Stereobasierte Videosensorik unter Verwendung einer stochastischen Zuverlässigkeitsanalyse

A. Suppes, S. Niehe, M. Hötter, E. Kunze

Fachhochschule Hannover, Fachbereich Elektrotechnik, Projekt AMIS,
Ricklinger Stadtweg 120, D-30459 Hannover
alexander.suppes@etech.fh-hannover.de

Abstract. The new stereo-based computer vision system presented here enables a robot to automatically navigate in an unknown environment by detecting obstructions. Development is aiming at a cheap, robust sensor which, apart from measuring object distance and direction, has the ability to judge and verify the validity of estimated data. Based on the assumption of planar robot motion, stochastical disparity measurement techniques are applied to make it insensitive to changes of illumination and contrast as well as to reflections and shadows. The software based technique runs on a standard PC in real time (about 5 Hz) and shows promising results.

1 Einleitung

Die fortschreitende Automatisierung in der heutigen Arbeitswelt erfordert von den eingesetzten Maschinen und Robotern im zunehmenden Maße autonomes Handeln.

Im Bereich der automatisierten Fördertechnik werden immer mehr Transportaufgaben durch autonome mobile Roboter ausgeführt. Für eine automatische Navigation erfaßt und vermißt eine leistungsfähige Sensorik die aktuelle Umgebung, um so eine Positionsbestimmung und Hinderniserkennung vornehmen und daraus den Fahrweg bestimmen zu können. Anwendung finden hierbei im wesentlichen Radarsysteme, Laserscanner, Ultraschallsysteme und optische Kameras mit Videoauswertung [8].

Die Videosensorik besitzt dabei wichtige Vorteile:

- keine künstlichen Landmarken notwendig,
- rein passive Sensorik ohne gegenseitige, störende Beeinflußung bei mehreren mit diesen Systemen ausgerüsteten mobilen Robotern,
- sehr preisgünstige Realisierung durch Standardkomponenten,
- leichte Überwachungsmöglichkeit durch Übertragung der Kamerabilder.

Das hier vorgestellte System ist eine stereobasierte Videosensorik, die eine Szene mit zwei Kameras aus unterschiedlichen Blickwinkeln aufnimmt und mit Hilfe der Triangulation eine Vermessung von relevanten Objekten ermöglicht, siehe z.B. [4]. Bei der 3D-Vermessung wird ein stochastisches Verfahren zur Disparitäts-schätzung [6] verwendet, welches sich robust gegen Störgrößen wie Helligkeitsänderungen, Schattenwürfe oder Spiegelungen verhält. Weiterhin erlaubt

der Ansatz eine automatische Überprüfung sowohl der Zuverlässigkeit der ermittelten Meßwerte als auch der Genauigkeit des gesamten Meßverfahrens beim Durchfahren und Vermessen bekannter 3D-Geometrie.

Es wird im folgenden das Gesamtkonzept der Videosensorik mit dem Schwerpunkt der Zuverlässigkeitsanalyse bei der Meßwerterfassung vorgestellt.

In Kapitel 2 wird zunächst das Gesamtsystem der Videosensorik kurz präsentiert, deren Komponenten in Kapitel 3 im Einzelnen beschrieben und diskutiert werden. In Kapitel 4 werden einige erste Ergebnisse der Leistungsfähigkeit der Videosensorik vorgestellt und weitere Arbeiten motiviert, um die bereits erzielte Robustheit und Zuverlässigkeit weiter zu erhöhen. Zusammenfassung und Ausblick finden sich in Kapitel 5.

2 Gesamtkonzept

Das Blockdiagramm der Videosensorik ist im Bild 1 dargestellt. Die analogen Videodaten zweier synchronisierter Schwarz-Weiß-CCD-Kameras werden über eine Framegrabberkarte digitalisiert und im Hauptspeicher eines Rechners abgelegt. Aus je einem Bildpaar wird in der Videoverarbeitung durch Rektifizierung, Texturanalyse, Disparitätsschätzung, 3D-Rekonstruktion und Segmentierung ein 3D-Modell der aufgenommen Szene berechnet.

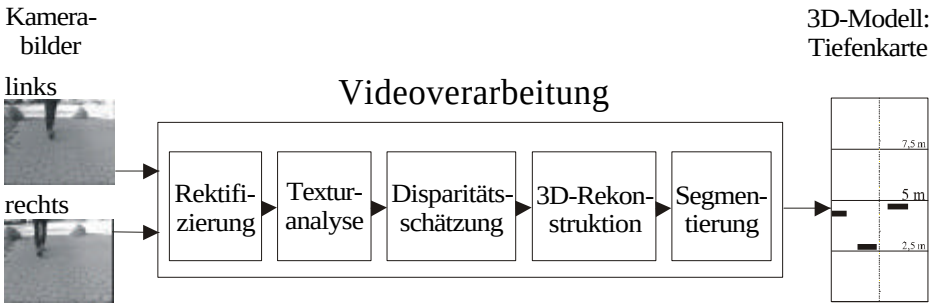

Bild 1 : Gesamtkonzept einer stereobasierten Videosensorik

Bei der Rektifizierung werden die digitalen Bilder der Kamera auf eine normierte Darstellung umgerechnet, welche die Meßwertverarbeitung von der tatsächlichen Kameraanordnung entkoppelt und dabei eine möglichst stabile und effiziente Berechnung des 3D-Modells ermöglicht.

Im nächsten Verarbeitungsschritt werden unter Verwendung einer Texturanalyse die Bildbereiche markiert, die ausreichende lokale Grauwertänderungen beinhalten und damit eine robuste und zuverlässige Disparitätsschätzung ermöglichen.

Die Disparitätsschätzung sucht korrespondierende Bildpunkte im Stereobildpaar. Ein Punkt aus der dreidimensionalen Szene wird aufgrund der unterschiedlichen Blickwinkel der beiden Kameras auf unterschiedliche Bildkoordinaten im rechten und linken Bild, die Punktkorrespondenzen, abgebildet. Diese Punktkorrespondenzen werden durch einen blockbasierten Vergleich lokaler Intensitätsverteilungen zwischen

den Bildern berechnet. Für den Vergleich wird hier ein Ähnlichkeitsmaß benutzt, welches unter Verwendung wahrscheinlichkeitstheoretischer Interpretationen auch Aussagen über die Zuverlässigkeit und Güte der gefundenen Punktkorrespondenzen erlaubt.

Unter Verwendung der Disparitätsschätzung wird aus den Grauwertbildern der beiden Kameras ein Disparitätsvektorfeld ermittelt. Die Punktkorrespondenzen beschreiben dabei die Differenz der Bildpunktkoordinaten eines abgebildeten 3D-Punktes auf das rechte und linke Kamerabild als einen Vektor, den sogenannten Disparitätsvektor. Im Weiteren werden für die 3D-Rekonstruktion ausschließlich das berechnete Disparitätsvektorfeld und dessen Zuverlässigkeitsanalyse verwendet.

Ausgehend von dem Disparitätsvektorfeld und der bekannten Kameraanordnung erfolgt durch Triangulation eine 3D-Rekonstruktion. Für jeden sicher geschätzten Disparitätsvektor, d.h. für jede zuverlässige Punktkorrespondenz, werden die Koordinaten des zugehörigen Punktes in der dreidimensionalen Szene zurückgerechnet.

Die Segmentierung faßt die berechneten 3D-Punkte mit Hilfe eines Abstandsmaßes zu Objekten zusammen und stellt diese Objekte in einer Tiefenkarte dar.

Im Folgenden werden die einzelnen Verarbeitungsschritte, die aus den beiden Kamerabildern ein 3D-Modell der Szene generieren, ausführlich dargestellt.

3 Systembeschreibung

3.1 Rektifizierung

Der Stereosensor besteht aus zwei Kameras, die so ausgerichtet sind, daß sie einen möglichst großen gemeinsamen Blickbereich haben (Bild 2a). Jede einzelne Kamera wird durch das Lochkameramodell angenähert, wobei die Abbildung jedes 3D-Punktes auf die Chipebene durch die Zentralperspektive erfolgt. Dieses Modell wird um eine radialsymmetrische Verzerrung erweitert, damit die nichtidealen Abbildungseigenschaften der Optik kompensiert werden können. Ausgehend von diesem Modell wird ein Verfahren zur Kamerakalibrierung verwendet [7], das sich aufgrund seiner hohen Stabilität und Robustheit für unterschiedliche Kameratypen bewährt hat.

Die so kalibrierte Kameraanordnung wird auf die sog. Standardanordnung gemäß Bild 2b umgerechnet. Punktkorrespondenzen liegen dann in den umgerechneten Bildern auf einer horizontalen Linie, der Epipolarlinie [3], d.h. der Disparitätsvektor enthält nur eine horizontale Komponente. Damit wird durch die Rektifizierung eine robuste und schnelle Disparitätsschätzung unterstützt (s. Kap. 3.3).

Bild 2 : Rektifizierung: a) reale Kameraanordnung b) Standardanordnung

3.2 Texturdetektion

Textur wird hier durch den örtlichen Grauwertgradienten beschrieben. Um eine zuverlässige Disparitätsschätzung zu ermöglichen, sind lokale Gradienten erforderlich. Da die Disparitätsschätzung ausschließlich in horizontaler Richtung (entlang der Epipolarlinie) erfolgt, wird das Vorhandensein vertikaler Strukturen durch einen Schwellwerttest für einen horizontalen Gradienten $T(X_L, Y_L)$ innerhalb eines Meßfensters überprüft. Für diesen gilt:

$$T(X_L, Y_L) = \frac{1}{N} \sum_{i,j} \left(s(X_L + i - 1, Y_L + j) - s(X_L + i + 1, Y_L + j) \right)^2 \qquad (i, j) \in \textit{Meßfenster} \tag{1}$$

mit N : Anzahl der Pixel im Meßfenster,
$s(i, j)$: Grauwert an der Pixelposition (i, j).

Unterschreitet der lokale Gradient $T(X_L, Y_L)$ eine vorgegebene Schwelle, wird der untersuchte Bildpunkt (X_L, Y_L) von der weiteren Verarbeitung ausgeschlossen, anderenfalls eine Disparitätsschätzung durchgeführt.

3.3 Disparitätsschätzung

Zur Disparitätsschätzung werden blockbasierte Verfahren eingesetzt, bei denen die Grauwertverteilung in rektifizierten Bildern direkt zur Bestimmung der Punktkorrespondenzen herangezogen wird.

Da die Disparitätsschätzung auf den rektifizierten Bildern durchgeführt wird, beschränkt sich die Suche von Punktkorrespondenzen auf die Suche entlang der horizontalen Epipolarlinie (Bild 3a, b). Den Arbeitspunkt, um den gesucht wird, bildet dabei die Disparität, die entstehen würde, wenn der betrachtete Bildpunkt auf der Ebene, auf der sich der Roboter bewegt, liegt und die aufgrund der bekannten Kamerageometrie für jeden Bildpunkt bekannt ist. Die eindimensionale Suche innerhalb eines begrenzten Bereiches der Epipolarlinie um den Arbeitspunkt spart erheblich Rechenzeit und eliminiert Mehrdeutigkeiten, was zur Erhöhung der Stabilität und Robustheit des Gesamtsystems führt.

Bei der hier eingesetzten blockbasierten Disparitätsschätzung wird ein rechteckiger Bildausschnitt (Meßfenster) an der Position $\vec{\mathbf{X}}_L = (X_L, Y_L)^T$ im Referenzbild (linkes Bild) mit Bildausschnitten an jeder Position $\vec{\mathbf{X}}_R = (X_R, Y_R = Y_L)^T$ im Suchbereich innerhalb des rechten Bildes durch Berechnung eines Gütemaßes verglichen (siehe Bild 3a, b). Einen Überblick über die im Bereich “Maschinelles Sehen” eingesetzten Gütemaße findet man z.B. in [2].

Bei den meisten für die Disparitätsschätzung benutzten Verfahren wird die Position $X_{R,\min}$, an der das Gütemaß die beste Übereinstimmung der Bildausschnitte zeigt, als der “wahre” zu X_L korrespondierende Punkt angenommen [2]. Weil aber die beiden Kameras der Stereoanordnung die Szene aus unterschiedlichen Blickwinkeln betrachten, kann diese Methode zu Fehlschätzungen in bestimmten Bildbereichen

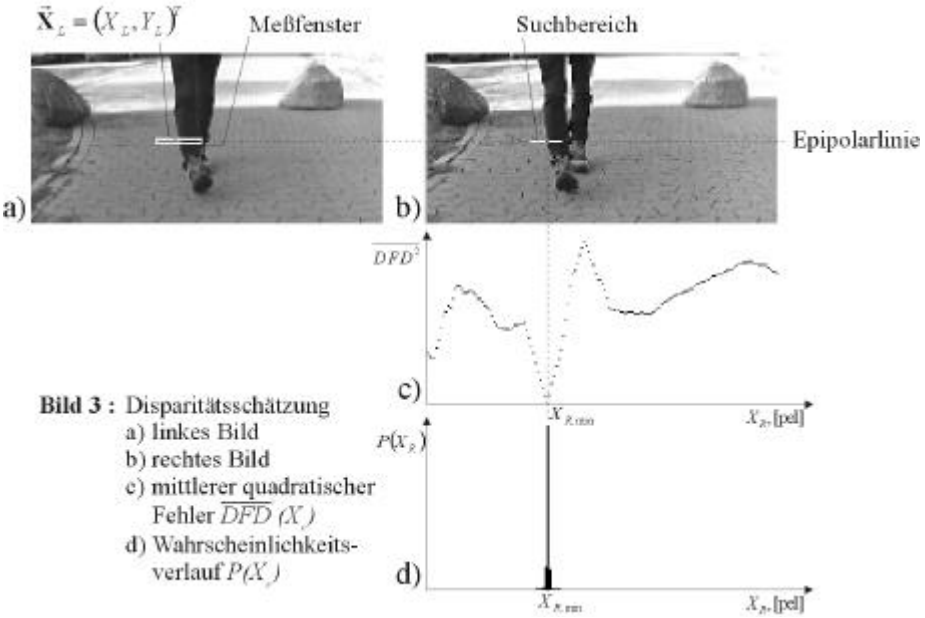

Bild 3 : Disparitätsschätzung
a) linkes Bild
b) rechtes Bild
c) mittlerer quadratischer Fehler $\overline{DFD}\,(X_R)$
d) Wahrscheinlichkeits-verlauf $P(X_R)$

führen, wie z.B. bei Verdeckungen. Durch die Anwendung eines stochastischen Verfahrens zur Disparitätsschätzung [6] ist es möglich, neben der Disparität noch eine Aussage darüber zu erhalten, wie sicher die Disparität geschätzt werden konnte. Somit lassen sich die unsicher geschätzten Punktkorrespondenzen von der weiteren Auswertung ausschließen, was die Stabilität des Gesamtsystems erhöht. Dieses Verfahren wird im Folgenden am Beispiel des mittleren quadratischen Fehlers $\overline{DFD^2}(X_R)$ als Gütemaß skizziert, der für jede Position X_R innerhalb des Suchbereiches wie folgt berechnet wird:

$$\overline{DFD^2}(X_R) = \frac{1}{N}\sum_{i,j}\left(s_L(X_L+i, Y_L+j) - s_R(X_R+i, Y_L+j)\right)^2;\ (i,j) \in \textit{Meßfenster}, \qquad \textbf{(2)}$$

mit N : Anzahl der Pixel im Meßfenster,
$s_L(i,j)$: Grauwert an der Pixelposition (i,j) im linken Bild,
$s_R(i,j)$: Grauwert an der Pixelposition (i,j) im rechten Bild.

Im Bild 3c ist ein typischer Verlauf von $\overline{DFD^2}(X_R)$ dargestellt. Das Hauptminimum mit dem kleinsten quadratischen Fehler $\overline{DFD^2}_{\min}$ tritt an der Stelle $X_{R,\min}$ auf. Um kleinere Nebenminima, die z.B. aufgrund von periodischen Strukturen im Bild auftreten können und leicht zu Fehlinterpretationen führen, von der weiteren statistischen Analyse auszuschließen, wird im Folgenden nur noch ein kleiner symmetrischer Ausschnitt $[X_{R,\min} - b,\ X_{R,\min} + b]$, $b \in N$, des Suchbereiches um $X_{R,\min}$ betrachtet.

Nach [6] läßt sich der Ausdruck

$$P'(X_R) = \exp\left(-\overline{DFD^2}(X_R) \Big/ \overline{DFD^2}_{\min}\right) \tag{3}$$

als ein Maß für die Wahrscheinlichkeit auffassen, mit welcher der Punkt X_R aus dem betrachteten Ausschnitt den "wahren" zu X_L korrespondierenden Punkt darstellt. Aus $P'(X_R)$ wird die Wahrscheinlichkeit $P(X_R)$ durch die Normierung berechnet:

$$P(X_R) = P'(X_R) \Big/ \sum_{X_R = X_{R,\min}-b}^{X_{R,\min}+b} P'(X_R). \tag{4}$$

Auf dem Bild 3d ist der Wahrscheinlichkeitsverlauf $P(X_R)$ dargestellt.

Für den Erwartungswert $\hat{X}_R$ des zu X_L korrespondierenden Punktes gilt dann:

$$\hat{X}_R = E[X_R] = \sum_{X_R = X_{R,\min}-b}^{X_{R,\min}+b} X_R \cdot P(X_R). \tag{5}$$

Die Unsicherheit der Schätzung von $\hat{X}_R$ wird als Varianz berechnet:

$$s^2_{\hat{X}_R} = E\left[\left(X_R - \hat{X}_R\right)^2\right] = E\left[X_R{}^2\right] - E^2[X_R] = \sum_{X_R = X_{R,\min}-b}^{X_{R,\min}+b} X_R{}^2 \cdot P(X_R) - \hat{X}_R{}^2. \tag{6}$$

Durch eine einfache Schwellwertbetrachtung werden nur Punkte mit einer Schätzvarianz unterhalb eines vorgegebenen Schwellwertes als sichere Schätzungen erkannt und bei der 3D-Rekonstruktion berücksichtigt.

3.4 3D–Rekonstruktion und –Segmentierung

Aus den Punktkorrespondenzen der sicheren Schätzungen läßt sich der entsprechende 3D-Punkt mittels Triangulation bestimmen [3], [5], wodurch jeder sicher geschätzten Disparität ein Objektpunkt im 3D-Raum zugeordnet wird. Die Aufgabe der Segmentierung besteht nun darin, diese Objektpunkte zu Objekten zusammenzufassen. Punkte, deren dreidimensionale Abstände unterhalb eines vorgegebenen Schwellwertes liegen, werden zu einem Objekt zusammengefaßt.

Für die Navigation von Fahrzeugen werden hier die Vorderkanten der detektierten Objekte als Hindernisse in einer Tiefenkarte dargestellt.

4 Experimentelle Ergebnisse

Das hier vorgestellte Verfahren wurde unter unterschiedlichen Bedingungen erprobt. Der Abstand zwischen den Kameras betrug ca. 60 cm, die Brennweite 4 mm bei einem 1/4"-Chip.

In einem vorab vermessenen Testfeld wurde die bei der Objektvermessung erzielte 3D-Genauigkeit überprüft. Bild 4a zeigt ein Bildpaar vom Testfeld sowie die vom Algorithmus errechnete Tiefenkarte mit Objekten. Ein Vergleich zwischen gemessenen und berechneten Objektkoordinaten zeigt eine Genauigkeit der 3D-Rekonstruktion von unter 3% bis zu einer Entfernung von 10 m.

Bild 4b zeigt ein in natürlicher Umgebung aufgenommenes Bildpaar sowie die daraus resultierende Tiefenkarke mit vermessenen Objekten. Die statischen und dynamischen Objekte wurden robust erfaßt und mit einer Genauigkeit <3% vermessen. Die Verarbeitungsrate beträgt auf einem PentiumII Rechner mit 500 MHz bei einer Bildgröße 186x137 pel (QCIF) etwa 5 Hz, was die Eignung des hier vorgestellten Stereosensors zur Roboternavigation zeigt.

Fehldetektionen treten z.B. bei Verdeckungen aufgrund der unterschiedlichen Blickwinkel der beiden Kameras auf. Durch die zeitliche Verküpfung von Informationen aufeinander folgender Bildpaare (Objektverfolgung), lassen sich diese Fehldetektionen vermeiden.

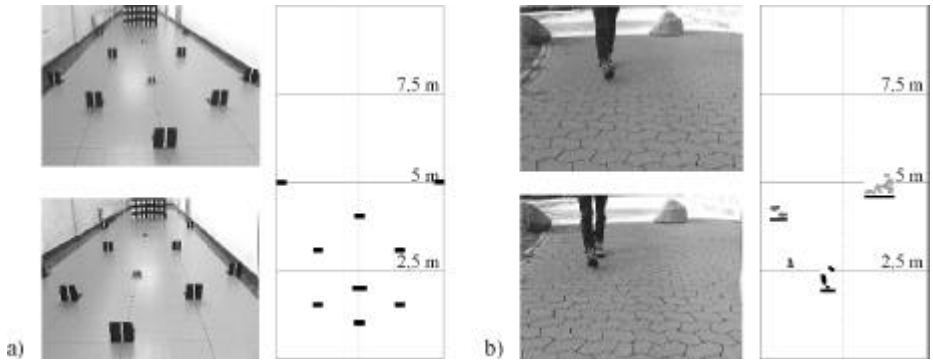

Bild 4 : Tiefenkarte für a) Testfeld und b) natürliche Umgebung

5 Zusammenfassung und Ausblick

Die vorgestellte stereobasierte Videosensorik ermöglicht es mobilen Robotern, ihre Umgebung automatisch zu erfassen und zu vermessen, um so eine Positionsbestimmung und Hinderniserkennung vornehmen und Transportaufgaben z.B. in Fertigungsstätten und im Servicebereich autonom ausführen zu können.

Durch den Einsatz von Standardkomponenten aus dem PC- und Multimediabereich wird ein kostengünstiges Sensorsystem realisiert. Eine einfache Adaption an verschiedene Anwendungen wird durch den rein softwarebasierten Ansatz ermöglicht.

Der hier vorgestellte Ansatz zeigt bei den Ergebnissen eine robuste Detektion und Vermessung von Objekten auch bei Störeinflüssen wie Helligkeitsänderungen, Schattenwürfe oder Spiegelungen. Durch stochastische Methoden erfolgt automatisch eine Überprüfung der Zuverlässigkeit für jeden erzielten Meßwert und damit auch des

gesamten Sensorsystems, was eine Früherkennung von Fehlfunktionen, wie z.B. zeitliche Veränderungen der Kameranordnung o.ä., erlaubt.

Zur Zeit werden drei weitere Entwicklungen zur Verbesserung bzw. Erweiterung der bestehenden Videosensorik verfolgt.

Zur Detektion von Objekten wird die Kenntnis der relativen Lage der Stereokameraanordnung zur Fahrebene benutzt. Im laufenden Betrieb kann sich diese Zuordnung z.B. beim Anfahren auf eine Rampe oder beim Fahren über kleinere Unebenheiten ändern. Um hier fehlerhafte Detektionen zu verhindern, wird die Bewegung der Stereokameraanordnung relativ zur Fahrebene in weiteren Arbeiten mitberücksichtigt.

Der hier vorgestellte Ansatz wertet immer nur ein Stereobildpaar aus. Durch Ausnutzen der Information aus der zeitlichen Abfolge der Stereobildpaare können Vorhersagen für das aktuell zu untersuchende Bildpaar getroffen werden. Damit wird das Meßverfahren schneller und die Meßwerterfassung bei Störeinflüssen und Mehrdeutigkeiten noch stabiler und robuster.

Ein weiterer Schwerpunkt künftiger Aktivitäten wird die Integration komprimierender Videoübertragungsverfahren und die Anbindung zu einer Schaltzentrale über Standardnetze (GSM, ISDN) sein. Der Empfänger kann dann direkt das Umfeld des mobilen Roboters visuell erfassen. Neben dem Monitoring des Roboters erschließt dies völlig neue Anwendungen z.B. im Bereich der Überwachungs- und Sicherheitstechnik oder bei der autonomen Erkundung von für den Menschen gefährlichen Umgebungen.

6 Literaturverzeichnis

[1] T. Aach, *Bayes-Methoden zur Bildsegmentierung, Änderungsdetektion und Verschiebungsvektorschätzung*, Diss. an der RWTH Aachen, VDI-Verlag, 1993.

[2] P. F. Aschwanden, *Experimenteller Vergleich von Korrelationskriterien in der Bildanalyse*, Diss. an der ETH Zürich, 1993.

[3] N. Ayache, *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*, MIT Press, 1991.

[4] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, 1993.

[5] R. Koch, *Automatische Oberfläfenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten*, Diss. an der Universität Hannover, VDI-Verlag, 1997.

[6] R. Mester, M. Hötter, „Robust Displacement Vector Estimation Including a Statistical Error Analysis“, in *Image Processing and its Applications*, 4-6 July, 1995, Conference Publication, pp 168-172.

[7] R.Y. Tsai, „A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using off-the-Shelf TV Cameras and Lenses“ in *IEEE Journal of Robotics and Automation*, Vol. RA-3, No.4, August 1987, pp. 323-344.

[8] *Automatisierung mit Augenmaß*, Tagungsband der Dritten Duisburger FTS – Fachtagung, 9. März 1995, Gerhard-Mercator-Universität Duisburg, Fertigungstechnisches Labor.