Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Axel Pinz
Institut für Elektrische Messtechnik und Messsignalverarbeitung
E-Mail: axel.pinz@tugraz.at
Tel: 0316 873 5021

„Smart Tracking“: Selbstlokalisation mit intelligenten Sensoren

Fusion of Stereo Vision and Inertial Sensors

Einleitung

Die Position und Orientierung eines bewegten Objektes im Raum soll in Echtzeit erfasst werden. Dieses Problem des „real-time tracking“ hat viele potenzielle Anwendungen in der Automatisierungstechnik, der Robotik, bei autonomen Fahrzeugen, und in HCI (Mensch-Maschine Kommunikation, beispielsweise in der Virtuellen Realität).

Das Thema „Tracking“ wird in einer von Axel Pinz geleiteten Gruppe am Institut für Elektrische Messtechnik und Mess-Signalverarbeitung (EMT) intensiv im Rahmen mehrerer Forschungsprojekte betrieben (siehe dazu auch http://www.emt.tugraz.at/~tracking). Im Vordergrund steht immer die bildgestützte Messtechnik, also das Erfassen der Trajektorie einer bewegten Kamera aus den Bildern dieser Kamera. Dabei gibt es problematische Konfigurationen – etwa eine rasche Rotation der Kamera – wo Bewegungsunschärfe und sehr rasche Änderungen des Gesichtsfeldes die benötigte Korrespondenzfindung behindern. Dies hat zur Entwicklung von hybriden Systemen geführt, in denen die bildgebenden Sensoren von komplementären Sensoren unterstützt werden.

Im vorliegenden Bericht wird vor allem das FWF Projekt P15748 „Tracking with Smart Sensors“ vorgestellt. Dieses Projekt wurde im Sommer 2002 begonnen. Es steht unter der Leitung von Axel Pinz und wird zu gleichen Teilen am EMT der TU Graz (Mitarbeiter: DI Ulrich Mühlmann) und am Institut für Automation und Regelungstechnik (ACIN, Partner: Dr. Markus Vincze, DI Stefan Chroust) der TU Wien durchgeführt. Allerdings wäre dieses Projekt ohne grundlegende Vorarbeiten in weiteren Forschungsprojekten nicht möglich gewesen. [1]

Problemstellung – „Tracking“ und „Structure and Motion“

Tracking einer bewegten Kamera kann durch die fortgesetzte Lösung des sogenannten „perspective-n-point“ Problems erreicht werden: In jedem Bild der Echtzeit-Video-Sequenz werden n Punkte gesucht, die mit n bekannten „landmarks“ (Referenzpunkten) in der Szene korrespondieren. Dieses Problem ist in der Photogrammetrie und in der Geodäsie als räumlicher Rückwärtsschnitt wohlbekannt. Um Tracking in Echtzeit zu erreichen, ist es notwendig, n möglichst klein zu halten, was allerdings die Genauigkeit und die Robustheit der Lösung gegen grobe Fehler sehr negativ beeinflussen kann. Dafür kann das Verfahren andererseits vereinfacht und beschleunigt werden, wenn man die Zeitachse berücksichtigt, also aus der bereits bekannten Trajektorie die aktuelle Position und den erwarteten Inhalt des nächsten Bildes schätzt.

Bisher sind wir davon ausgegangen, dass die Struktur der Szene, also die von der Kamera beobachteten Referenzpunkte, bekannt sein muss, um die Trajektorie messen zu können. Umgekehrt ist es auch möglich, die 3D Geometrie der Szene aus einer Bildfolge zu rekonstruieren, wenn die Trajektorie bekannt ist. Diese Aufgabe heißt in der Computer Vision „Structure from Motion“. Dazu müssen Punktkorrespondenzen zwischen mehreren Bildern der Bildfolge hergestellt werden („Matching“). Wenn die Trajektorie genügend genau bekannt ist, dann ist diese Problemstellung sehr ähnlich zur Stereo-Rekonstruktion (Matching von zwei Bildern von zwei Kameras mit einer bekannten, festen Basis). Allerdings ist die Echtzeitanforderung auch in diesem Fall ein hoher Anspruch an die Leistungsfähigkeit des Systems.

Das Projekt „Smart Tracking“ hat sich eine Kombination der beiden oben genannten Problemstellungen („Tracking“ und „Structure from Motion“) zum Ziel gesetzt: „Structure and Motion“ in Echtzeit. Es soll also sowohl die Trajektorie bestimmt werden, als auch die 3D Struktur der Szene. Das Konzept der oben beschriebenen bewegten Kamera muss dafür allerdings erweitert werden. Wir wollen ein neues, intelligentes Tracking-System aufbauen, welches aus zwei Kameras (Stereo-Vision) und aus Inertialsensoren besteht.

Neue Hardware-Komponenten

Seit einigen Jahren gibt es neuartige Inertialsensoren, welche vorwiegend im Automobil eingesetzt werden (Beschleunigungssensoren für das Auslösen von Airbags, Drehratensensoren zur Erfassung etwa des Neigungswinkels). Aus derartigen „silicon micromachines“ haben wir einen neuen Inertialsensor (Abb.1) bestehend aus drei Beschleunigungssensoren und drei Gyroskopen aufgebaut, welcher alle sechs Freiheitsgrade der Position und Orientierung – drei Richtungen der Translation, und drei Achsen der Rotation – misst. Allerdings ist ein solcher Sensor im Vergleich zu herkömmlichen, wesentlich teureren und größeren INS (Inertial-Navigations-Systemen) viel zu ungenau, um alleine zur verlässlichen Messung einer Trajektorie eingesetzt werden zu können. Es muss ja die Beschleunigung zwei mal und die Drehrate einmal integriert werden, um Position bzw. Orientierung zu erhalten. Vor allem zu hohes Rauschen und hohe Drift führen dazu, dass die Messwerte schon nach etwa einer Sekunde unbrauchbar werden. Ein derartiger Sensor liefert zwar hohe Datenraten von rund 1kHz, muss aber für einen sinnvollen Einsatz mit einem anderen Verfahren – in unserem Fall der Bildanalyse – kombiniert werden.

Abb.1: Ein neuer Inertialsensor (Eigenentwicklung) bestehend aus drei Beschleunigungs- und drei Drehratensensoren kann zur Messung aller sechs Freiheitsgrade der Position und Orientierung eingesetzt werden.

Grundsätzlich gibt es mehrere Konfigurationen von Kamerasystemen, welche für Tracking eingesetzt werden können. CCD-Kameras mit Framegrabber oder mit FireWire Interface erreichen Frameraten von 25-60Hz. Derartige Systeme haben den Vorteil, dass jeweils das gesamte Bild im Speicher des Rechners abgelegt wird, also auch im gesamten Bild nach Punktkorrespondenzen gesucht werden kann. Andererseits ist unter Echtzeit-Bedingungen ohnedies nur eine Suche in relativ kleinen „Fenstern“, beispielsweise von 11 x 11 Pixeln je Punkt möglich.

CMOS Kamera Technologie basiert auf dem Prinzip eines „Active Pixel Sensors“, in dem auf jedes Bildelement wahlfrei direkt zugegriffen werden kann. Während die meisten CMOS Kameras für das Auslesen eines gesamten Bildes niedrigere Frameraten erreichen als CCD Kameras, bietet sich hier die Möglichkeit, nur die für das Tracking wirklich benötigten Regionen (also die oben besprochenen kleinen Fenster) auszulesen. Auf diese Weise erreicht man wesentlich höhere Datenraten von mehreren 100 Hz bis zu einigen kHz. Diese hohen Wiederholraten erlauben noch kleinere Fenster, da sich

[1] Grundlagenforschung zu hybridem Tracking im FWF Projekt „Studierstube“ (gemeinsam mit Prof. Gervautz von der TU Wien), und zu mobilem Tracking im FWF Projekt „MCAR – Mobile Collaborative Augmented Reality“ (gemeinsam mit Prof. Schmalstieg, TU Wien). Anwendungen dieser Tracking Technologien im VRVis Kplus Kompetenzzentrum, im CD-Labor für Kraftfahrzeug-Messtechnik (Leitung: Prof. Brasseur, TU Graz), und im EU-IST Projekt „VAMPIRE – Visual Active Memory Processes and Interactive Retrieval“.

ein Punkt im Bild in kürzerer Zeit weniger weit fortbewegt und seine Position im nächsten Fenster besser geschätzt werden kann.

Unser hybrides Tracking-System basiert auf zwei CMOS-Kameras. Da eine solche Kombination, bei der aus beiden Kameras korrespondierende Fenster zeitsynchron ausgelesen werden können, nicht als fertiges Produkt erhältlich ist, wird im ersten Projektjahr des Projektes „Smart Tracking" eine derartige neue Ansteuerung für ein CMOS-Kamera Stereo Setup entwickelt (Abb.2).

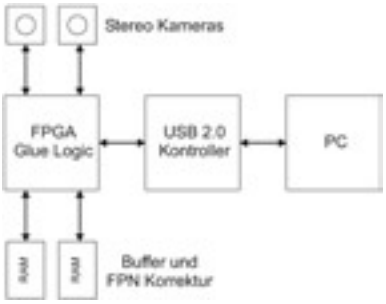

Abb.2: Blockdiagramm für eine neue Ansteuerungslogik für zwei CMOS-Kameras. Zeitsynchrones Stereo von kleinen Fenstern ist möglich. Beide Kameras können kalibriert werden, und die Daten werden über eine USB-2 Schnittstelle übertragen.

Ein derartiger CMOS-Stereo-Sensor erfüllt zwar die meisten Anforderungen an ein bildgestütztes Tracking. Dennoch können Situationen auftreten, wo Punkte zeitweilig nicht detektiert werden können (Bewegungsunschärfe, oder temporäre Verdeckung eines Zieles durch ein anderes Objekt im Vordergrund), oder überhaupt verloren gehen (bei sehr großen, sich plötzlich ändernden Beschleunigungen). Deshalb muss das bildgestützte Tracking mit einem anderen Verfahren – in unserem Fall Inertialtracking – kombiniert werden.

Insgesamt soll im Rahmen des Projektes ein neuartiger Sensorkopf entwickelt werden, welcher aus einem Stereo-Vision-System und aus dem oben beschriebenen Inertialtracker besteht. Je nach Anwendung muss das Gesamtsystem leicht sein (z.B. am Kopf des Benutzers montiert), und die Stereo-Basis kann variieren (z.B. 20cm für Augmented Reality am Kopf, mehr als 1m im automobilen Einsatz). Alle Komponenten werden in einer fixen Konstellation montiert und kalibriert sein, d.h. die innere und relative Orientierung der Kameras, Linsenverzerrung, sowie weitere Parameter („fixed pattern noise" der CMOS-Kameras, diverse geometrische und elektrische Parameter der sechs Inertialsensoren, etc.) werden als bekannt vorausgesetzt.

Neue Algorithmen

Um mit dem oben beschriebenen Sensorkopf letztlich das Projektziel – „Structure and Motion in Real-Time" zu erreichen, werden zu verschiedenen Teilproblemen neue Algorithmen benötigt und auf der obersten Systemebene zu einem Gesamtsystem für Hybrides Tracking zusammengefügt. Die geplante Funktionsweise einiger dieser Algorithmen wird nachfolgend skizziert[2] . Von einigen Komponenten existieren bereits erste Prototypen. Vieles befindet sich noch in Entwicklung, da das Projekt ja erst ein halbes Jahr „jung" ist.

Gesamtsystem

In einer Initialisierungsphase wird mit Hilfe des Stereo Systems die Lage einiger markanter Punkte (erste „Landmarks") relativ zum Sensorkopf geschätzt. Sobald das System bewegt wird, erfolgt eine erste Schätzung der Trajektorie mit Hilfe der Inertialsensoren. Weitere bildgestützte Messungen aus neuen Blickwinkeln verbessern die Genauigkeit der „Landmarks", sodass diese bald für bildgestütztes Tracking benutzt werden können. Bewegt sich der Messkopf weiter, so verschwinden die ersten „Landmarks" aus dem Gesichtsfeld. Daher müssen laufend weitere neue „Landmarks" bestimmt werden.

Erkennen von "Landmarks"

In natürlicher Umgebung kommen verschiedene visuelle Merkmale vor, welche als „Landmark" oder „Target" für bildgestütztes Tracking verwendet werden können. Unser System verwendet Ecken (sog. „corner-detection") und kompakte Flecken (z.B. annähernd kreisförmige Objekte, sog. „blobs"). In beiden Fällen wird die Position der Merkmale subpixelgenau bestimmt, um die erforderliche Messgenauigkeit zu erreichen. Abbildung 3 zeigt ein Beispiel eines kleinen Bildausschnittes mit einer Ecke. Diese wird zunächst mit einem herkömmlichen Eckendetektor geschätzt (rotes Kreuz), und nachfolgend als Konsensus der Geraden, welche diese Ecke begrenzen, subpixel-genau gefunden.

Abb.3: Eine Ecke („corner") in einem 7x7 Pixel kleinen Bildausschnitt wird subpixel-genau detektiert.

Fusion von bildgestützter und Inertial-Messtechnik

Bildgestütztes Tracking und Inertialtracking sind zwei komplementäre Verfahren, welche einander sehr gut ergänzen. Bei langsamen Bewegungen und bei gut bekannter, relativ statischer Szene (d.h. wenig Verdeckungen von „Landmarks" durch Objekte im Vordergrund, viel statischer Hintergrund und wenig dynamischer Vordergrund) ist die Bildanalyse sehr gut geeignet. Kommt es zu raschen Bewegungen, insbesondere Rotationen des Sensorkopfes, oder zu temporären Verdeckungen, so soll der Inertialsensor für eine gewisse, kurze Zeit (< 1 Sekunde!) die tragende Rolle übernehmen. Nur durch ein reibungsloses und sehr gut synchronisiertes Zusammenspiel der verschiedenen Komponenten kann ein stabiles Verhalten des Gesamtsystems erreicht werden. Unser erster Prototyp mit

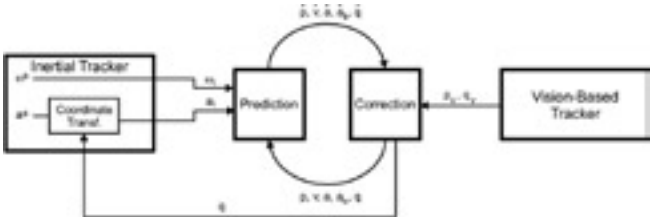

Abb.4: Modifizierter „Extended Kalman Filter" zur Fusion von Inertialtracking und bildbasiertem Tracking.

Inertialsensor und nur einer Kamera verwendet ein modifiziertes „Extended Kalman Filter" (Abb.4).

Anwendungen

Mobile „Augmented Reality"

Da unser Sensorkopf (anders als etwa bei herkömmlichem magnetischen oder optischem „outside-in" Tracking, oder bei GPS-Systemen) keine aktive Information von außen benötigt, kann das System an beliebigen Orten eingesetzt werden. Mit einer geeigneten Stromversorgung ist man also völlig mobil und kann durch den „Structure and Motion" Ansatz gleichzeitig unbekanntes Terrain betreten und tracken. Dies ist in Innenräumen und außen möglich. Eine sehr attraktive Anwendung ist die mobile „Augmented Reality" (AR). Einer oder mehrere Benutzer tragen AR-Systeme (siehe Abb.5) bestehend aus einem Subsystem für das Tracking und einem Subsystem

[2] Weitere wichtige Komponenten, welche hier nicht genauer dargestellt werden können sind: Vorhersage-Modul (erwartete Position von „landmarks" im 2D Bild und in 3D Szenenkoordinaten), Komplexitätsreduktion für Szenen mit vielen „landmarks", Auswahl von „guten landmarks", also Merkmalen welche sich für die Tracking-Aufgabe besonders eignen, sowie die eigentliche Schätzung der „Pose" einer Kamera.

Abb.5: Mobiles AR-System bestehend aus Helm mit Datenbrille und Tracking Sensoren, und Rucksack mit zwei Computersystemen (single-board computer für Tracking, Laptop für 3D Graphik). (a) Gesamtsystem in Verwendung. (b) Rucksack. (c) Helm.

für 3D Graphik Visualisierung. „Augmented Reality" bedeutet, dass der Benutzer gleichzeitig die reale 3D Szene und die überlagerte virtuelle 3D Graphik wahrnimmt. Diese Wahrnehmung kann nur befriedigend funktionieren, wenn die räumliche Übereinstimmung von Realität und Virtualität (also das Tracking) perfekt funktioniert.

Autonome Robot-Navigation, Selbstlokalisation im RoboCup

Während AR die Hauptanwendung bei den meisten Projekten unserer Gruppe darstellt, betreiben unsere Partner im Projekt „Smart Tracking" an der TU Wien mobile Roboternavigation. In diesem Bereich kann es sein, dass ein Modell der 3D Szene bereits existiert, „Structure and Motion" wird aber in vielen Anwendungen dringend benötigt, etwa wenn sich eine an sich bekannte Szene verändert hat, oder beim Navigieren in unbekanntem Terrain.

An dieser Stelle möchte ich noch ein weiteres Projekt erwähnen: Die Teilnahme der TU Graz an der Roboter-Fussball-Weltmeisterschaft RoboCup. Dies ist eine Aktivität, welche von derzeit 12 Instituten aus 3 verschiedenen Fakultäten der TU Graz sowie von diversen Sponsoren aus der Industrie unterstützt wird (siehe: http://www.robocup.tugraz.at) . In der Middle Size League spielen Teams von jeweils vier autonomen Robotern auf einem Spielfeld von rund 5x9m gegeneinander. Die Roboter messen rund 60cm im Durchmesser und sind 80 cm hoch. Es wird ein normaler Winterfussball (rot) verwendet. Die Robot-Selbstlokalisation auf dem Spielfeld erfolgt mit Hilfe mehrerer Sensorsysteme (Ultraschall, Laser, und Bildverarbeitung). Unsere Gruppe hat die Entwicklung der bildgestützten Selbstlokalisation in das RoboCup Projekt eingebracht. Abbildung 6 zeigt einen ersten Prototypen des TU Graz Fussballroboters „Keksi" und Details zum Bildverarbeitungs-Sensor, welcher ein 360 Grad Panorama des Spielfeldes aufnimmt und analysiert.

Zusammenfassung

Dieser Bericht gibt einen Überblick über die Ziele und den aktuellen Stand der Arbeiten im FWF-Projekt „Tracking with Smart Sensors". Das angestrebte System funktioniert rein „inside-out", das heißt es wird die eigene Position und Orientierung im Raum über visuelle Stereo-Information und zusätzlich mithilfe von Inertialsensoren bestimmt. Von außen ist keine zusätzliche Information wie etwa GPS oder signalisierte „Landmarks" nötig (die Szene muss nur visuell wahrnehmbar sein, es darf also nicht völlig dunkel sein). Gleichzeitig nimmt das System die umgebende Szene wahr und erfasst und verfeinert laufend 3D Strukturinformation über die Szene. Diese Vorgangsweise modelliert sowohl sensorisch (Stereo-Sehen + Gleichgewichtsorgan) als auch algorithmisch (Fusion verschiedener Sinneswahrnehmungen) die menschliche Raumwahrnehmung.

Danksagung

Prof. Brasseur hat in den vergangenen Jahren den Aufbau einer Gruppe für Bildgestützte Messtechnik an seinem Institut ermöglicht und stets voll unterstützt. Neben dem FWF Projekt P15748 „Tracking with Smart Sensors", Mitarbeiter DI Ulrich Mühlmann und Partner DI Stefan Chroust und Dr. Markus Vincze an der TU Wien haben noch folgende Mitarbeiter zu den hier gezeigten Arbeiten beigetragen: DI Brandner, Dr. Ganster, Dr. Ribo, DI Siegl, DI Stock. Der Inertialsensor wurde im Rahmen einer von Prof. Brasseur betreuten Diplomarbeit (DI Lang) entwickelt. Die Arbeiten am Thema „Tracking" werden auch aus folgenden weiteren Projekten finanziert: FWF-MCAR P14470-INF, EU-VAMPIRE IST-2001-34401, CD-Labor für Kraftfahrzeug-Messtechnik. Die Selbstlokalisation im RoboCup entstand im Rahmen der Diplomarbeit DI Wolf. Das Gesamtkonzept im RoboCup wurde vom „Scientist in Charge" DI Gerald Steinbauer, Univ.Ass. bei Prof. Wotawa entwickelt.

Smart Tracking: Fusion of Stereo Vision and Inertial Sensors (FWF Project P15748):

In this project, a new, generic approach to real-time tracking using a combination of „smart sensors" is searched. We plan to use a sensor suite consisting of a fixed, calibrated stereo rig together with an „inertio-tracker" based on accelerometers and gyroscopes. These two sensor types provide complimentary characteristics: visual sensing is very accurate at low velocities while inertial sensors can track fast motions but suffer from drift particularly at low velocities. This inside-out tracking system will be subjected to arbitrary motion and shall operate in cluttered multi-object scenes with multiple and independent motion and some static, but 3D background. A typical example might be a person carrying the sensors and walking through a city. The primary goal is a reliable reconstruction of the trajectory of the system itself, as well as the recovery of 3D structure required for successful tracking. Similar to the perceptive capabilities of a human, the system shall operate autonomously, without requiring additional information about its localisation and pose.

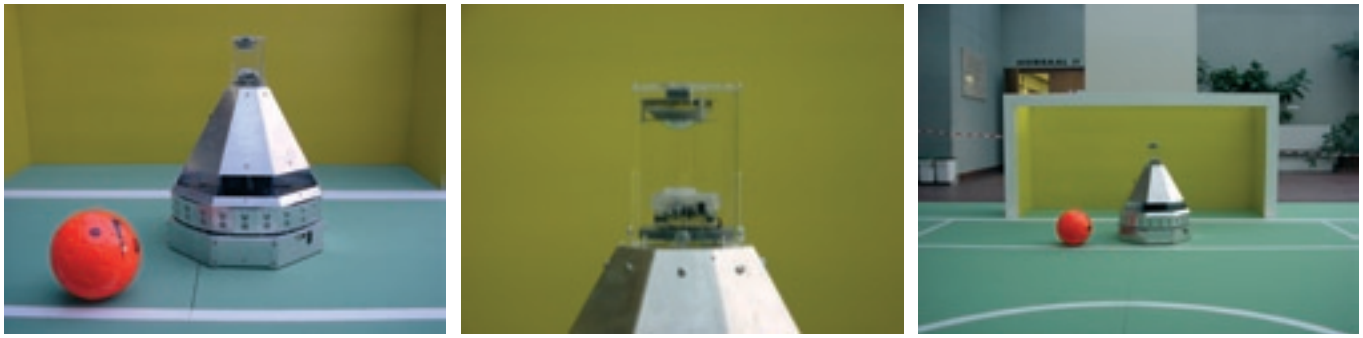

Abb. 6: (a) Autonomer Fussballroboter „Keksi" des Teams der TU Graz für die Roboter-Fussball-Weltmeisterschaft RoboCup. (b) Detail des Bildverarbeitungs-Sensors (360 Grad Panorama). (c) RoboCup Spielfeld im Foyer der Inffeldgasse 18.