Auf dem Weg zum persönlichen Assistenten – Multisensorielle Personendetektion zur Realisierung von robustem Folgeverhalten für einen mobilen Serviceroboter

Stefan Feyrer, Andreas Zell
Universität Tübingen
Wilhelm-Schickard-Institut für Informatik
Arbeitsbereich Rechnerarchitektur
Köstlinstraße 6, D-72074 Tübingen
{feyrer, zell}@informatik.uni-tuebingen.de

Zusammenfassung. Natürliche Mensch-Maschine-Schnittstellen für mobile Serviceroboter setzen seitens des Roboters die Fähigkeit voraus, Personen und deren Verhalten wahrzunehmen. Wir präsentieren in diesem Beitrag ein System, das einen mobilen Roboter in die Lage versetzt, Personen zu detektieren, aktiv zu tracken und unter Vermeidung von Hindernissen schritthaltend zu verfolgen. Die Personendetektion erfolgt multisensoriell. Zum einen kommen mehrere bildverarbeitungsbasierte Verfahren zum Einsatz, zum anderen zusätzlich ein Ansatz zur Detektion menschlicher Beine mit Hilfe von Laser-Entfernungsdaten. Die Kombination beider Modalitäten ermöglicht ein flexibles und robustes Folgeverhalten.

Abstract. Natural man-machine interfaces for mobile service robots presuppose on the part of the robot the ability to perceive persons and their behavior. In this paper we present a system, which enables a mobile robot to detect persons, to track them acitvely, and to pursue them in real-time. Obstacles are detected and avoided during pursuit. The process of person detection is based on multisensory information. Besides several vision-based methods an additional approach to detect human legs in laser scans is used. Combining both modalities enables a flexible and robust following behavior.

1 Einleitung

Mobilen Servicerobotern im Einsatz als persönliche Assistenten wird im allgemeinen ein beträchtliches mittel- bis langfristiges Marktpotential beigemessen. Die möglichen Einsatzfelder sind vielfältig. Produktionsassistenten im industriellen Bereich beispielsweise könnten durch Erledigung von Aufgaben in Interaktion mit dem Menschen Verbesserungen sowohl im Hinblick auf eine Steigerung der Produktivität als auch hinsichtlich der Humanisierung von Arbeitsplätzen erzielen. Shopping-Assistenten wären ein wesentlicher Beitrag, um Kunden das Zurechtfinden in Supermärkten und Kaufhäusern zu erleichtern. Im häuslichen Bereich ist ein Einsatz als mobiler Heim- und Pflegeassistent, der einfache Tätigkeiten ausführt und gleichzeitig als Mobilitätshilfe und Kommunikationsplattform dient, denkbar.

Eine große Anzahl von Forschungsprojekten beschäftigen sich derzeit mit solchen oder ähnlichen Assistenzsystemen [1] [2] [8]. Unabhängig vom konkreten Szenario ist stets eine robuste Erfassung und Interpretation der Umgebung, ein hohes Maß an Adaptivität und nicht zuletzt eine natürliche Mensch-Maschine-Schnittstelle notwendig. Eine wichtige Voraussetzung in diesem Zusammenhang ist die Fähigkeit des Roboters, den Bediener als Interaktionspartner wahrzunehmen, ihn zu lokalisieren und seine Aufmerksamkeit auf ihn zu richten, um gegebenenfalls auf verbale oder gestische Anweisungen angemessen reagieren zu können. Idealerweise sollte eine solche Benutzererkennung nicht entkoppelt von der momentan aktuellen Aufgabenstellung, sondern vielmehr in diese integriert und an den jeweiligen Aufgabenkontext angepaßt erfolgen.

Das hier beschriebene System setzt an genau dieser Stelle an. Es ermöglicht es, mit Hilfe eines mobilen Roboters Personen zu detektieren und zu lokalisieren. Ferner kann der Roboter auf Wunsch einer Person schritthaltend folgen, wobei dynamische und stationäre Hindernisse erkannt und umfahren werden.

Im Unterschied zu Vorgängerversionen des Systems [4] [5] wird hier eine neue Qualität der Robustheit und Flexibilität erreicht, indem die bisher rein visuelle Detektion von Personen durch eine laserbasierte Methode ergänzt wird. Eine intelligente Integration dieser zusätzlichen Modalität in das bisherige System ermöglicht es zum einen, die Vorteile einer Datenfusion auf Modellebene zu nutzen, um auch unter schwierigen Umgebungsbedingungen ein stabiles Detektionsverhalten zu realisieren. Zum anderen entfällt durch eine flexible automatische Gewichtung der Detektionsmodalitäten die Beschränkung, daß Personen nur dann erkannt werden, wenn ihr Gesicht dem Roboter zugewandt ist. Das neue System kann Personen in beliebiger Orientierung detektieren und ist zudem in der Lage, deren Orientierung bezüglich des Roboters zu bestimmen. Für eine vorausgehende Person besteht somit auch während des Folgevorgangs die Möglichkeit, die Orientierung bezüglich des Roboters beliebig zu wechseln.

2 Visuelle Detektion von Personen

Die visuelle Komponente der Personendetektion arbeitet mit dem Ziel, Kopf- und insbesondere Gesichtsregionen aus den Kamerabildern zu extrahieren. Sie basiert ihrerseits auf mehreren Einzelmerkmalen. Neben einem farb- und einem konturbasierten Verfahren wird eine Bewegungsdetektion durchgeführt, um bewegte Objekte in der Szene vom Hintergrund zu trennen. Ferner kommt ein Stereoverfahren zum Einsatz, um den Abstand zwischen Roboter und Person zu bestimmen. Die verschiedenen Module und ihr Zusammenspiel sind in [4] und [5] beschrieben. Aus diesem Grund erfolgt hier lediglich eine grobe Skizzierung.

Der farbbasierte Teil des Systems realisiert eine Hautfarbdetektion im HSI-Farbraum, wobei die eigentliche Farbklassifikation ausschließlich im helligkeitsinvarianten zweidimensionalen HS-Unterraum stattfindet. Die zu detektierenden Farbbereiche werden über Lookup-Matrizen definiert, die durch manuelle Segmentierung von umfangreichem Bildmaterial gewonnen wurden. Die Aufnahme des Bildmaterials erfolgte mit 13 Personen bei unterschiedlichen Beleuchtungsverhältnissen, so daß die entstehende Klassifikationsmatrix bereits inhärente Invarianzeigenschaften bezüglich individueller Hautfarbausprägungen und Umgebungsfarbtemperaturen enthält.

Um bewegte Vordergrundobjekte vom stationären Hintergrund zu trennen, wird ein Differenzbildverfahren angewandt, welches bei stationärem Roboter zwei aufeinanderfolgende Bilder direkt, bei bewegtem Roboter das aktuelle Bild mit einem Prädiktionsbild, das unter Berücksichtigung der Eigenbewegung des Roboters berechnet wird, verglichen. Die schnellen Schwenk- und Neigebewegungen der Kameras werden nicht kompensiert, was eine sakkadische Kameranachführung zur Folge hat.

Abbildung 1: Ergebnisse der Farb-, Bewegungs- und Konturdetektion sowie der Tiefenanalyse.

Zur Detektion von Konturen kommt ein Verfahren zum Einsatz, das unter dem Begriff *Aktive Strahlen* bekannt ist [3]. Ausgehend von einem Referenzpunkt innerhalb der Objektkonturen werden die Grauwertverläufe auf Geraden durch diesen Punkt untersucht. Mit Hilfe einer Energieminimierung wird eine Punktkette ermittelt, die den tatsächlichen Objektkonturverlauf approximiert. Die Schulterkonturen werden separat extrahiert. Der Vergleich mit einer empirisch ermittelten Durchschnittskontur liefert einen Konfidenzwert für das Vorliegen einer Kopf-Schulterpartie.

Tiefeninformation zur Lokalisation der Person im Raum und zur Kompensation der Eigenbewegung des Roboters wird über ein Stereoverfahren mit automatischer regelmäßiger Rekalibrierung gewonnen.

In das sich anschließende strukturbasierte Auswahlverfahren fließen neben geometrischen und strukturellen Merkmalen der ermittelten „Regions of interest" die Ergebnisse der Bewegungs- und Konturdetektion sowie die ermittelte Raumposition ein. Anhand einer stochastischen Bewertungsfunktion findet anschließend die Auswahl der Kopfregionen statt.

3 Personendetektion mit Hilfe von Laserscans

Aus zulassungsrechtlichen Gründen sind zweidimensionale Laserscanner an Roboterfahrzeugen in der Regel in sehr geringem Abstand vom Boden angebracht. Für die Personendetektion stehen somit lediglich Abstandsinformationen in einer horizontalen Ebene in Beinhöhe zur Verfügung. Dies hat zur Folge, daß eine Diskriminierung von anderen Umgebungsobjekten alleinig anhand bestimmter Charakteristika menschlicher Beine erfolgen kann. In dem hier vorgestellten Ansatz wird dabei sowohl charakteristische Forminformation als auch Bewegungsinformation verwendet.

Im ersten Schritt wird ein eingelesener radialer Scan $S = (s_1, \ldots, s_n)$ nach einzelnen Beinen durchsucht. Dazu werden Scanintervalle $I_m = [i_m, j_m]$ ermittelt, an deren Rändern signifikante Tiefendifferenzen auftreten und die sich zudem von ihrer lokalen Umgebung durch geringere Tiefenwerte abheben:

$$(s_{i_m-1} - s_{i_m} > D) \wedge (s_{j_m+1} - s_{j_m} > D) \wedge (\forall i_m \le i < j_m : |s_i - s_{i+1}| \le D) \quad (1)$$

Anschließend werden die so gefundenen Segmente einer Plausibilitätsprüfung unterzogen, die zum einen die räumliche Ausdehnung des Bereichs in tangentialer Richtung, zum anderen die Form des Segments berücksichtigt. Segmente, die ein gewisses Konvexitätsmaß unterschreiten, werden ebenso verworfen wie Segmente, deren Größe eine zu starke Abweichung von üblichen Beinausmaßen aufweisen.

Die verbleibenden Segmente werden einem Bewertungsalgorithmus zugeführt, mit Hilfe dessen der Übergang von einzelnen Beinpositionen zu Personenpositionen vollzogen wird. Die Bewertung erfolgt zunächst in Abhängigkeit davon, ob zwei Beinpositionen so plaziert sind, daß sie zu einer Personenposition zusammengefaßt werden können. Darauf aufbauend wird die Änderung d bezüglich der im vorausgehenden Detektionsschritt ermittelten Position in Kombination mit der Stärke $m \in [0, 1]$ von aufgetretenen Bewegungen in dem betreffenden Bereich bewertet:

$$p(d, m) = \alpha \cdot \exp(-(1 - e^{-(d/\sigma)^2} - m)^2) + \beta \quad (2)$$

α und β sind Hilfswerte, die p auf den Wertebereich $[0, 1]$ skalieren. Die Wahl dieser Bewertungsfunktion gewährleistet, daß kontradiktorische Indikatoren – beispielsweise geringe Positionsänderung bei starker Bewegung – zu niedrigen Bewertungsergebnissen führen. Anhand der sich ergebenden Gesamtbewertungen wird unter Berücksichtigung der Ergebnisse der

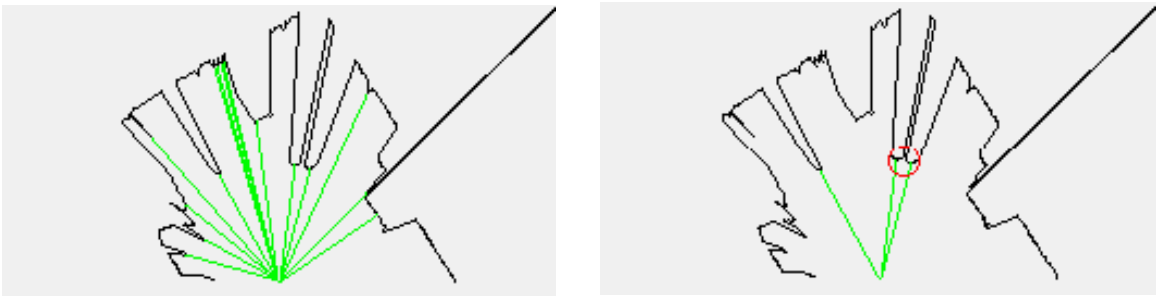

Abbildung 2: Zwischenergebnis und Resultat der laserbasierten Detektion.

visuellen Detektion diejenige Position ausgewählt, an der sich mit der größten Wahrscheinlichkeit eine Person befindet.

Zur Detektion von Bewegungen kommt ein Korrelationsverfahren zum Einsatz, das auf einer Trennung von Positions- und Rotationskorrelation basiert und somit einen Echtzeiteinsatz möglich macht [7]. Die initiale relative Position und Orientierung der zu korrelierenden Scans erfolgt unter Zuhilfenahme der Odometriedaten des Roboters. Dadurch ist bereits zu Anfang eine gute Schätzung der differentiellen Änderung gegeben, was eine Einschränkung des Suchbereichs der Korrelation und somit eine weitere Effizienzsteigerung bewirkt.

Zwischenergebnisse sowie das Resultat eines Detektionsvorgangs sind in Abbildung 2 dargestellt: links das Ergebnis nach der Grobsegmentierung, rechts die als Beine erkannten Segmente und die Position der erkannten Person.

Abbildung 3 zeigt das Systemverhalten für drei unterschiedliche Situationen. Links der Fall einer Person, die sich in frontaler Orientierung zum Roboter befindet, in der Mitte dieselbe Person in seitlicher Stellung. In beiden Fällen wird die Person korrekt detektiert. Befinden sich mehrere Personen in der Nähe des Roboters, werden alle Personen von der formbasierten Detektionskomponente registriert (Abbildung 3 rechts), vorausgesetzt daß keine Person durch ein Objekt oder eine andere Person im Scan vollständig verdeckt ist. Die Entscheidung, bei welcher Person es sich um die zu trackende handelt, wird unter Einbeziehung der Bewegungsdetektion und der Ergebnisse der visuellen Komponente in dem sich anschließenden Fusions- und Selektionsschritt getroffen.

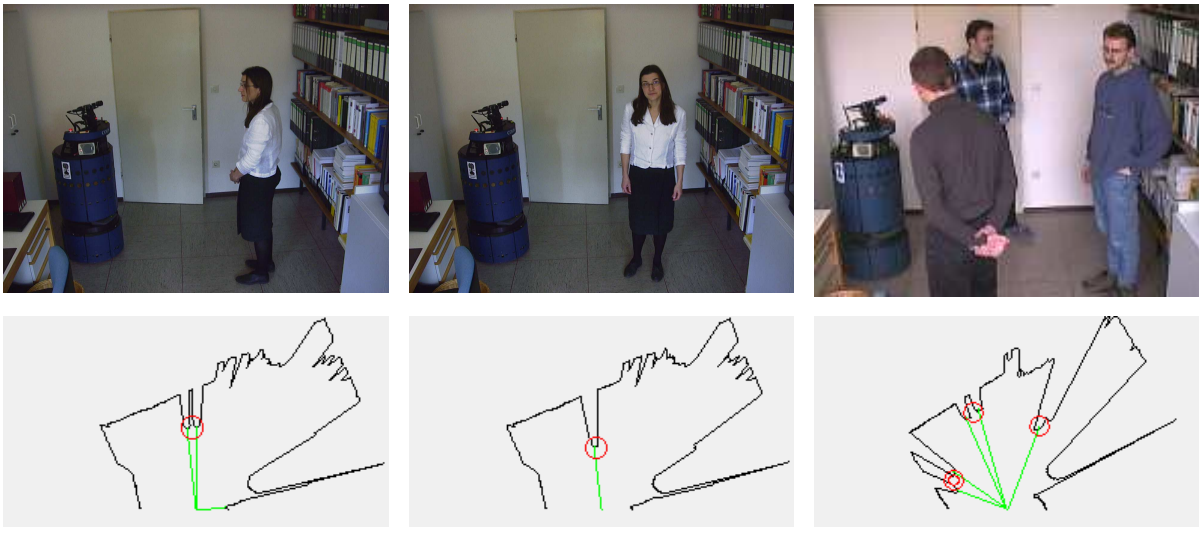

Abbildung 3: Ergebnisse der laserbasierten Personendetektion für verschiedene Situationen.

4 Robuste Lokalisation durch Fusion multisensorieller Information

Die bereits hohe Robustheit der visuellen wie auch der laserbasierten Detektion wird durch deren Kombination insofern weiter gesteigert, als sich die Rate der irrtümlicherweise als Person detektierten Objekte deutlich verringert. Dies setzt allerdings voraus, daß von der Person beide Detektionssysteme angesprochen werden, was abgesehen von Ausnahmen (z.B. im Fall von Rollstuhlfahrern oder Personen mit sehr langen Kleidern oder Röcken) im allgemeinen gewährleistet ist. Aufgrund der starken Abhängigkeit der Erkennungsleistung von der Ausprägung und der Dynamik der Umgebung sind allgemeine quantitative Aussagen an dieser Stelle nicht sinnvoll.

Der Fusionsvorgang selbst läuft in Form einer Maximumsuche innerhalb eines Potentialfeldes ab, das aus der Überlagerung zweidimensionaler Gaußfunktionen gebildet wird. Jede Funktion repräsentiert eine Position, an der entweder von der visuellen oder der laserbasierten Komponente eine Person detektiert wurde. Zusätzlich wird dem Feld eine Funktion an einer über ein Kalman-Filter aus den bisherigen Pfadpositionen geschätzten Position überlagert. Dadurch wird die zeitliche Kohärenz der Positionswerte mit in die Selektion einbezogen. Die Streuungsparameter der einzelnen Gaußfunktionen bestimmen sich durch die Genauigkeit der Positionsmessung. Diese ist bei Lasermessung im allgemeinen höher als bei Messungen auf Basis der Stereobildverarbeitung. Die Skalierungsfaktoren für die Funktionen ergeben sich durch die von den Detektionskomponenten gelieferten Konfidenzwerten.

Durch die Fusion ergibt sich neben einer Verbesserung der Erkennungsleistung auch ein Zuwachs an Flexibilität, insbesondere bezüglich möglicher Orientierungsänderungen der vorausgehenden Person relativ zum Roboter. Beispielsweise kann die farbbasierte Detektionskomponente keinen vernünftigen Beitrag zur Lokalisation mehr liefern, sobald die Person ihr Gesicht von den Kameras abwendet. In diesem Fall wird durch die Sensorfusion der Detektionsprozeß implizit von den übrigen Komponenten übernommen.

5 Folgeverhalten und Hindernisvermeidung

Die Ergebnisse der Erkennungskomponente werden in Bewegungen des Roboters mit dem Ziel umgesetzt, der erkannten Person zu folgen bzw. sich ihr bis auf einen definierten Abstand zu nähern. Zur Navigation wird auf eine echtzeitfähige Potentialfeld-Methode [6] zurückgegriffen, die um Hindernisse herum eine Potentialbarriere aufbaut und mit einer zum jeweiligen Zielpunkt hin gerichteten parabelförmigen Senke überlagert. Da für die hier vorgestellte Anwendung keine globale Hinderniskarte vorliegt, werden Hindernisse anhand der aktuellen Laserdaten durch lokale Minima im Laserscan modelliert.

Die aktuelle Zielposition wird einer Liste entnommen, welche die Raumpositionen enthält, die in regelmäßigen zeitlichen Abständen von der Lokalisierungskomponente geliefert werden. Ein spezielles Auswahlverfahren der Zielposition stellt sicher, daß keine Sackgassensituationen für den Roboter entstehen. Für Details sei an dieser Stelle auf [4] verwiesen.

6 Praktischer Einsatz

Das System kommt auf einem mobilen Roboter des Typs RWI B21 zum Einsatz. Der Roboter besitzt zwei Doppelprozessor-Pentium-II-PCs, von denen einer für die Bildverarbeitung, der andere für die Navigation und die laserbasierte Personendetektion verwendet wird. Die beiden Systemkomponenten kommunizieren über ein schnelles Interprozeß-Kommunikationsprotokoll. Die Bildakquisition erfolgt über zwei CCD-Farbkameras, die auf einer Pan-Tilt-Unit montiert sind. Der 2D-Laserscanner mit einem Abtastbereich von 180° und einer Winkelauflösung von 0,5° ist im Roboterrumpf in einer Höhe von 38 cm über dem Boden eingebaut.

Das System erreicht eine Verarbeitungsgeschwindigkeit von 8 Hz. Mit einer maximalen Fahrgeschwindigkeit von 0,5 m/s ist der Roboter damit in der Lage, einer Person schritthaltend zu folgen und dabei auch mit schwierigen Situationen wie beispielsweise engen Raumverhältnissen oder dynamischen Umgebungen zurechtzukommen. Abbildung 4 illustriert das Folgeverhalten anhand einer kurzen Bildsequenz.

Abbildung 4: Verfolgungsfahrt mit zweimaligem Passieren einer Türöffnung.

Literatur

[1] H.-J. Boehme, H.-M. Groß: Ein interaktives mobiles Service-System für den Baumarkt, in G. Schmidt, U. Hanebeck, F. Freyberger (Hrsg.): *Autonome Mobile Systeme 1999*, Informatik aktuell, Springer, S. 344-353, 1999.

[2] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, S. Thrun: The Museum Tour-Guide Robot RHINO, in H. Wörn, R. Dillmann, D. Henrich (Hrsg.): *Autonome Mobile Systeme 1998*, Informatik aktuell, Springer, S. 245-254, 1998.

[3] J. Denzler, H. Niemann: Active Rays: A New Approach to Contour Tracking, *International Journal of Computing and Information Technology*, Bd. 4, Nr. 1, S. 9-16, 1996.

[4] S. Feyrer, A. Zell: Tracking and Pursuing Persons with a Mobile Robot, *Proceedings of the International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time-Systems*, Korfu, Griechenland, S. 83-88, 1999.

[5] S. Feyrer, A. Zell: Detection, Tracking, and Pursuit of Humans with an Autonomous Mobile Robot, *Proceedings of the International Conference on Intelligent Robots and Systems*, Kyongju, Korea, S. 864-869, 1999.

[6] O. Khatib: Real-Time Obstacle Avoidance for Manipulators and Mobile Robots, *International Journal of Robotics Research*, Bd. 5, Nr. 1, S. 90-98, 1986.

[7] A. Mojaev, A. Zell: Online-Positionskorrektur für mobile Roboter durch Korrelation lokaler Gitterkarten, in H. Wörn, R. Dillmann, D. Henrich (Hrsg.): *Autonome Mobile Systeme 1998*, Informatik aktuell, Springer, S. 93-99, 1998.

[8] R. D. Schraft, C. Schaeffer, T. May, A. Traub: Mobiler Serviceroboter Care-O-bot® – ein Hilfssystem im häuslichen Bereich – Aufbau, Funktionsweise und Einsatzmöglichkeiten, in H. Wörn, R. Dillmann, D. Henrich (Hrsg.): *Autonome Mobile Systeme 1998*, Informatik aktuell, Springer, S. 234-244, 1998.