate.

Damit sind aber Rotation und Translation zwischen den Kameras bekannt.

2.4.3 Rekonstruktion

Indem man die äußeren Parameter der ersten Kamera auf I_4 setzt und für weitere Kameras/Bilder k ($1 < k \leq n$) die Transformation $[\tilde{RT}]^{(1,k)}$ bestimmt, sind die äußeren Parameter aller Kameras (bzgl. der ersten) bekannt. Mit

$$\tilde{K} = \begin{pmatrix} \alpha_x & s & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

und

$${}^{(j)}P = \tilde{K} \cdot [\tilde{RT}]^{(1,j)}$$

ist die vollständige Abbildungsmatrix einer jeden Kamera verfügbar.

Für Bildpunkte (Korrespondenzen) ${}^{(j)}x$ und deren unbekanntes Urbild X im Raum gilt ${}^{(j)}\tilde{x} \sim {}^{(j)}P \tilde{X}$, folglich

$${}^{(j)}\tilde{\mathbf{x}} \times ({}^{(j)}P \tilde{\mathbf{X}}) = 0. \quad (4)$$

Da aber ein Bildpunkt nur zwei Koordinaten hat, kann er höchstens zwei unabhängige Gleichungen liefern, die für die Triangulation verwendbar sind. Wenn man nun die Abbildungsmatrix der j -ten Kamera aufschreibt in der Form

$${}^{(j)}P = \begin{pmatrix} {}^{(j)}\mathbf{p}_1^T \\ {}^{(j)}\mathbf{p}_2^T \\ {}^{(j)}\mathbf{p}_3^T \end{pmatrix},$$

dann wählt man aus dem Vektorprodukt (4) z. Bsp. die ersten beiden Zeilen und erhält als Ausgleichsproblem das homogene lineare Gleichungssystem $A \tilde{\mathbf{X}} = 0$ mit

$$A = \begin{pmatrix} {}^{(1)}\mathbf{p}_1^T - {}^{(1)}\tilde{x}_1 {}^{(1)}\mathbf{p}_3^T & {}^{(1)}\tilde{x}_2 {}^{(1)}\mathbf{p}_3^T - {}^{(1)}\mathbf{p}_2^T \\ {}^{(1)}\tilde{x}_2 {}^{(1)}\mathbf{p}_3^T - {}^{(1)}\mathbf{p}_2^T & \vdots \\ {}^{(n)}\mathbf{p}_1^T - {}^{(n)}\tilde{x}_1 {}^{(n)}\mathbf{p}_3^T & {}^{(n)}\tilde{x}_2 {}^{(n)}\mathbf{p}_3^T - {}^{(n)}\mathbf{p}_2^T \end{pmatrix}.$$

Gesucht ist $\tilde{\mathbf{X}} \neq (0,0,0)^T$ als Lösung im Sinne kleinster Fehlerquadrate. Dies wird erreicht mit Hilfe der Singulärwertzerlegung,

$$A = UDV^T,$$

und $\tilde{\mathbf{X}}$ ist die zum betragsmäßig kleinsten Singulärwert gehörende Spalte von V . Zwischen den rekonstruierten Punkten \mathbf{X} kann man nun leicht den Euklidischen Abstand berechnen.

3 Ergebnisse

Zum Testen des Verfahrens wurden Messungen in verschiedenen Szenen durchgeführt. Dabei haben sich vor allem die Winkel zwischen den optischen Achsen der Kameras sowie zwischen optischer Achse und der Ebene des Referenzvierecks als einflußreiche Größen gezeigt. Mit dem ersten Parameter für den Fehler (Winkel zwischen optischen Achsen) ist jede photogrammetrische Methode konfrontiert, der zweite ist verfahrensspezifisch.

In einer Straßenszene (s. Abb. 3) wurde ein



Abb. 3: Vermessung in einer Straßenszene⁵, ermittelte Werte (relativer Fehler bzgl. Ground-Truth-Daten), **A**: 305,7 cm (2,2 %), **B**: 152,3 cm (2,2 %), **C**: 49,4 cm (2,7 %)

⁵ Mit freundlicher Unterstützung der Fiebig GmbH Jena – Vertragshändler der BMW AG

Fahrzeug mit ca. 10 m Abstand zur Kamera in zwei Bildern vermessen. Der Winkel zwischen den optischen Achsen lag bei $13,8^\circ$, die Winkel zwischen den optischen Achsen und dem Viereck bei $12,2^\circ$ bzw. $16,9^\circ$. Diese Werte erwiesen sich als für die Vermessung suboptimal und bedingten relative Fehler von 2% bis 3%.

Es kann jedoch gezeigt werden, daß bei günstigen Bedingungen der relative Meßfehler unter 1% liegt (s. Abb. 4). Diese günstigen Bedingungen sind praktisch leicht zu realisieren, indem das Referenzviereck aufgerichtet und damit aus steilem Winkel ($>30^\circ$) abgebildet wird und zwischen mindestens zwei Aufnahmen eine ausreichend große Bewegung stattfindet (ebenso $>30^\circ$ zwischen optischen Achsen bei Ausrichtung auf das Objekt). Die in Abb. 4 dargestellten Fehlerwerte wurden erhalten, indem ein 70,2 cm großes Objekt in einem Abstand von 7,76 m zur ersten Kamera in jeweils zwei Bildern vermessen wurde.

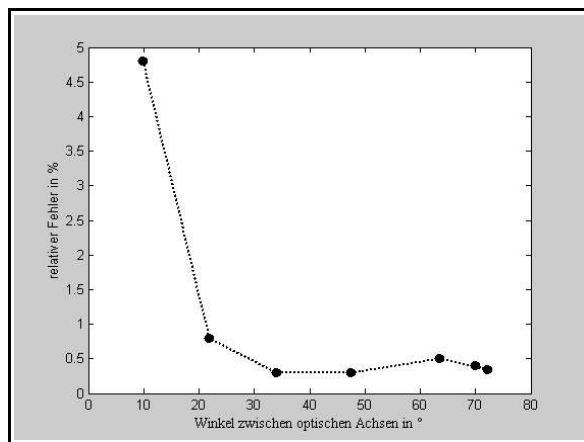


Abb. 4: Vermessung in zwei Bildern mit unterschiedlichen Kamerapositionen

4 Zusammenfassung

Das hier vorgestellte Verfahren hat zum Ziel, Spezialwissen anzuwenden, um metrische 3-D-Vermessung mit absolut minimaler Interaktion am Einsatzort und bei der Nachbereitung am Rechner zu realisieren, dabei ma-

thematisch stabile und gegen Fehler robuste Ergebnisse zu liefern.

Dieses Ziel wird durch die Separierung des Problems erreicht (mon. Rekonstruktion – Euklidische Transformation – Rekonstruktion). Durch diese Zerlegung erhält man mathematisch voll beherrschbare Teilprobleme, kann diese direkt und stabil lösen (jeweils Lösungen im Sinne kleinster Fehlerquadrate) und beherrscht damit das gesamte Problem. Hier unterscheidet sich das Verfahren von vergleichbaren, z. Bsp. ZHANG, 2000. Die benötigten Eingaben (Kameraparameter, Koordinaten des Vierecks) müssen nur einmal im Labor ermittelt werden. Die Interaktion vor Ort beschränkt sich auf das Plazieren des Referenzvierecks und ist damit minimal.

5 Literatur

- K. VOSS, R. NEUBAUER, M. SCHUBERT, 1995: Monokulare Rekonstruktion für Robotvision – Verlag Shaker, Aachen, S. 76 ff.
- R. HARTLEY, A. ZISSERMAN, 2002: Multiple View Geometry in Computer Vision, Second Edition – Cambridge University Press, Cambridge, S. 312 f.
- E. TRUCCO, A. VERRI, 1998: Introductory Techniques for 3D Computer Vision – Prentice-Hall, New Jersey
- O. FAUGERAS, 1996: Three-Dimensional Computer Vision: A Geometric Viewpoint – MIT Press, Cambridge
- M. W. WALKER, L. SHAO, R. A. VOLZ, 1991: Estimating 3D location parameters using dual number quaternions – Computer Vision, Graphics and Image Processing: Image Understanding, Vol. 54, S. 358 – 367
- Z. ZHANG, 2000: A flexible new technique for camera calibration – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22(11): S. 1330-1334