Videobasierte Erkennung von Personen und hindernisvermeidendes Folgeverhalten mit einem mobilen Serviceroboter

Stefan Feyrer, Andreas Zell

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

Zusammenfassung. Es wird ein System vorgestellt, das einen mobilen Serviceroboter in die Lage versetzt, Personen visuell zu detektieren und mittels Kamera- und Fahrbewegungen zu verfolgen. Die Detektion basiert auf einer Kombination von Farb-, Bewegungs- und Konturinformation. Ein Stereoverfahren zur Gewinnung von Tiefendaten ermöglicht eine Lokalisation der Person im dreidimensionalen Szenenraum. Der daraus ermittelte durchlaufene Pfad der vorangehenden Person wird ebenso für die Navigation herangezogen wie eine über Abstandssensoren gewonnene lokale Umgebungskarte, mit Hilfe derer stationäre und dynamische Hindernisse umgangen werden können.

1 Einleitung

Die Einsatzmöglichkeiten mobiler Serviceroboter sind vielfältig. Allen gemeinsam ist jedoch die Notwendigkeit von Mensch-Maschine-Schnittstellen, die ein möglichst natürliches Interagieren mit dem Roboter erlauben. Herkömmliche Ein-/Ausgabegeräte wie Tastatur, Maus und Bildschirm sind dafür denkbar ungeeignet. Eine einfache und intuitive Bedienung bedarf vielmehr anderer, natürlicher Kommunikationsmedien. Eine Reihe von Forschergruppen beschäftigt sich deshalb intensiv mit der technischen Realisierung solcher Schnittstellen. Das Verstehen natürlicher Sprache [7] fällt dabei ebenso in diesen Bereich wie die Erkennung von Personen und die Interpretation menschlicher Gesten [1] [5] [16]. Für einen mobilen Serviceroboter ist es dabei von besonderer Wichtigkeit, eine derartige Erkennungsleistung nicht nur passiv zu vollziehen, sondern sie aktiv in seinen aktuellen Aufgabenkontext einzubeziehen.

Das hier beschriebene System leistet einen Beitrag in diese Richtung, indem es eine visuell arbeitende Komponente zur Erkennung von Personen mit einer abstandssensorbasierten Navigationseinheit kombiniert. Aufgabe der Navigationskomponente ist es, einer erkannten Person zu folgen. Dies ist beispielsweise dann sinnvoll, wenn der Roboter zu einem bestimmten Einsatzort geführt oder mit einer ihm unbekannten Umgebung vertraut gemacht werden soll. Die Orientierung erfolgt dabei sowohl anhand des von der Person durchlaufenen Pfades als auch mit Hilfe einer lokalen Umgebungskarte, die über Abstandssensoren ermittelt wird. Stationäre und dynamische Hindernisse können somit umgangen werden.

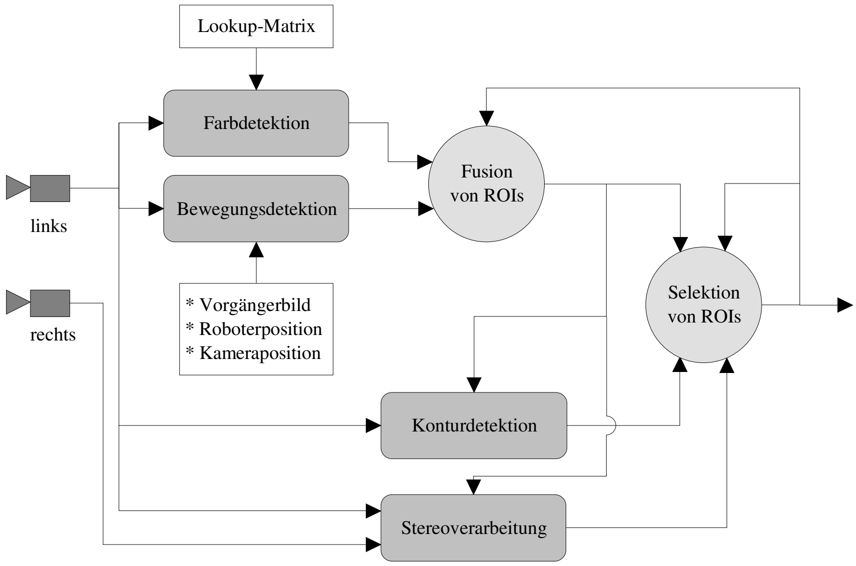

Abbildung 1. Struktureller Aufbau der Detektionskomponente

Die beiden Hauptkomponenten des Systems – Detektion und Blicksteuerung auf der einen, Navigation und Fahrsteuerung auf der anderen Seite – sind nicht nur auf der logisch-strukturellen Ebene als zwei eigenständige Einheiten zu betrachten. Auch auf der Implementierungsebene sind sie als zwei separate Prozesse realisiert, die über ein schnelles Interprozeß-Kommunikationsprotokoll Informationen austauschen. Diese Entkopplung hat den Vorteil, daß die Kameranachführung und somit die Blickfixierung auch bei schnelleren Bewegungen der Person schritthaltend erfolgt, während die Navigationskomponente davon unabhängig die Auswertung der Abstandssensoren und die Pfadplanung in einer dafür optimalen Verarbeitungsfrequenz vornehmen kann.

Im folgenden wird zunächst die Detektionskomponente beschrieben. Anschließend wird in jeweils eigenen Abschnitten auf die Stereoverarbeitung und die Kamerasteuerung eingegangen. Nachfolgend werden die bei der Navigation eingesetzten Verfahren vorgestellt. Ergebnisse und Erfahrungen beim praktischen Einsatz des Systems sowie ein Ausblick schließen den Beitrag ab.

2 Visuelle Detektion von Personen

Während in früheren Systemen zur visuellen Detektion von Personen oftmals lediglich ein einzelnes Erkennungsmerkmal verwendet wurde [17] [18], kommen heute zunehmend Kombinationen mehrerer Merkmale zum Einsatz [2] [12]. Dadurch wird die Robustheit erhöht und eine größere Unabhängigkeit von einschränkenden Randbedingungen, beispielsweise Beleuchtungsverhältnissen, erzielt. Auch in dem hier vorgestellten System erfolgt eine mehrschichtige Detektion. Neben einer Hautfarberkennung kommt eine Detektion bewegter Objekte und ein Verfahren zur Extraktion von Konturpunktketten zum Einsatz. Darüber hinaus fließt die durch ein Stereoverfahren gewonnene Abstandsinformation in Form einer zusätzlichen Plausibilitätsprüfung mit ein. Eine schematische Übersicht der Detektionskomponente ist in Abb. 1 gegeben.

2.1 Farbbasierte Detektion

Die Detektion von Hautfarbe geschieht anhand der in den HSI-Raum transformierten Farbwerte der einzelnen Bildpunkte. Da für die Klassifikation nur die reine Farbinformation von Bedeutung ist, wird der Intensitätsanteil vernachlässigt. Diejenigen Elemente im helligkeitsinvarianten HS-Unterraum, bei denen es sich um Hautfarbwerte handelt, sind durch eine Lookup-Matrix definiert. Diese Matrix ist das Ergebnis einer manuellen Segmentierung von mehr als 50 Bildsequenzen, die mit 15 Personen bei unterschiedlichen Beleuchtungsverhältnissen aufgenommen wurden. Da sich die Bereiche im HS-Raum weder durch eine Normalverteilung beschreiben ließen noch durch Überdeckungen von geometrischen Primitiven einfach abzugrenzen waren, hat sich die Verwendung einer Lookup-Matrix als notwendig erwiesen [6]. Die konkrete Realisierung der Farbdetektion verwendet eine in den RGB-Raum zurücktransformierte Tabelle, um die aufwendige Online-Umrechnung der RGB- in HSI-Werte zu umgehen.

2.2 Bewegungsdetektion

Das Problem bewegungsorientierter Detektionsverfahren ist, daß bei nicht stationärer Kamera zur Extraktion bewegter Vordergrundobjekte eine Kompensation der Eigenbewegung stattfinden muß, die einen nicht unerheblichen rechnerischen Aufwand erfordert. Aus diesem Grund werden bei diesem System zwei Fälle unterschieden.

Bewegt sich lediglich die auf einer Pan-Tilt-Unit montierte Kamera, nicht aber der gesamte Roboter, so findet eine Bewegungsdetektion mittels eines Differenzbildverfahrens statt, das mit zwei bei temporär stationärer Kamera aufgenommenen Bildern arbeitet. Dies erfordert eine sakkadische Kameranachführung, hat aber den Vorteil, daß sich wesentlich weniger Fehldetektionen ergeben.

Bewegen sich sowohl die Pan-Tilt-Unit als auch der Roboter, wird eine Eigenbewegungskompensation durchgeführt, die sich an [4] und [14] anlehnt. Dabei wird zunächst die Positions- und Orientierungsänderung der Kamera bezüglich des Zeitpunkts, zu dem das Vorgängerbild aufgenommen wurde, berechnet und in Form einer homogenen Transformationsmatrix dargestellt:

$$T = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

Dies ist leicht möglich, da die Bewegungen von Kamera und Roboterbasis vom Programmsystem selbst gesteuert werden und somit bekannt sind. Anschließend wird mittels folgender Gleichungen ein Prädiktionsbild für die neue Kameraposition berechnet, welches anschließend mit dem tatsächlich aufgenommenen Bild verglichen wird.

$$x_2 = f \cdot \frac{r_{11}x_1 + r_{12}y_1 + fr_{13} + f\frac{t_x}{Z}}{r_{31}x_1 + r_{32}y_1 + fr_{33} + f\frac{t_z}{Z}} \qquad y_2 = f \cdot \frac{r_{21}x_1 + r_{22}y_1 + fr_{33} + f\frac{t_y}{Z}}{r_{31}x_1 + r_{32}y_1 + fr_{33} + f\frac{t_z}{Z}} \tag{2}$$

x_1 und y_1 sind dabei die Bildkoordinaten des Originalbildes, x_2 und y_2 beziehen sich auf das Prädiktionsbild. Mit f ist die Brennweite der verwendeten Kamera bezeichnet. Der Wert Z gibt den zum Vorgängerzeitpunkt gültigen Abstand zwischen Kamera und Szenenobjekt für den jeweiligen Bildpunkt an und wird über das in Abschnitt 3 beschriebene Stereoverfahren ermittelt. Die Berechnung der Abstandswerte erhöht zusätzlich die Komplexität der Berechnungen und erfordert deshalb an dieser Stelle eine Reduktion der verwendeten Bildauflösung.

2.3 Segmentierung, Fusion und Selektion von Regionen

Zur Segmentierung der detektierten Bereiche wird ein Region-growing-Verfahren eingesetzt, das verzahnt mit einer schnellen Variante des Flood-fill-Algorithmus und einem morphologischen Open-Filter abläuft [6].

Die so gewonnenen *Regions of interest* (ROIs) werden anschließend einer Fusion unterzogen, die speziell oftmals durch die Augenlinie unterbrochene Gesichtsbereiche zusammenfügt. Dazu wird ein Algorithmus verwendet, der auf einer beschränkten Tiefentraversierung eines Nachbarschaftsgraphen von ROIs basiert. Jede im Laufe der Traversierung auftretende Kombination von benachbarten Regionen wird mittels des im folgenden beschriebenen Bewertungsverfahrens untersucht. Diejenige Kombination mit der höchsten Bewertung wird fusioniert und aus dem Adjazenzgraphen entfernt.

Das strukturbasierte Auswahlverfahren arbeitet mit dem Ziel, Kopfbereiche von anderen Segmentierungsclustern zu trennen. Neben geometrischen Kriterien wie dem Breiten-Höhen-Verhältnis des umgebenden achsenparallelen Rechtecks und der Kompaktheit des Objekts wird der Durchschnitt der relativen Häufigkeiten der HS-Hautfarbkombinationen der zum Bereich gehörenden Bildpunkte herangezogen. Dadurch wird die Abweichung von der durch die manuelle Segmentierung ermittelte „mittlere Hautfarbe“ mitbewertet. Die Kohärenz von Bildfolgen wird insofern ausgenutzt, als zusätzlich die sich ergebende Größenveränderung und der Abstand der Schwerpunkte bezüglich des im Vorgängerbild selektierten Kopfbereichs einfließen.

Die endgültige Selektion eines Kopfbereiches erfolgt unter Verwendung der Ergebnisse der Bewegungs- und Konturdetektion sowie der mittels der Stereoverarbeitung gewonnenen Abstandsinformation.

2.4 Integration von Konturinformation

Zur Detektion von Konturen kommt ein Verfahren zum Einsatz, das unter dem Begriff *Aktive Strahlen* bekannt ist [3]. Ausgangspunkt ist ein Referenzpunkt innerhalb der Objektkonturen, der sich idealerweise in der Nähe des Objektschwerpunkts befinden sollte. In diesem Fall werden dazu die Mittelpunkte der von der Farbdetektion gelieferten ROIs verwendet. Untersucht werden nun die Grauwertverläufe auf Geraden durch diesen Punkt. Mit Hilfe einer Energieminimierung wird eine Punktkette ermittelt, die den tatsächlichen Objektkonturverlauf annähert. Eine Besonderheit des hier eingesetzten Verfahrens ist, daß bei der Minimierung nicht nur Grauwertkonturen, sondern auch die Konturen einbezogen werden, die sich aus dem von der Bewegungsdetektion ermittelten binären Differenzbild ergeben. Einzelheiten dazu finden sich in [6]. Die berechneten Konturverläufe werden anschließend bezüglich ihrer Ähnlichkeit zu einer statistisch ermittelten Durchschnittskontur bewertet. Abb. 2 zeigt die ermittelten Konturverläufe in einer Beispielsequenz.

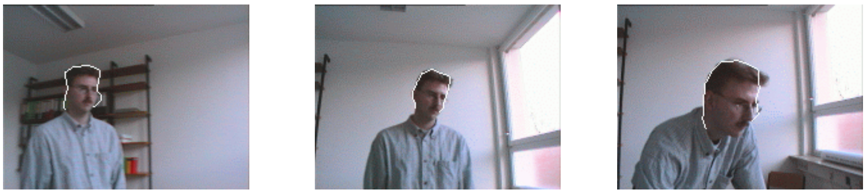

Abbildung 2. Detektierte Konturpunktketten, verbunden durch Polygonzüge

3 Stereobildverarbeitung

Jede visuelle Wahrnehmungsleistung benötigt ab einem bestimmten Komplexitätsgrad neben zweidimensionalen Bilddaten auch Informationen über den räumlichen Aufbau der betrachteten Szene. Die Gewinnung von Tiefeninformation spielt in diesem System eine besonders wichtige Rolle, da erst durch eine Lokalisation der Person im Raum eine Verfolgung des von ihr zurückgelegten Pfades stattfinden kann. Ferner kann die Abstandsinformation einen wichtigen Beitrag im Selektionsprozeß leisten, da davon ausgegangen werden kann, daß zwischen zwei nacheinander aufgenommenen Bildern keine signifikante Änderung des Abstandes zur Person auftritt. Für beide Fälle – Pfadermittlung und Plausibilitätsprüfung – ist es lediglich erforderlich, innerhalb der extrahierten ROIs Abstandsdaten zu berechnen. Wesentlich aufwendiger wird die Stereoverarbeitung, wenn für alle Bildpunkte Abstandswerte vorliegen müssen, wie dies bei der Kompensation der Eigenbewegung der Fall ist.

3.1 Stereogeometrie und Kalibrierung

Die Gewinnung von Tiefeninformation basiert bei diesem System auf einer Standard-Stereogeometrie [9]. Das bedeutet, daß die optischen Achsen der beiden Kameras parallel verlaufen und die Zeilen beider Bilder parallel zur Verbindungsgeraden der optischen Zentren liegen. Der Vorteil, daß epipolare Linien entlang der Bildzeilen verlaufen und somit eine einfache Korrespondenzanalyse möglich ist, wird erkauft durch eine relativ hohe Empfindlichkeit des Aufbaus gegenüber Abweichungen von der Idealgeometrie. Bereits geringe Richtungsdifferenzen der optischen Achsen, vor allem in vertikaler Richtung, führen zu inakzeptablen Ergebnissen. Deshalb ist eine automatische Kalibrierung des Stereoaufbaus unabdingbar. Dazu wird zunächst das Bild in Spalten gleicher Breite aufgeteilt. Innerhalb jeder Spalte werden mit Hilfe des Moravec-Feature-Detektors [13] die signifikantesten Punkte ermittelt. Für diese Punkte wird eine Korrespondenzanalyse durchgeführt, die nicht nur auf einer Bildzeile, sondern auf einem aus mehreren Bildzeilen bestehenden Streifen arbeitet. Die beste Korrespondenz liefert somit für jeden Featurepunkt einen vertikalen Versatz. Der maximale und minimale Versatz jeder Spalte wird verworfen, alle anderen Werte werden einer linearen Regression zugeführt. Die eigentliche Korrespondenzanalyse wird unter Berücksichtigung des Versatzes, der durch die entstehende Regressionsgerade definiert ist, durchgeführt. Sowohl Justierungsfehler in vertikaler Richtung als auch Rotationen um die optischen Achsen werden somit kompensiert (siehe Abb. 3 rechts).

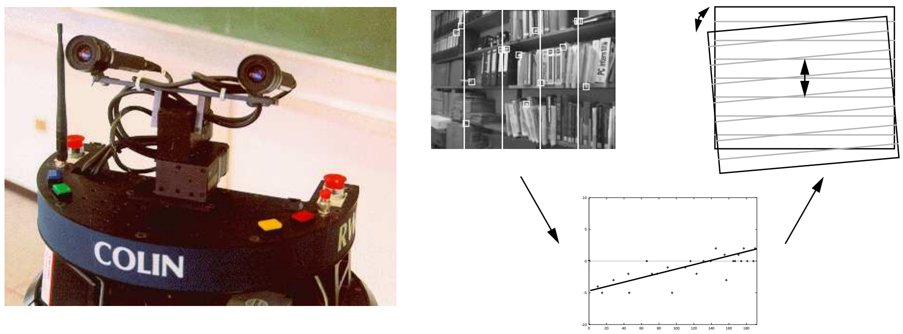

Abbildung 3. links: Kameraplattform mit Schwenk-Neige-Einheit; rechts: Kalibrierung der Stereogeometrie

3.2 Korrespondenzanalyse

Es gibt eine Reihe von unterschiedlichen Möglichkeiten, die Korrespondenzanalyse durchzuführen [11]. In diesem System erfolgt sie mittels einer Korrelation von helligkeitsnormierten Bildausschnitten der Größe 5x5. Als Korrelationsmaß wird folgende Funktion verwendet [8]:

$$s = \max(0, 1 - c) \tag{3}$$

mit

$$c = \frac{\sum_{i,j}((L(x+i, y+j) - \bar{L}) - (R(x+dx+i, y+dy+j) - \bar{R}))^2}{\sqrt{\left(\sum_{i,j}(L(x+i, y+j) - \bar{L})^2\right) \cdot \left(\sum_{i,j}(R(x+dx+i, y+dy+j) - \bar{R})^2\right)}} \tag{4}$$

Dabei bezeichnen L und R die linke bzw. rechts Bildmatrix, $\bar{L}$ und $\bar{R}$ den Mittelwert der Grauwerte innerhalb des jeweiligen Korrelationsfensters. Die Indizes i und j laufen über die Spalten bzw. Zeilen des Korrelationsfensters.

Um falsch ermittelte Korrespondenzen weitestgehend zu eliminieren, werden die ermittelten Tiefenbilder mit einem Medianfilter und diversen morphologischen Operatoren nachbearbeitet. In Abb. 4 sind ein Stereobildpaar und das daraus ermittelte Tiefenbild dargestellt.

Abbildung 4. Stereobildpaar mit daraus ermitteltem Tiefenbild

4 Blicksteuerung

Die verwendete binokulare Kameraplattform (Abb. 3 links) besitzt zwei Freiheitsgrade, die durch einen Schwenkwinkel ψ und einen Nickwinkel θ beschrieben sind. Die Steuerung der Kamerabewegung erfolgt sakkadenartig. Geht man von dem Modell einer Lochkamera aus und setzt man zusätzlich voraus, daß sowohl die Schwenk- als auch die Nickachse durch das Projektionszentrum verlaufen, so lassen sich die nötigen Rotationswinkel für die Zentrierung eines Objekts in der Bildmitte durch folgende Gleichungen bestimmen:

$$\psi = \Delta x \cdot \frac{\psi_{max}}{w} \qquad \theta = \Delta y \cdot \frac{\theta_{max}}{h} \tag{5}$$

w und h bezeichnen dabei die Bildbreite und -höhe in Pixel. Δx und Δy sind die Abweichungen des Mittelpunkts des detektierten Objekts bezüglich der Bildmitte. Die beiden Winkelmaximalwerte werden durch eine Kalibrierung des Kamerasystems gewonnen und beschreiben den Öffnungswinkel des sichtbaren Bereichs.

Bei der verwendeten Kameraplattform sind Schwenk- und Nickachse um jeweils 6 cm bezüglich des optischen Zentrums versetzt. Der dadurch entstehende Fehler ist durch folgende Gleichung, die durch einfache geometrische Überlegungen hergeleitet werden kann, definiert:

$$\Delta l = d(\sin\beta\tan\beta + \cos\beta - 1) \tag{6}$$

d bezeichnet den Versatz bezüglich des Projektionszentrums, β einen mittels (5) berechneten Winkel. Konkret ergibt sich für die Schwenkachse ein maximaler Fehler von 0,4 cm, für die Nickachse von 0,3 cm (jeweils in Weltkoordinaten). Beide Werte liegen unterhalb der Meßgenauigkeit selbst bei nahen Objekten. Aus diesem Grund wird auf eine Kompensation des Achsenversatzes verzichtet.

5 Navigation

Die Navigationseinheit setzt die Ergebnisse der Erkennungskomponente in Bewegungen des Roboters um mit dem Ziel, der erkannten Person zu folgen bzw. sich ihr bis auf einen Abstand von ca. einem Meter zu nähern. Zur Navigation wird auf die sogenannte Potentialfeld-Methode [10] [15] zurückgegriffen. Es handelt sich dabei um eine Echtzeit-Pfadplanungsmethode, die mittels der Auswertung der jeweils aktuellen Sensorwerte einem vorgegebenen Pfad kollisionsfrei folgt. Dazu wird ein Potentialfeld definiert, das sich aus einer anziehenden und einer abstoßenden Komponente zusammensetzt:

$$U = U_{attr} + U_{rep} \tag{7}$$

Das abstoßende Potential baut um Hindernisse herum eine Potentialbarriere auf, die der Roboter nicht überwinden kann und die den Pfad möglichst weit an den Hindernissen vorbeiführt. Da keine globale Hinderniskarte vorliegt, fließen die von den Sensoren gelieferten Abstandswerte ρ_k direkt in die Berechnung des abstoßenden Potentials ein:

$$U_{rep} = \sum_{k=1}^{r} U_k \qquad U_k = \begin{cases} \frac{1}{2}\eta\left(\frac{1}{\rho_k} - \frac{1}{\rho_0}\right)^2 & \text{für} \quad \rho_k \le \rho_0 \\ 0 & \text{sonst} \end{cases} \tag{8}$$

ρ_0 ist die Entfernung, bis zu welcher ein Hindernis noch Einfluß auf die Fahrt des Roboters nimmt. Das anziehende Potential wird durch eine parabelförmige Senke definiert:

$$U_{attr} = \frac{1}{2}\xi\rho_{goal}^2 \tag{9}$$

ρ_{goal} ist dabei der Abstand zum Zielpunkt. η und ξ sind positive Skalierungsfaktoren, die das Verhältnis zwischen anziehendem und abstoßendem Potential festlegen. Der Zielpunkt wird ermittelt, indem alle Punkte des Pfades bezüglich ihrer direkten Erreichbarkeit überprüft werden. Unter den direkt erreichbaren Punkten wird der aktuellste ausgewählt. Ist kein Punkt des Pfades direkt erreichbar, wird als Zielpunkt diejenige Position verwendet, an der zuletzt eine Person erkannt wurde.

Die anzustrebende Bewegungsrichtung des Roboters ergibt sich aus der Richtung einer aus dem Potential ermittelten Kraft:

$$\vec{F} = -\vec{\nabla} U \tag{10}$$

Die Translationsgeschwindigkeit wird innerhalb bestimmter Bereichsgrenzen in linearer Abhängigkeit vom Abstand zur jeweiligen Zielposition geregelt, falls sich keine Hindernisse in unmittelbarer Nähe befinden. Sind Hindernisse vorhanden, wirken sich diese hemmend auf die Translationsgeschwindigkeit aus.

Sobald der Roboter eine im Pfad verzeichnete Position mit genügender Genauigkeit erreicht hat, wird diese und alle vor ihr im Pfad liegenden Einträge aus der entsprechenden Datenstruktur entfernt.

6 Ergebnisse und Ausblick

Das System kommt auf einem mobilen Roboter des Typs RWI B21 zum Einsatz. Der Roboter besitzt zwei Doppelprozessor-PentiumPro-PCs, von denen einer für die Bildverarbeitung, der andere für die Navigation verwendet wird. Die Bildakquisition erfolgt über zwei CCD-Farbkameras, die auf einer Pan-Tilt-Unit montiert sind. Für die Navigation können wahlweise die 24 ringförmig angebrachten Ultraschallsensoren oder der eingebaute 2D-Laserscanner verwendet werden.

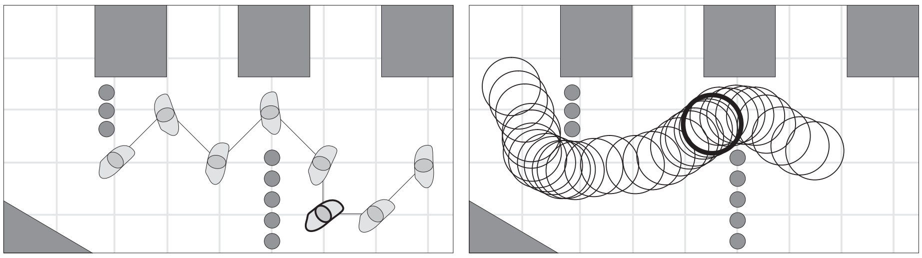

Abbildung 5. links: zurückgelegter Weg der Person, rechts: Bahn des nachfolgenden Roboters

Bei einer Auflösung von 192x144 Bildpunkten erreicht das System eine Verarbeitungsgeschwindigkeit von 8 Bildern pro Sekunde bei stationärem Roboter. Durch die bei bewegter Roboterbasis nötige Kompensation der Eigenbewegung sinkt die Verarbeitungsfrequenz auf 4 Bilder pro Sekunde. Damit ist ein schritthaltendes Tracking möglich, wenn die Person keine ungewöhnlich schnellen Bewegungen ausführt. Die Geschwindigkeit der Fahrbewegungen hängt stark von der Umgebung und insbesondere von der jeweiligen Distanz zu Hindernisobjekten ab. Momentan läßt die Verarbeitungsgeschwindigkeit noch kein Nachfolgen in Schrittgeschwindigkeit zu.

Abb. 5 zeigt die Aufzeichnung einer Beispielfahrt. Links ist der zurückgelegte Weg der Person dargestellt, rechts die Bahn des nachfolgenden Roboters. Um das zeitliche Verhältnis der Bewegungen zu illustrieren, ist eine Momentaufnahme hervorgehoben, die den Roboter noch vor und die Person bereits hinter einem Hindernis zeigt. Die Person hat sich während der Aufzeichnung an am Boden befindlichen Markierungsstreifen orientiert und befand sich stets dem Roboter zugewandt. Diese Notwendigkeit resultiert aus der starken Abhängigkeit der Detektionskomponente von der Farberkennung.

Um zu erreichen, daß Personen auch von hinten detektiert werden, muß eine Flexibilisierung der Detektion dahingehend vorgenommen werden, daß sie sich nach einer Initialisierungsphase stärker auf andere, eventuell zusätzliche Merkmale stützt. Weitere Schwerpunkte der Weiterentwicklung werden eine Verbesserung der Bildfolgenanalyse und eine Positionsprädiktion im dreidimensionalen Szenenraum sein.

Literatur

1. H.-J. Böhme, A. Brakensiek, U.-D. Braumann, M. Krabbes, H.-M. Groß: *Visually-Based Human-Machine-Interaction in a Neural Architecture*, Workshop SOAVE 97, Selbstorganisation von adaptivem Verhalten, Fortschritt-Berichte VDI, Reihe 8, Nr. 663, S. 166-175, 1997
2. J. L. Crowley, F. Berard: *Multi-Modal Tracking of Faces for Video Communications*, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997
3. J. Denzler: *Aktives Sehen zur Echtzeitobjektverfolgung*, Dissertation, Technische Fakultät der Universität Erlangen-Nürnberg, 1997
4. M. Ebner: *Extraction of Moving Objects with a Moving Mobile Robot*, 3rd IFAC Symposium on Intelligent Autonomous Vehicles, Madrid, 1998
5. S. Eickeler, G. Rigoll: *Kontinuierliche Erkennung von spontan ausgeführten Gesten mit neuen stochastischen Dekodierverfahren*, Workshop Dynamische Perzeption, Proceedings in Artificial Intelligence, Bd. 8, infix-Verlag, S. 175-182, 1998
6. S. Feyrer, A. Zell: *Ein integrierter Ansatz zur Lokalisation von Personen in Bildfolgen*, Workshop Dynamische Perzeption, Proceedings in Artificial Intelligence, Bd. 8, infix-Verlag, S. 183-190, 1998
7. C. Fischer, P. Havel, G. Schmidt, J. Müller, H. Stahl, M. Lang: *Kommandierung eines Serviceroboters mit natürlicher, gesprochener Sprache*, in G. Schmidt, F. Freyberger: Autonome Mobile Systeme 1996, Informatik aktuell, Springer, S. 248-259, 1996
8. P. Fua: *A Parallel Stereo Algorithm That Produces Dense Depth Maps and Preserves Image Features*, Machine Vision and Applications, Bd. 6, Nr. 1, S. 35-49, 1993
9. X. Jiang, H. Bunke: *Dreidimensionales Computersehen – Gewinnung und Analyse von Tiefenbildern*, Springer 1997
10. O. Khatib: *Real-Time Obstacle Avoidance for Manipulators and Mobile Robots*, International Journal of Robotics Research, Bd. 5, Nr. 1, S. 90-98, 1986
11. K. Konolige: *Small Vision Systems: Hardware and Implementation*, 8th International Symposium on Robotics Research, Hayama, Japan, 1997
12. S. McKenna, S. Gong: *Tracking Faces*, Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, S. 271-277, 1996
13. H. Moravec: *Towards Automatic Visual Obstacle Avoidance*, Proceedings of the 5th International Joint Conference on Artificial Intelligence, Vision-I, S. 584, 1977
14. D. Murray, A. Basu: *Motion Tracking with an Active Camera*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Bd. 16, Nr. 5, S. 449-459, 1994
15. J.-C. Latombe: *Robot Motion Planning*, Kluwer Academic Publishers Group, 1991
16. R. Rae, H. J. Ritter: *Recognition of Human Head Orientation Based on Artificial Neural Networks*, IEEE Transactions on Neural Networks, Bd. 8, Nr. 2, S. 257-265, 1998
17. K. Rohr: *Towards Model-Based Recognition of Human Movements in Image Sequences*, CVGIP: Image Understanding, Bd. 59, Nr. 1, S. 94-115, 1994
18. M. Sommerau, G. Mamier, A. Zell, M. Vogt, P. Levi: *Fast Face Localization and Tracking with Model-Based Time Synchronization of a Head System*, in G. Sagerer, S. Posch, F. Kummert (Hrsg.): Mustererkennung 1995, Springer, S. 516-523, 1995