

Sequentielle Schätzung von Sensorpositionen aus Bildsequenzen für Robot Vision Anwendungen

Th. Kersten, E. Baltsavias

Institut für Geodäsie und Photogrammetrie, ETH Zürich
ETH-Hönggerberg, CH-8093 Zürich, Schweiz

Kurzfassung

Unter einer Vielzahl von Sensortechnologien sind mobile Visionsysteme in der Lage, ihre Beziehung und Orientierung bezüglich ihrer Umgebung aus Bildsequenzen abzuleiten. In diesem Beitrag präsentieren wir einen Algorithmus/Methode, um für mono- bzw. binokulare vision-basierte mobile Systeme die Sensororientierungen aus Bildsequenzen zu bestimmen. In zwei praktischen Tests wird die Datenakquisition, die Extraktion charakteristischer Punkte in den Bildern, sowie das Tracking und Matching dieser Punkte entlang der Sequenz beschrieben. Die Bestimmung der Sensorpositionen erfolgt durch sequentielle Schätzungen basierend auf Givens Transformationen. Voraussetzung für genaue 3-D Positionierungen sind präzise und zuverlässige Kalibrierungen der Visionsysteme. Die Resultate der Kamerakalibrierungen und der Sensororientierungen für zwei praktische Beispiele werden präsentiert.

1. Einführung

Robot Vision Systeme sind wichtige Bestandteile moderner Roboter. Sie haben die Fähigkeit, ihre Umgebung wahrzunehmen, in ihr zu messen, bedingt Schlußfolgerungen daraus abzuleiten und die Informationen an Führungssysteme weiterzuleiten. Daher erfordern die dynamischen Aufgaben eines Roboters ständigen "feedback" vom Visionsystem, das durch vereinfachte Bildverarbeitungsfunktionen oder/und durch Implementation der entsprechenden Algorithmen in Hardware erreicht werden kann. Direkte Rückmeldung ("robot guidance") verlangt eine schnelle Verarbeitung der Bilddaten, um durch Analysen und 3-D Beschreibungen der Roboterumgebung rechtzeitig agieren zu können.

Für zeitkritische Anwendungen solcher Robot Vision Systeme, bei denen eine Verarbeitung von Bildsequenzen und hochgenaue 3-D Positionierung der Objekte und Sensoren durchgeführt wird, sind sorgfältige Abstimmungen und optimale Wechselwirkung der Systemhardware, komplexe Algorithmen und entsprechend optimierte Software erforderlich. Da sich bei der Bildaufnahme und -verarbeitung Fehler nicht vermeiden lassen, sind neben einer robusten 3-D Positionierung auch Fehlersuche und -analyse wichtige Systemkomponenten. Anforderungen für ein photogrammetrisches Robot Vision System, wie z.B. praktikable Geschwindigkeit, Leistung und Fehlertoleranz sind in [8] diskutiert. Die erforderliche Zeit für die Verarbeitung von Bildsequenzen, um solche Operationen wie die Extraktion charakteristischer Punkten, das Verfolgen der Punkte (Tracking) und die Bestimmung der Sensorpositionen durchzuführen, hängt natürlich von den Anforderungen jeder Applikation ab. Für visuelle Echtzeitnavigation sollten alle Operationen für ein Bild in weniger Zeit, als für die Bildaufnahme zur Verfügung steht (1/25 Sekunden), erfolgen. Um solche Robot Vision Anwendungen durchführen zu können, müssen einige Operationen durch Hardware, andere durch Software

erfolgen. Spezielle Algorithmen, die bestimmte Aufgaben erledigen, sind als Hardwaremodule bis zu 1000 mal schneller als reguläre Software.

Obwohl wir uns bewußt sind, daß sehr schnelle Lösungen von Positionierungsaufgaben in Videoechtzeit trotz leistungsfähiger Computerhardware immer noch grundlegende Vereinfachungen der Messungen erfordern, möchten wir in unserem Beitrag eine generelle Lösung einer sequentiellen Punktbestimmung und Sensorpositionierung, die auf einer photogrammetrischen Bündellösung [11] basiert, vorstellen. Der präsentierte sequentielle Schätzalgorithmus benutzt dabei die Givens Transformationen, um das obere Dreieck des reduzierten Normalgleichungssystem mit den Informationen der neuen Beobachtungen bzw. Messungen zu aktualisieren (update), aus dem die Objektpunkte und die Kameraorientierungsdaten als Unbekannte bestimmt werden. Wir verwenden die Givens Transformationen (GT) im Gegensatz zu dem Kalman Filter, der bei Analysen von Bildsequenzen hauptsächlich im Computer Vision Bereich ([6], [20], [15], [23]) eingesetzt wird, weil einerseits in der Robotik die Aktualisierung des Parametervektors nicht in jedem Stadium gebraucht wird und andererseits die Variation der Vektorgröße (Hinzufügen und Löschen von neuen Objektpunkten, Hinzufügen von Kameraorientierungen, etc.) zu schwachen Rechenleistungen des Kalman Filters führt. In der Vergangenheit wurden verschiedene sequentielle Schätzalgorithmen miteinander verglichen. Zuerst wurde der TFU-Algorithmus (Triangular Factor Update) leistungsfähiger als der Kalman-Filter hinsichtlich Rechenzeiten und Speicherplatzausnutzung bewertet ([10], [22]). Später zeigten Runge [21] und Holm [16], daß die GT in der Rechenleistung, Speicherplatzausnutzung und in der einfachen Handhabung und Softwareimplementation besser als der TFU-Algorithmus sind. Unser algorithmischer Ansatz zur Bestimmung der Sensorpositionen beinhaltet auch Selbstkalibrierungsparameter zur Modellierung systematischer Fehler und statistische Test für die grobe Fehlersuche. Blais [5] stellte die GT erstmals für photogrammetrische Anwendungen vor. Die sequentielle Schätzung in der Bündelausgleichung mit GT wird in [11], [21], [16] und [14] detailliert beschrieben. Den Einsatz von GT in einem mobilen Mapping-System, das mit einem Stereo-Visionssystem ausgestattet ist, stellen Edmundson und Novak [7] vor.

2. Photogrammetrische Punktbestimmung durch die Bündelmethode

Die Bündelausgleichung ist ein modernes und effizientes Verfahren der photogrammetrischen Punktbestimmung. Die mathematischen Grundlagen der Bündelausgleichung basieren auf dem Modell der Zentralprojektion. Die funktionalen Beziehungen der Bündelmethode lassen sich aus der Kollinearitätsbedingung ableiten, die in Parameterform die Gleichungen darstellt, in denen Bildpunkt, Projektionszentrum und Objektpunkt auf einer Geraden liegen. Für jeden Bildpunkt gibt es zwei Kollinearitätsgleichungen, in denen jeweils neun Unbekannte (3 Objektpunktkoordinaten und 6 Parameter der Kamerastation) zu bestimmen sind. Jede Beobachtung eines Kontrollpunktes (in der Photogrammetrie Paßpunkt genannt) ergibt eine zusätzliche Gleichung. Alle Gleichungen der Bild- und Paßpunkte führen zu einem System von Beobachtungsgleichungen. Zur Lösung des Gleichungssystems durch die Kleinste-Quadrate-Methode werden die Gleichungen linearisiert. Das linearisierte Gleichungssystem ist in dem Gauß-Markov-Modell formuliert als:

$$-e = Ax - l; P \quad ,$$

mit e als wahren Fehler, A als Designmatrix mit den partiellen Ableitungen der Gleichungen nach den Unbekannten, x als Parameter/Unbekanntenvektor, l als Absolutgliedvektor und P als Gewichtsmatrix für die Beobachtungen. Die simultane Lösung der Normalgleichungen und die Berechnung des Parametervektors x erfolgt durch Bündelausgleichung nach der Kleinste-Quadrate-Methode:

$$\hat{x} = (A^T P A)^{-1} A^T P l$$

Bei der sequentiellen Lösung wird das Gleichungssystem nicht relinearisiert und faktorisiert, sondern durch orthogonale Transformationen wird die Matrix des vorherigen Stadiums direkt mit den Daten der neuen Beobachtungen aktualisiert.

3. Algorithmischer Ansatz zur Bestimmung der Sensororientierung

Der Einsatz und die Funktionalität des präsentierten Algorithmus wird durch zwei praktische Beispiele aufgezeigt. In dem ersten Beispiel wurde eine einfach zu handhabende Standardvideokamera benutzt, um eine Bildsequenz von unserem 3-D Testfeld aufzunehmen, wobei die Kamera einen beweglichen Roboter simuliert. Die Bildaufnahme, die Bilddatendigitalisierung, die Punktmessungen durch Matching, die präzise Kamerakalibrierung und die Bestimmung der Sensorpositionen werden beschrieben. Im zweiten Beispiel wird ein Stereo-Visionssystem, das auf einer mobilen Plattform installiert ist, verwendet, um Bildsequenzen in einer Büroumgebung zu generieren. Die Bestimmung der Sensorpositionen läuft hier in einer zyklischen Prozedur ab (Abb. 1), die folgende Schritte beinhaltet: (1) Bildaufnahme und Digitalisierung, (2) Bildverarbeitung zur Extraktion charakteristischer Punkte (Landmarken), (3) Matching und Tracking dieser Punkte und (4) sequentielle Schätzung der Sensorpositionen mit anschließender grober Fehlersuche. Um die Genauigkeit unseres Ansatzes zu testen, wurde am Anfang und am Ende der Stereosequenz das 3-D Testfeld aufgenommen, so daß die errechneten Sensorpositionen zum Schluß mit Referenzwerten verglichen werden konnten. Für beide praktischen Beispiele werden Angaben über Rechenzeiten gemacht, die natürlich als Softwarelösungen keine Echtzeitverarbeitung bieten.

4. Eingesetzte Visionsysteme und Aufnahme der Bildsequenzen

Für die Aufnahme der Bildsequenzen in beiden praktischen Tests wurden ein mono- bzw. binokulares Visionsystem eingesetzt.

4.1 Monokulares Visionsystem

Im ersten Beispiel diente unser 3-D Testfeld (Abb. 2) als Aufnahmeobjekt für die Simulation eines mobilen Roboters. Aber an Stelle eines Roboters filmte ein Operateur, der sich mit einem JVC Videocamcorder GR-S77E (S-VHS) entlang des Testfeldes bewegte, das Testobjekt. Der Camcorder eignet sich für solche Aufnahmen sehr gut, da er sehr leicht zu bedienen ist, einen Sucher hat und auf einer internen Videokassette sehr lange Bildsequenzen speichern kann. Der Weg des mobilen Visionsystems während der Aufnahme ist in Abb. 3 abgebildet. Für die Aufnahme wurde der Autofokus abgeschaltet und die Kamera auf unendlich fokussiert. Mit einer Brennweite von 8.5 mm wurde die Schärfentiefe zur scharfen Abbildung des Objektes als genügend erachtet. Von der aufgenommen Bildsequenz wurden 53 Sekunden mit einem Videopix Framegrabber an einer SparcStation 1+ von SUN Microsystems digitalisiert. Die

generierten Bilder wurden während des Digitalierungsprozesses mit einer Tiefpaßfilter verbessert, um u.a. Bewegungseffekte zu kompensieren. Insgesamt wurden 88 Bilder digitalisiert, was eine Bildrate von 1.8 Bilder pro Sekunde ergab.

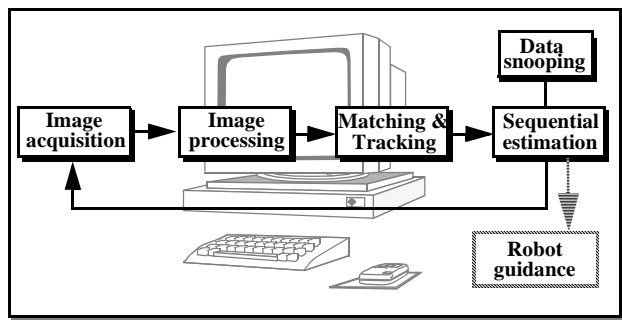


Abb. 1: Prozedur für die Bestimmung der Sensororientierung



Abb. 2: 3-D Testfeld

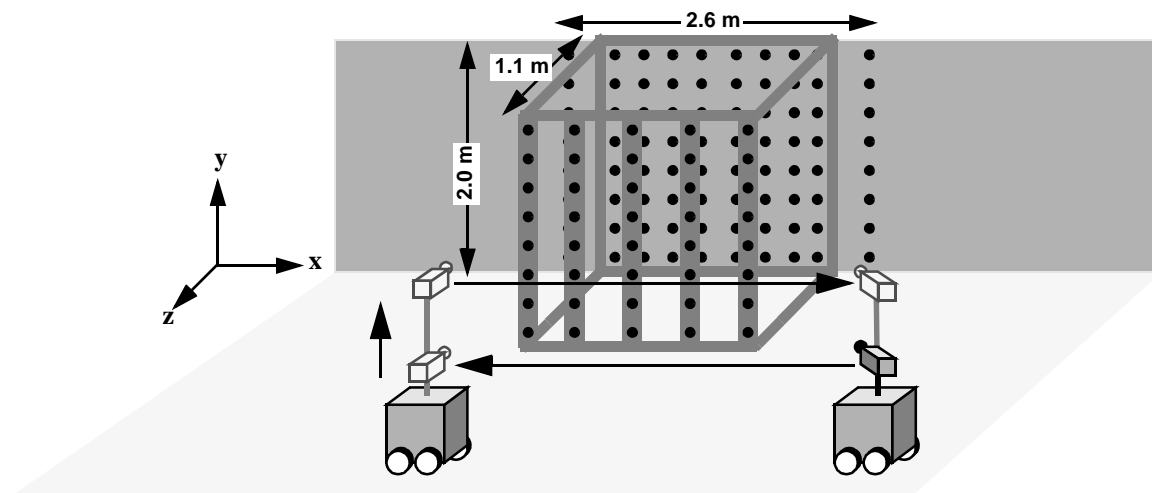


Abb. 3: Dimensionen des Testfeldes und Kameraweg bei der Aufnahme der JVC Sequenz

4.2 Binokulares Visionssystem

Die Kameras waren auf einer mobilen Plattform (Abb. 4) in einem Abstand von 56 cm zueinander befestigt. Die Aufnahme der Bildsequenzen erfolgte mit zwei synchronisierten Sony XC77CE CCD-Kameras (9 mm Fujinon-Objektiv) und die Bilddaten wurden auf zwei Videorekordern aufgezeichnet und gespeichert. Da eine Datenrate des Visionssystems von ca. 13 MB/Sekunde bei der Aufnahme anfällt und digitale Echtzeitspeichergeräte für schnelle Speicherung von langen Bildsequenzen noch zu teuer sind, wurden die Bildsequenzen auf zwei U-matic Videorekordern zwischenzeitlich gespeichert und später digitalisiert. Um korrespondierende Bilder (Frames) später von den Videobändern lesen und zuordnen zu können, erzeugte ein Bildmustergeber ein binäres Muster mit einer sequentiellen Codenummer in den Bildern.

Die Bewegung der Plattform wurde von einem Operateur durchgeführt und konnte auf einem Monitor überprüft werden. Die Plattform wurde durch eine Büroumgebung bewegt, wobei jeweils am Anfang und am Ende der aufgenommenen Sequenz das Testfeld abgebildet

wurde (Abb. 5). Die Bildsequenzen wurden an einem PC, der mit einem Matrox MVP Framegrabberboard ausgestattet war, automatisch in Videoechtzeit (25 Frames/sec) digitalisiert. Die Zwischenspeicherung von Bildsequenzen auf Videobändern verursachen gegenüber der direkten Digitalisierung einen Qualitäts- und Genauigkeitsverlust von ca. 30% [18]. Insgesamt wurden pro Sequenz 742 Bilder (Sequenzlänge 29.7 Sekunden) digitalisiert, die eine Speicherkapazität von 195 MB pro Kamera verlangten. Wegen der Bewegungseffekte in den Bildern, die Probleme bei der Extraktion charakteristischer Punkte verursachen, wurden für die weitere Verarbeitung der Sequenzen in diesem Fall nur die ungeraden Halbbilder verwendet.

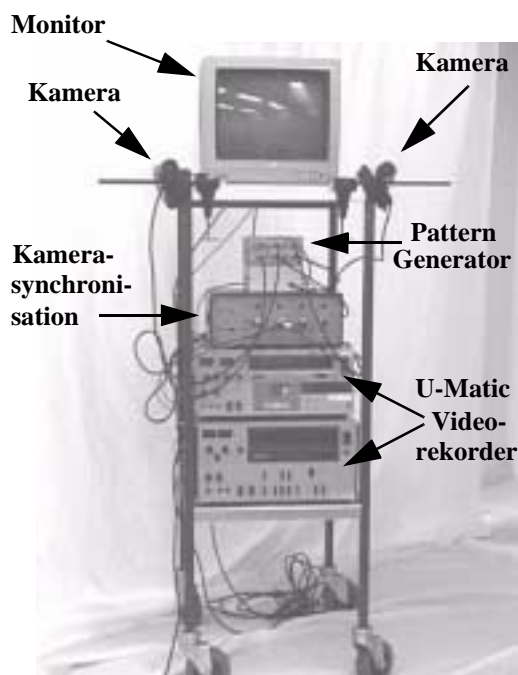


Abb. 4: Mobile Plattform mit Stereovision

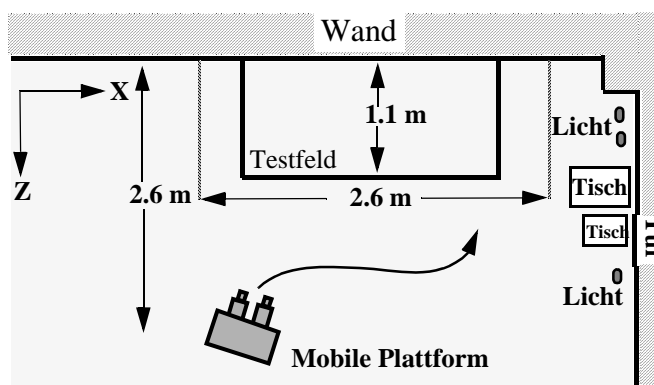


Abb. 5: Weg der mobilen Plattform

5. Kalibrierung der Visionsysteme

Vor der Verarbeitung der Bildsequenzen müssen die Kameras kalibriert werden. Eine Einführung in die photogrammetrische Kamerakalibrierung wird in [3] beschrieben. Die relevante Software wurde auch hier benutzt. Mit der JVC und den beiden CCD-Kameras (Sony) wurden von dem 3-D Testfeld aus vier verschiedenen Positionen Aufnahmen gemacht. Die Sony-Kameras wurden zusätzlich noch auf jeder Station um 90 Grad gedreht. Bei der Kalibrierung werden 9 zusätzliche Parameter bestimmt, die die innere Orientierung der Kamera (Kammerkonstante und Lage des Bildhauptpunktes) und andere systematische Fehler (Maßstabfaktor in x, Scherung, 2 Parameter für die radialsymmetrische und 2 Parameter für die tangentiale Verzeichnung) modellieren. In den digitalen Bildern wurden 130 (JVC) bzw. 123 (Sony) Zielmarken des Testfeldes mit dem Verfahren des Least Squares Template Matching (LSTM) [2] mit Subpixelgenauigkeit gemessen, während deren 3-D Objektkoordinaten durch Theodolitmessungen bestimmt wurden. Die Bildpunktmessungen wurden in einer Bündelausgleichung mit Selbstkalibrierung (zusätzliche Parameter als Unbekannte) ausgeglichen. Die Ergebnisse der Kamerakalibrierungen (Version 1 für JVC, Version 2 bis 4 für Sony) sind in Tabelle 1 zusammengefaßt. In Version 1 stehen die Ergebnisse der JVC-Kalibrierung und in Version 2 bis 4 die Ergebnisse der Kalibrierung einer Sony-Kamera. In Version 2 wurden die

Tabelle 1: Ergebnisse der Kamerakalibrierungen

Version	Ka	T	ZP	Ko	$\hat{\sigma}_0$ [μm]	Theoretische Genauigkeit aus Ausgleichung			Praktische Genauigkeit aus Vergleich mit Kontrollpunkten ¹⁾		
						Objektraum [mm]			Objektraum [mm]		
						σ_X	σ_Y	σ_Z	μ_X	μ_Y	μ_Z
1	JVC	1	9	67	0.84	0.29	0.27	0.62	0.39	0.31	0.75
2	Sony	1	9	62	0.73	0.15	0.15	0.37	0.26	0.20	0.41
3	Sony	2	9	62	0.97	0.19	0.19	0.49	0.29	0.21	0.54
4	Sony	1	0	62	6.57	1.31	1.31	3.35	3.15	3.47	5.93

Ka.....Kamera
 T.....Typ (1=Voll, 2=Halbbild)
 ZP.....Anzahl der Zusatzparameter
 Ko.....Anzahl der Kontrollpunkte
¹⁾ Werte nach einer räumlichen Ähnlichkeitstransformation aller Kontrollpunktkoordinaten auf ihre Referenzwerte.

σ_0 Standardabweichung der Gewichtseinheit a posteriori
 σ_{XYZ}Theoretische Genauigkeit im Objektraum
 μ_{XYZ}Empirische Genauigkeit im Objektraum

Vollbilder und in Version 3 nur die Halbbilder verwendet. Die Ergebnisse einer Kalibrierung ohne Modellierung der systematischen Fehler sind in Version 4 zusammengefaßt.

Für die Qualitätsbeurteilung der photogrammetrischen Punktbestimmung und Kalibrierung werden u.a. die theoretische und praktische Genauigkeit als Kriterien verwendet. Die theoretische Genauigkeit beschreibt die statistische Qualität der in eine Ausgleichung nach der Kleinst-Quadrate-Methode bestimmten Unbekannten/Parameter. Die Standardabweichungen der geschätzten Unbekannten (σ_X , σ_Y , σ_Z) werden durch die Inversion der Normalgleichungsmatrix und durch die Standardabweichung der Gewichtseinheit a posteriori $\hat{\sigma}_0$ berechnet. Die praktische Genauigkeit zeigt, wie nah die geschätzten Parameter den wahren Werten sind. Sie ist unabhängiges Maß für die Genauigkeit und kann durch Vergleich mit Referenzwerten (Kontrollpunkte) erhalten werden. Die praktische Genauigkeit (μ_X , μ_Y , μ_Z), die nach einer Ähnlichkeitstransformation auf die Kontrollpunkte berechnet wurde, zeigt, daß in der Kalibrierung eine Genauigkeit von besser als 1 Millimeter erreicht wurde.

Der Vergleich von Version 2 und 3 (Sony) zeigt, daß die Kalibrierung der Kamera mit Voll- bzw. Halbbildern nur einen geringfügigen Unterschied aufweist. Um den Einfluß der Zusatzparameter auf die Genauigkeit abzuschätzen, wurde die Version 4 (eine der beiden Sony-Kameras) ohne Selbstkalibrierung gerechnet. Das Ergebnis ist um den Faktor 10 schlechter als die Ergebnisse mit Selbstkalibrierung und zeigt, wie wichtig die Modellierung systematischer Fehler ist. Beyer hat in [4] nachgewiesen, daß die Zusatzparameter das Ergebnis bis zu einen Faktor 70 verbessern können. Im Bildraum ergaben die Kalibrierungen eine Genauigkeit von 1/10 (JVC), 1/15 (Sony-Halbbild) bzw. 1/20 (Sony-Vollbild) des Pixelabstandes.

6. Bestimmung der Sensororientierungen

6.1 Monokulares Visionssystem (JVC Camcorder)

Für die Camcorder-Sequenz wurden die Zielmarken des Testfeldes in allen 88 Vollbildern mit LSTM bestimmt. Durch Messungen von einigen gut verteilten Punkte im Bild konnte die Kameraposition durch Rückwärtsschnitt genähert berechnet werden. Mit diesen Näherungen und den bekannten 3-D Objektkoordinaten der restlichen Zielmarken konnten die genäherten Bild- bzw. Pixelkoordinaten rückprojiziert werden, so daß mit diesen Näherungen bildweise die anderen Landmarken (zwischen 79 und 146 Punkte pro Bild) automatisch gemessen werden

konnten. Die Messdaten wurden anschließend im On-Line Triangulations System OLTRIS [17] verarbeitet. OLTRIS bietet die Möglichkeit sowohl im sequentiellen Modus als auch simultan auszugleichen. Die Sensorpositionen wurden in der fortlaufenden Triangulation durch sequentielle Schätzung bestimmt, in dem das Normalgleichungssystem laufend durch neue Beobachtungen erweitert und an den entsprechenden Stellen aktualisiert wurde. Für die ganze Sequenz wurden 20860 Beobachtungen verarbeitet und 1026 Unbekannte durch sequentiellen Ausgleich bestimmt. Der berechnete Weg der bewegten Videokamera ist in Abb. 6 geplot-

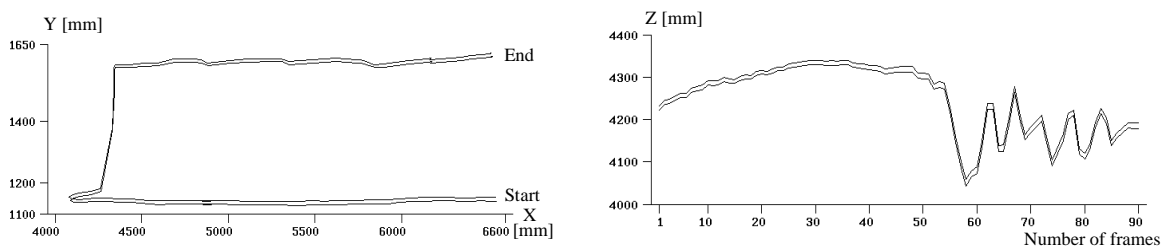


Abb. 6: Weg der JVC Videokamera

tet. Der untere Pfad ist durch OLTRIS bestimmt worden, während der obere durch eine simultane Ausgleichung mit Selbstkalibrierung berechnet wurde. Die Unterschiede beider Kamerawege sind 4.5 mm in x-, 11 mm in y- und 12 mm in z-Richtung, das durch die fehlende Kompensation der systematischen Fehler in der sequentiellen Schätzung zu erklären ist, die zu diesem Zeitpunkt in OLTRIS nicht implementiert war.

Wichtige Unterschiede zwischen sequentieller und simultaner Ausgleichung sind die unterschiedlichen Methoden für den "Update" des Normalgleichungssystems und Berechnung des Lösungsvektors, wie in Kapitel 2 erwähnt, und die Rechenzeiten (CPU). Abb. 7 zeigt die stei-

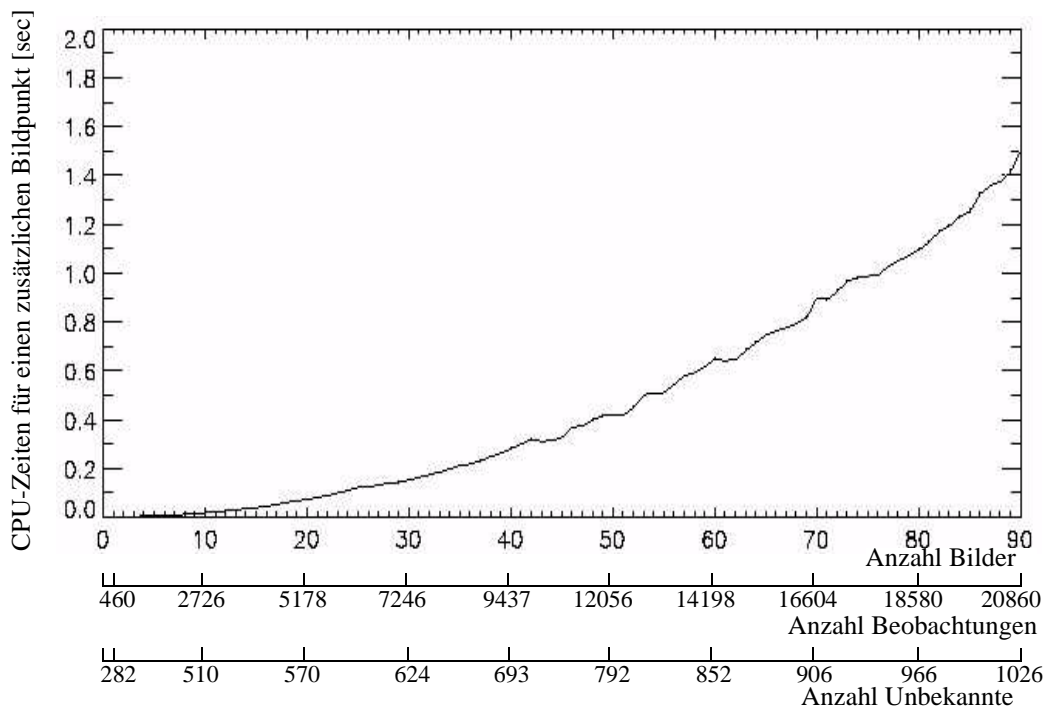


Abb. 7: CPU-Zeiten für das Einfügen eines zusätzlichen Bildpunktes in die Sequenz (Sequentielle Schätzung in OLTRIS auf einer SUN SparcStation 1+)

genden Berechnungszeiten (CPU) für das Einfügen einer neuen Beobachtung eines Bildpunktes in das Normalgleichungssystem und die sequentielle Lösung des Unbekanntenvektors mit

zunehmenden Bildern in der Sequenz. Die CPU-Zeiten lagen zwischen 0.01 sec zu Beginn und 1.54 sec am Ende der JVC-Sequenz pro zusätzlichen gemessenen Bildpunkt. Wegen der zeitaufwendigen Relinearisierung des Normalgleichungssystems, konnten diese Rechenzeiten mit der simultanen Ausgleichung natürlich nicht erreicht werden. Zum Zeitpunkt von 10 (40) triangulierten Bildern braucht die simultane Ausgleichung 3 Sekunden (20 Sek.) für jede Iteration, was ungefähr um den Faktor 70 schlechter als die sequentielle Methode ist.

6.2 Binokulares Visionsystem (Sony CCD-Kameras)

Die in Abb. 1 dargestellte Prozedur zur Bestimmung der Sensororientierung wurde mit den aufgenommenen Sequenzen des Stereo-Visionsystems durchgeführt.

(a) Extraktion charakteristischer Punkte

Nach der Aufnahme der Bilddaten wurden in den Halbbildern charakteristische Punkte (Landmarken) mit dem Förstner Interest Operator [9] extrahiert. Die Parameter für den Operator waren so gesetzt, daß Punkte mit großen Grauwertgradienten in zwei orthogonalen Richtungen und Mittelpunkte von kleinen Kreisen ausgewählt wurden. Ungefähr 60 Punkte wurden mit dieser Methode im ersten linken Bild detektiert, während die korrespondierenden Punkte im rechten Bild durch das Least Square Matching (LSM) mit geometrischen Bedingungen [2] bestimmt wurden. Die Patchgröße war 9 x 9 Pixel. Zwei Translation und zwei Maßstäbe wurden als geometrische Parameter sowie Gain und Offset als radiometrische Korrekturen gesetzt. Anschließend wurden diese Punkte unabhängig voneinander in jeder Bildsequenz verfolgt. Das Tracking der Punkte wurde nach ungefähr 50 Bildern, was 2 Sekunden Aufnahmezeit entspricht, gestoppt, da wegen der Bewegung der Kameras neue Objekte in den Bildern erschienen und deshalb neue Landmarken extrahiert werden sollten. In der neuen Bildregion des letzten linken Bildes wurden mit dem Förstner-Operator neue Punkte detektiert und die korrespondierenden Punkte wurden in der entsprechenden rechten Bildregion durch Matching bestimmt. Das letzte Stereobildpaar bildete somit das erste Bildpaar in einer neuen Teilsequenz, in der alle vorhandene Punkte weiter verfolgt wurden.

(b) Tracking der Landmarken und erste grobe Fehlersuche

In den folgenden Stereobildpaaren wurden die Landmarken in jeder Sequenz unabhängig voneinander mit LSM verfolgt. Das LSM lieferte auch eine Genauigkeitsabschätzung und erlaubte damit grobe Fehlersuche. Die Matching-Methode kann Punkte verfolgen, deren Verschiebung von Bild zu Bild innerhalb der Sequenz nicht größer als 4-5 Pixel ist. Natürlich hängt das auch von der Patchgröße ab, die in diesem Fall 9 x 9 Pixel groß war. Wegen des geringen geometrischen und radiometrischen Unterschieds innerhalb der Bildsequenz wurde das Matching nur mit zwei Translationen und ohne radiometrische Korrekturen durchgeführt. Diese Vereinfachung führte nicht zu einer Verschlechterung der Matching-Genauigkeit, sondern erhöhte signifikant die Verarbeitungsgeschwindigkeit (120 "gematchte" Punkte pro Sekunde auf einer SUN SparcStation 10). Für solche Tracking-Anwendungen ist im Allgemeinen eine kleine Patchgröße erforderlich, um Störungen durch den sich verändernden Hintergrund (bei beweglichen Objekten) und Verdeckungen (bei mobilen Sensoren) zu verringern und um die Rechenzeiten zu beschleunigen. Die durchschnittliche Verschiebung der Punkte von Bild zu Bild lag bei 2 Pixel (Maximum 7 Pixel), welche natürlich innerhalb der Teilsequenzen variierte. Probleme beim Tracking tauchten bei Punkten, die Mittelpunkt eines kleinen Kreises waren, auf,

wenn die Plattform sich frontal auf das Objekt zu bewegte. Durch die Veränderung des Maßstabes kam es zu fehlerhaften Matchings. Daher ist es für das Verfolgen von Punkten besser, wenn weniger maßstabempfindliche Punkte wie z.B. Ecken verwendet werden.

Um den Einfluß fehlerhafte Matching-Ergebnisse auf die Bestimmung der Sensorpositionen so gering wie möglich zu halten, wurden die Resultate gleich nach dem Matching getestet. Folgende Kriterien wurden für die Qualitätskontrolle des LSM verwendet, um automatisch grobe Fehler zu detektieren und auszusortieren: Standardabweichungen aus dem LSM, Korrelationskoeffizient der Grauwerte zwischen den zwei Bildern, Anzahl Iterationen, x- und y-Translationen und deren Standardabweichungen. Die Schwellenwerte als Kriterium für das Aussortieren einer Messung wurden durch robuste statistische Werte der Messungen bestimmt, welche auf den Median und dessen Standardabweichung basierten. Durchschnittlich wurden 5% der gemessenen Punkte als Fehler aussortiert.

(c) Sequentielle Schätzung

Die Sensororientierungen wurden ebenfalls wie bei der JVC-Sequenz durch die sequentielle Schätzung in OLTRIS bestimmt. Die Messungen wurden durch Addition der Kalibrierungswerte um den Einfluß systematischer Fehler korrigiert.

Da die sehr große Anzahl von Bildpunktmessungen für die gesamte Stereosequenz (66656 Punkte) ungewöhnlich lange Berechnungszeiten verursachen würde, wurde die Sequenz mit den 742 Stereobildpaaren in 10 Teilsequenzen (mit einer Anzahl von 65 bis 105 Bildpaaren) aufgeteilt. Um eine zuverlässige Verknüpfung der Teilsequenzen zu garantieren, wurde eine Überlappung von 5 Stereobildpaaren gewählt.

Die Bildsequenz startete für das erste Stereobildpaar mit bekannten Sensororientierungen. Die Näherungskordinaten der Objektpunkte (Landmarken), die durch den Förstner-Operator detektiert wurden, wurden durch räumlichen Vorwärtsschnitt bestimmt, während die folgenden Sensororientierungen als Näherungswerte durch räumlichen Rückwärtsschnitt bestimmt wurden. Alle berechneten Näherungswerte wurde als gewichtete Beobachtungen in das Normalgleichungssystem eingeführt. Der Lösungsvektor für die ersten beiden Stereobildpaare wurde durch eine simultane Ausgleichung berechnet, um am Anfang der Sequenz eine stabile Ausgangssituation zu erhalten. In diesem Stadium konnte die CPU-Zeit für die Ausgleichung vernachlässigt werden. Für alle folgende Sequenzbildpaare wurde der Lösungsvektor, der aus den laufenden Sensororientierungen jedes hinzukommenden Bildpaares und den 3-D Objektpunktkoordinaten besteht, durch sequentielle Ausgleichung bestimmt. Nach jedem sequentiellen Update der Normalgleichungsmatrix wurde alle Beobachtungen auf grobe Fehler untersucht. Für die grobe Fehlersuche, die ein wichtiger Bestandteil automatischer Meßsysteme ist, wird Baarda's Data Snooping [1] als Standardverfahren in photogrammetrischen Anwendungen eingesetzt. Das Data Snooping ist auch in OLTRIS implementiert. Das Verfahren setzt voraus, daß grobe Fehler im Beobachtungsdatensatz geringe Korrelationen haben. Entsprechend wurden nur die fünf größten Fehler aussortiert, ohne deren Korrelationen untereinander zu überprüfen, und die Normalgleichungsmatrix wurde erneut aktualisiert. Doch eine visuelle Überprüfung einiger aussortierter Beobachtungen zeigte, daß sie korrekt gemessen waren. Damit liegt die Vermutung nahe, daß zwischen den groben Fehlern hohe Korrelationen vorliegen müßten. In zukünftigen Untersuchungen müssen die Korrelationen durch Berechnung der Q_{VV} -Matrix, wie es in [11] beschrieben ist, getestet werden.

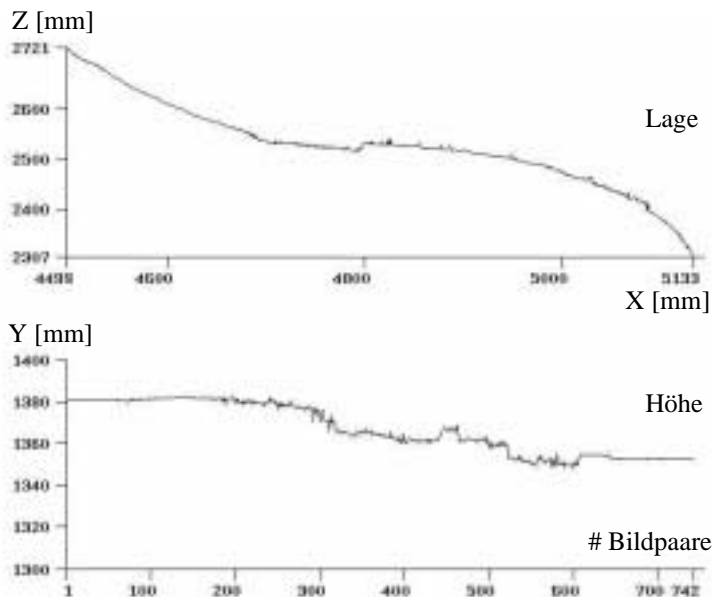


Tabelle 2: Differenz der Sensorposition des letzten Stereopaars im Vergleich zu Referenzwerten

Kamera	dx [mm]	dy [mm]	dz [mm]
links	16.6	20.7	6.7
rechts	2.8	36.3	-1.4
Mittel	9.7	28.5	2.7

Abb. 8: Berechneter Weg der mobilen Plattform in seiner Umgebung

Der berechnete Weg der mobilen Plattform in seiner Umgebung ist in Abb. 8 dargestellt. Die Abbildung zeigt, daß wegen der Fehlerfortpflanzung mit zunehmender Bildanzahl (ca. ab Bildpaar 200) die Sensororientierungen leicht zu variieren anfangen und keinen geglätteten Verlauf mehr nehmen. Da keine Kontrollpunkte innerhalb der Sequenz verwendet wurden, konnten diese Schwankungen in der Sensororientierung erwartet werden. Die Berechnungen mit OLTRIS wurden auf einer SUN SparcStation 10 durchgeführt und für jede Teilsequenz dauerten sie zwischen 4 und 23 Stunden je nach Größe und Anzahl Beobachtungen. Das zeigt, das wir mit der benutzten Hardware weit von einer Echtzeitverarbeitung sind. Aber wie wir am Anfang bereits erwähnt haben, möchten wir kein Echtzeit-Navigationssystem vorstellen, sondern zeigen, daß unserer algorithmischer Ansatz in der Lage ist, Sensororientierungen aus Bildsequenzen abzuleiten.

Um die erreichte Genauigkeit zu testen, wurde folgender Vergleich vorgenommen. Die Sensororientierung des letzten Stereobildpaares, die im letzten Schritt der sequentiellen Ausgleichung bestimmt wurde, wurde mit Referenzwerten verglichen, die aus einer simultanen Bündellösung mit den Punkten des 3-D Testfeldes abgeleitet wurden. Die Differenz beider Lösungen ist in Tabelle 2 zusammengefaßt. In der Lage ist die Abweichung besser als ein Zentimeter und in der Höhe liegt die Differenz bei 2.8 cm.

6. Mögliche Anwendungen

Für Anwendungen und Aufgaben in der Robotik muß man die Robot Vision im generellen Zusammenhang mit den Bereichen Computer Vision und Machine Vision eng verknüpft sehen (Abb. 9). Während die Computer Vision Theorie und Grundlagen für Bild- und Szenenanalyse liefert, bietet die Machine Vision, in deren Bereich die Robot Vision mit ihren zeitbedingten Lösungen liegt, Aspekte der Sensormodellierung und Systeme. Typische Anwendungen und Aufgaben für Visionsysteme liegen in der Inspektion und Qualitätssicherung von Industrieprodukten, in der Produktion beim Schneiden, Schweißen, Zusammenbauen, Oberflächenbehandlung, etc., bei Transport, Navigation, Überwachung und Objekterkennung. Zuech und Miller

[24] beschreiben kurz 23 Robot Vision Systeme. Doch viele der Systeme erfüllen die photogrammetrischen Anforderungen [13] wie z.B. die Fähigkeit der Selbstdiagnose (Qualitätskontrolle), das Potential für hohe Genauigkeit und Zuverlässigkeit durch redundante Sensordaten und die Flexibilität bei der Aufgabenstellung durch Funktionen für die 3-D Objektrekonstruktion nicht.

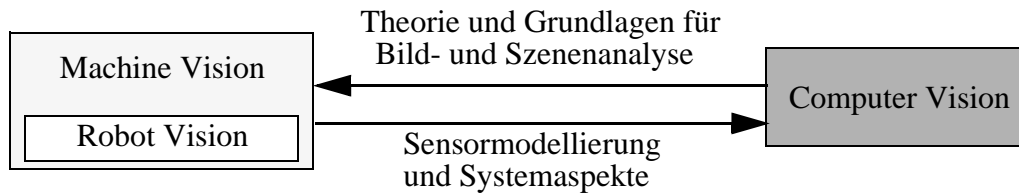


Abb. 9: Robot Vision im Zusammenhang mit Machine und Computer Visi-

Für den präsentierten Algorithmus stellen wir uns mögliche Anwendungen in der industriellen Inspektion und Überwachung von Objekten, z.B. in der Nuklearindustrie, durch mobile Meßroboter, die mit einem optischen Visionsystem ausgerüstet sind, vor, bei denen die Geschwindigkeit keine vorderrangige Rolle spielt. Dabei kann ein Expertensystem [19] die notwendigen Informationen für eine genaue Positionierung der Sensoren zur Vermessung oder Überwachung der Objekte bereitstellen, die dann von dem Meßroboter angesteuert werden, so daß dort die erforderlichen Aufnahmen gemacht werden können.

7. Zusammenfassung

In zwei praktischen Tests mit einem Mono- und Stereo-Visionsystem haben wir gezeigt, daß unser algorithmischer Ansatz in der Lage ist, Sensororientierungen aus Bildsequenzen abzuleiten. In dem praktischen Beispiel mit dem Stereo-Visionsystem wurde durch die Vorgabe der Startposition auf Kontrollpunkte verzichtet. Doch der Verzicht auf Kontrollpunkte in der aufgenommenen Sequenz kann durch Fehlerfortpflanzung zunehmende Ungenauigkeit in der Bestimmung der Sensororientierungen verursachen. Zu wenig Punkte und/oder schlechte Punktverteilung im Bild bzw. in der Sequenz erzeugen auch Ungenauigkeiten der Sensororientierungen. Deshalb sollte das System in unterschiedlichen Umgebungen getestet werden, in dem auch Kontrollpunkte und eine größere Anzahl von extrahierten Punkten verwendet werden. Allerdings bietet das System mit den verwendeten Softwarelösungen keine Echtzeitverarbeitung. Die einzelnen Softwaremodule sind hinsichtlich der Rechenzeiten noch nicht optimiert. Durch Integration der Softwaremodule in einem System könnte der Verarbeitungsprozeß der Sequenzen auch stärker automatisiert werden. Außerdem muß getestet werden, ob eine ausreichende Bestimmung der Sensororientierungen durch Reduktion der Bildrate (keine Videoechtzeit) möglich ist.

Literaturverzeichnis

- [1] Baarda, W., 1968. A Testing Procedure for Use in Geodetic Networks. Netherlands Geod. Com. Publications on Geodesy. Vol.2, No.5, Delft.
- [2] Baltsavias, E.P., 1991: Multiphoto Geometrically Constrained Matching. Dissertation, ETH Zürich, Institut für Geodäsie und Photogr., Mitteilungen Nr. 49, Dezember, 221 S.
- [3] Beyer, H.A., 1991: An Introduction to Photogrammetric Camera Calibration. Seminar Orasis, St. Malo, Frankreich, September, pp. 37-42.

- [4] Beyer, H.A., 1992: Geometric and Radiometric Analysis of a CCD-Camera Based Photogrammetric Close-Range System. Dissertation, ETH Zürich, Institut für Geodäsie und Photogrammetrie, Mitteilungen Nr. 51, Mai, 186 S.
- [5] Blais, J.A.R., 1983: Linear Least-Squares Computations Using Givens Transformations. *The Canadian Surveyor*. Vol. 37, No. 4, pp. 225-233.
- [6] Deriche, R., Faugeras, O., 1990: Tracking Line Segments. Proceedings of the First European Conference of Computer Vision, Antibes, France, April, pp. 259-268.
- [7] Edmundson, K.L., Novak, K., 1992: On-Line Triangulation for Autonomous Vehicle Navigation. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXIX, Part B5, Commission V, pp. 916-922.
- [8] El-Hakim, S.F., 1985: A Photogrammetric Vision System for Robots. *Photogrammetric Engineering & Remote Sensing*, Vol. 51, No. 5, pp. 545-552.
- [9] Förstner, W., Gülch, E., 1987: A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. Proc. of ISPRS Intercommission Conf. on "Fast Processing of Photogrammetric Data", Interlaken, June 2-4, pp. 281-305.
- [10] Grün, A., 1982. An Optimum Algorithm for On-Line Triangulation. Symposium of ISPRS Commission III, Helsinki, June 7-11.
- [11] Grün, A., 1985: Algorithmic Aspects in On-Line Triangulation. *Photogrammetric Engineering and Remote Sensing*, Vol. 51, No. 4, pp. 419-436.
- [12] Grün, A., 1986: Photogrammetrische Punktbestimmung mit der Bündelmethode. ETH Zürich, Institut für Geodäsie und Photogrammetrie, Mitteilungen Nr. 40, März, 87 S.
- [13] Grün, A., 1991: Recent Advances of Photogrammetry in Robot Vision. Proceedings of First Australian Photogrammetric Conference, Sydney, November, Paper No. 1.
- [14] Grün, A., Kersten, Th., 1992: Sequential Estimation in Robot Vision. *Inter. Archives of Photogrammetry and Remote Sensing*, Vol. XXIX, Part B5, Commission V, pp. 923-931.
- [15] Hahn, M., 1991: 3-D Egomotion from Long Image Sequences. Proceedings of the Second Int. Workshop on High Precision Navigation, Stuttgart/Freudenstadt, Nov., pp. 241-251.
- [16] Holm, K.R., 1989: Test of Algorithms for Sequential Adjustment in On-Line Phototriangulation. *Photogrammetria*, 43, pp. 143-156.
- [17] Kersten, Th., Grün, A., Holm, K.R., 1992: On-Line Point Positioning with Single Frame Camera Data. DoD Final Technical Report, Nr. 7, IGP-Bericht, Nr. 197, ETH-Zürich.
- [18] Maas, H.-G., 1992: Digitale Photogrammetrie in der dreidimensionalen Strömungsmesstechnik. Dissertation, ETH Zürich, Institut für Geodäsie und Photogrammetrie, Mitteilungen Nr. 50, Februar, 140 S.
- [19] Mason, S., Grün, A., 1993: Design of Spatial Image Sequences for Vision-based Inspection. NRP23-Symp. on Artificial Intelligence and Robotics, ETH Zürich, Nov., pp. 41-46.
- [20] Matthies, L., Szeliski, R., Kanade, T., 1989: Kalman Filter-based Algorithms for Estimating Depth from Image Sequences. *Int. Journal of Comp. Vision*, Vol. 3, No. 3, pp. 209-238.
- [21] Runge, A., 1987: The Use of Givens Transformation in On-Line Triangulation. Proceedings of ISPRS Intercommission Conference on "Fast Processing of Photogrammetric Data", Interlaken, Switzerland, June 2-4, pp. 179-192.
- [22] Wyatt, A.H., 1982: On-Line Triangulation - An Algorithmic Approach. Master Thesis, Department of Geodetic Sciences and Surveying, The Ohio State Univ., Columbus, Ohio.
- [23] Zhang, Z., Faugeras, O., 1990: Motion Tracking in a Sequence of Stereo Frames. Proceedings 9th European Conference of Artificial Intelligence, Stockholm, August, pp. 747- 752.
- [24] Zuech, N., Miller, K.R., 1987: Machine Vision. Fairmont Press, Lilburn, GA 30247, USA, p. 209.