

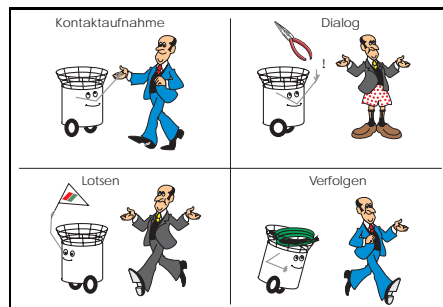
# Ein Interaktives Mobiles Service-System für den Baumarkt\*

Hans-Joachim Boehme & Horst-Michael Gross

Fachgebiet Neuroinformatik, Technische Universität Ilmenau,  
98684 Ilmenau (Thüringen)  
email: {hans,homi}@informatik.tu-ilmenau.de

**Zusammenfassung** In diesem Beitrag wird das Konzept für den Einsatz eines Roboters als intelligenter, interaktiver Shopping-Assistent in einer Baumarkt-Umgebung vorgestellt und diskutiert. Vorrangig behandelt werden die Schwerpunkte *visuell basierte Mensch-Roboter-Interaktion, lokale und globale Roboternavigation* sowie Fragen der *Steuerarchitektur zur Verhaltenskoordination* vor dem Hintergrund dieses realen Einsatzszenarios. Als Forschungsplattformen kommen dabei die beiden Robotersysteme MILVA und PERSES des Fachgebiets zum Einsatz. Betont werden soll an dieser Stelle, daß es sich hier um ein langfristig angelegtes Forschungsvorhaben handelt und lediglich erste methodische Konzepte und vorläufige Ergebnisse aus der realen Baumarkt-Umgebung präsentiert werden können.

## 1 Einleitung und Anwendungsszenario



**Abb. 1.** Veranschaulichung der elementaren Verhaltensleistungen eines mobilen interaktiven Shopping-Assistenten

Um möglichst schnell die Einbettung der Forschungsaktivitäten in ein reales Anwendungsfeld zu gewährleisten, streben wir den Einsatz einer mobilen Plattform als intelligenten und interaktiven Shopping-Assistenten in einer

Baumarkt-Umgebung an. Ein deutsches Baumarktunternehmen als Kooperationspartner stellt uns zu diesem Zweck eine Filiale für Testuntersuchungen zur Verfügung. Vor diesem Szenario lassen sich verschiedene Teilleistungen eines solchen Systems definieren, die in Abb. 1 veranschaulicht sind und insbesondere das Auffinden und kontinuierliche Verfolgen eines potentiellen Benutzers, die robuste lokale und globale Navigation mittels Kameras und/oder Entfernungssensoren, die Integration aller Verhaltensmodule des Systems in eine geeignete Steuerarchitektur, die Artikulation des Systems gegenüber dem Benutzer sowie

\* gefördert durch die Projekte PERSES (TMWFK, Nr. B 611-98041) und DYGEST (EU TMR, Nr. ERB FMBI CT 97 2613)

eine Reihe von baumarktspezifischen Service-Angeboten umfassen. Der aktuelle Realisierungsstand dieser Teilleistungen, die dabei zur Anwendung kommenden Methoden sowie die Funktionalität des Gesamtsystems werden in den folgenden Abschnitten dargestellt.

Abb. 2 zeigt die beiden Robotersysteme MILVA und PERSES, die als Forschungsplattformen des Fachgebiets zum Einsatz kommen und jeweils über mehrere Kameras, Entfernungssensorik und On-Board-Rechentechnik verfügen.



**Abb. 2.** Links: MILVA, eine Plattform mit Dreiradkinematik, Front(Navigations)kamera, Doppelkamarasystem mit PTU's und schwenkbarem „Hals“, Stereomikrofonanordnung, Laserscanner, Bumperleiste  
Rechts: PERSES, eine B21-Plattform (RWII, USA) mit Ultraschallsensorik, Stereokamerasystem und Omnikamera

## 2 Personenlokalisierung und Personentracking

Unter weitgehend uneingeschränkten Umgebungsbedingungen ist das Auffinden und Verfolgen von Personen in Bildsequenzen in Echtzeit auch zum aktuellen Zeitpunkt noch eine sehr anspruchsvolle Aufgabenstellung. Während sich zunächst die Mehrzahl der zu diesem Problem publizierten Beiträge auf den Einsatz eines speziellen Merkmals (Hautfarbe, Gesicht, Bewegung etc.) beschränkte und dabei entweder weitgehend definierte, vorab bekannte oder zumindest näherungsweise unveränderliche Randbedingungen voraussetzte, verwendet man heute typischerweise Kombinationen verschiedener visueller Cues, um die notwendige Robustheit unter variablen realen Umgebungsbedingungen zu erzielen ([5]). Vor diesem Hintergrund wurde ein MULTI-CUE-ANSATZ entwickelt und implementiert, der parallel verschiedene visuelle Cues extrahiert und diese zu einem personenspezifischen Auffälligkeitssystem fusioniert (siehe auch [2]).

## 2.1 Multi-Cue-Ansatz zur Personenlokalisierung

Einen Überblick über den implementierten Multi-Cue-Ansatz zur Personenlokalisierung gibt Abb. 3. Auflösungspyramiden (Abtastverhältnis  $1/\sqrt{2}$ ) transformieren das Bild in eine Multiskalenrepräsentation mit 5 Ebenen.

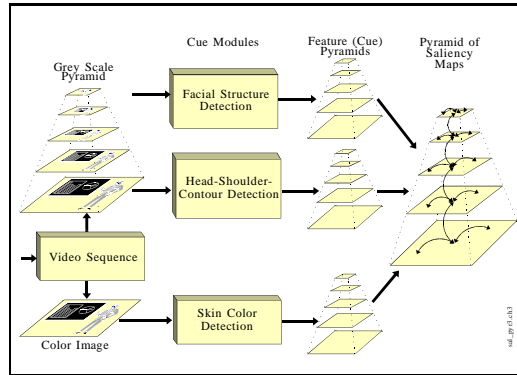


Abb. 3. Schematischer Aufbau des personenspezifischen Auffälligkeitssystems

Die beiden Cue-Module, die strukturbasiert arbeiten und für die Detektion einer *Gesichts* bzw. einer *Kopf-Schulter-Silhouette* verantwortlich sind, operieren auf allen Ebenen der Auflösungspyramide, während die Detektion von *Hautfarbe* im

Originalbild erfolgt. Danach wird das hautfarbsegmentierte Bild ebenfalls in eine Pyramide transformiert, um eine einheitliche Datenstruktur zur Fusion der einzelnen Cues zu erhalten. Die folgenden Abschnitte geben eine Kurzbeschreibung der einzelnen Cue-Module.

**Hautfarbsegmentierung** Der von uns implementierte Ansatz lehnt sich an [16] und [7] an und verwendet die chromatische r-g-Farbebene des RGB-Farbraumes, um Intensitätsunterschiede zu kompensieren. Ausgehend von einem manuell erstellten Hautfarb-Datensatz (verschiedene Personen unter verschiedenen Beleuchtungsbedingungen) wurde ein Modell dieser kompakten Datenverteilung, bestehend aus Mittelwert und Kovarianzmatrix, berechnet. Die eigentliche Hautfarbsegmentierung erfolgt über die Mahalanobis-Distanz des aktuellen Farbwerts zum Hautfarbmodell. Abb. 4 zeigt ein Beispiel zur Hautfarbklassifikation.



Abb. 4. Beispielergebnis zur Hautfarbsegmentierung: links das Originalbild aus der Baumarkt-Umgebung, rechts das Segmentierungsergebnis mit einem groben, universellen Hautfarbmodell

**Gesichtsdetektion** Der zunächst favorisierte Ansatz zur Gesichtsdetektion verwendete Eigengesichter (siehe [12]) als Faltungsmasken und klassifizierte in dem entstehenden Merkmalsraum, erwies sich jedoch hinsichtlich Berechnungskomplexität und Geschwindigkeit für den Echtzeit-Einsatz auf Standard-Hardware als ungeeignet. Das Verfahren wurde deshalb dahingehend modifiziert, daß ausschließlich ein Mittelwertgesicht als Faltungsmaske benutzt wird, wobei in beiden genannten Varianten eine Normierung des Bildausschnitts und eine Elimination

des Mittelwertes erfolgte. Die aktuell implementierte Methode der *normierten mittelwertfreien Skalarprodukte* (NMS) hat neben dem enormen Effektivitätsgewinn den Vorteil, daß das erhaltene Faltungsergebnis direkt als Ähnlichkeitsmaß benutzt werden kann. Abb. 5 stellt anhand eines Beispielbildes die beiden genannten Methoden gegenüber: (v.l.n.r.) Originalbild, Segmentierungsergebnis mittels Eigengesichtern und anschließender Fitwertklassifikation und Segmentierungsergebnis mit dem NMS-Verfahren. Das NMS-Verfahren liefert mehr falsch positive Detektionsergebnisse, die jedoch durch die Fusion der verschiedenen Cues wieder eliminiert werden können und sich somit nicht negativ auf das Gesamtlokalisationsergebnis auswirken.

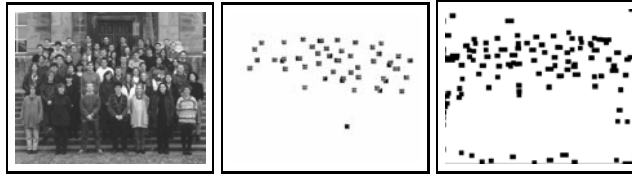


Abb. 5. v.l.n.r. Originalbild, Gesichtsdetektion mittels Eigengesichtern und Ergebnis mittels NMS-Verfahren

**Detektion einer Kopf-Schulter-Silhouette** Der Konturdetektionsalgorithmus lokalisiert die Kopf-Schulter-Silhouette einer annähernd frontal zur Kamera ausgerichteten Person. Abb. 6 illustriert die wesentlichen Verarbeitungsschritte. Die Grundidee des Verfahrens beruht auf der Approximation eines Kontur-

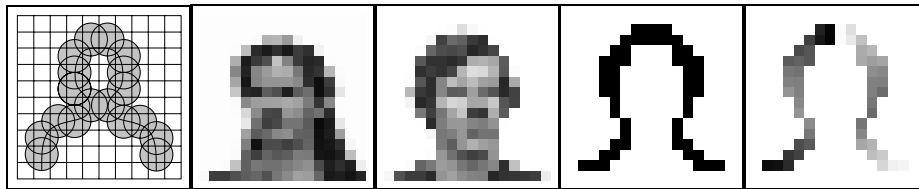


Abb. 6. v.l.n.r.: Arrangement orientierungssensitiver Filter; zwei Beispielbilder ( $16 \times 16$ ) des Trainingsdatensatzes; daraus ermittelter binärer Konturverlauf; eigentliches Konturmodell (Prototyp) als Arrangement der dominanten Orientierungen entlang des binären Konturverlaufs, wobei die Orientierungswinkel als Grauwerte kodiert wurden ( $0^\circ$ : schwarz;  $90^\circ$ : mittelgrau;  $180^\circ$ : weiß); das Konturmodell ist symmetrisch, da der Datensatz entlang der y-Achse gespiegelt wurde

verlaufs durch ein Arrangement orientierter Filter, welches auf alle Auflösungs-ebenen angewendet wird. Um dieses Arrangement möglichst allgemeingültig zu bestimmen, wird zunächst mittels eines Satzes von Beispielbildern ein mittlerer binärer Konturverlauf gelernt, woran sich die Ermittlung der dominanten Orientierung an jedem Konturpunkt nach [8] anschließt. Das so erhaltene Konturmodell wird jetzt mit dem Orientierungsverlauf verglichen, der an der aktuellen Bildposition ermittelt wurde. Eine detaillierte Beschreibung des Verfahrens findet sich in [3].

**Fusion der Cues und Lokalisation der Person** Die Fusion der Cues und die sich daran anschließende Selektion der Region, die mit größter Wahrscheinlichkeit eine Person enthält, soll durch Abb. 7 illustriert werden. Alle Cues liefern



**Abb. 7.** Lokalisierungsergebnis in einer Szene aus der realen Baumarkt-Umgebung: oben links das Originalbild und rechts die Regionen, für die nach der Cue-Fusion eine starke Hypothese für das Vorhandensein einer Person vorliegt; deutlich sichtbar sind neben den eigentlichen Personenregionen weitere Hypothesen, die sich nahezu zwangsläufig aufgrund der Szenenstruktur ergeben; unten v.l.n.r. exemplarisch die Ausgaben der Cue-Module für Hautfarbe, Silhouette und Gesichtsstruktur für die erste Ebene (höchste Auflösung) der Multiskalenrepräsentation

parallel ihren Beitrag in die Pyramide der Auffälligkeitskarten (siehe Abb. 3), die aus nichtlinearen dynamischen Neuronen aufgebaut ist, welche sowohl innerhalb einer Ebene als auch zwischen unmittelbar benachbarten Ebenen miteinander wechselwirken. Ein WTA-ähnlicher 3D-Selektionsprozess ([1]) ermittelt jetzt die Regionen, in denen die Cues die stärksten korrespondierenden Beiträge liefern. Vorteilhaft daran ist die implizite Hystereseeigenschaft des Selektionsprozesses infolge der nichtlinearen Dynamik, weshalb damit quasi bereits ein Tracking einer einmal selektierten Region erfolgt. Für eine detaillierte Beschreibung des Verfahrens sei wiederum auf [3] verwiesen.

Die Personenlokalisierung ist zweifelsfrei eine wesentliche Verhaltensleistung, die das Gesamtsystem erbringen muß. Durch die parallele Integration verschiedener visueller Cues ist ein robustes dynamisches Auffälligkeitssystem realisiert worden. In einem Entfernungsbereich von 0.5 bis etwa 2.5 Meter, den die 5 Auflösungsebenen bei Verwendung einer CHUGAI BOYEKI CD 08 Videokamera mit maximaler Weitwinkelseinstellung abdecken, ist mit dem vorgestellten System eine sehr sichere Personenlokalisierung möglich.

**Integration weiterer Cues** Zur Beschleunigung der Personenlokalisierung wurden zwei weitere Methoden entwickelt. Zum einen wird das Bild der Omnikame-

ra, welches den gesamten Umgebungsbereich der PERSES-Plattform als 360°-Aufnahme erfasst, hinsichtlich Bewegungsinformation segmentiert. Zu diesem Zweck wurde ein Segmentierungsverfahren entwickelt, das eine kontinuierliche Adaption der Modelle für Szenenhintergrund und Person realisiert und Ähnlichkeiten zu dem von WREN in [15] vorgeschlagenen Ansatz aufweist. Damit ist zwar keine exakte Segmentierung der Person möglich, jedoch läßt sich auf diese Weise sehr effizient und einfach die Richtung bestimmen, in der sich möglicherweise eine Person befindet (siehe Abb. 8). Nachdem das Doppelkamera-Active-Vision-System in diese Richtung geschwenkt wurde, schließt sich dann die oben beschriebene Personenlokalisierung an, da eine Lokalisation direkt im Bild der Omnikamera aufgrund der geringen Detailauflösung unmöglich ist. Durch die Kopplung beider Kamerasysteme kann der gesamte Umgebungsbereich der Plattform kontinuierlich beobachtet werden.



**Abb. 8.** Beispiel für die Bewegungsdetektion im Bild der Omnikamera; links: Originalbild der Einsatzumgebung; rechts: bewegungsbasiert segmentierte Regionen, in denen sich eine Person befindet bzw. befinden könnte

Das zweite Verfahren verwendet eine Stereo-Mikrofonanordnung zur Geräuschlokalisierung [11]). Auch hier besteht das Ziel darin, die Aufmerksamkeit der Plattform in die entsprechende Richtung zu lenken und diesen Bereich der Umgebung nach einer möglicherweise vorhandenen Person, die mit der Plattform in Interaktion treten will, abzusuchen.

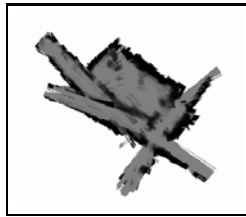
## 2.2 Personentracking

Die Personenlokalisierung kann als der Initialisierungsschritt für das anschließende Personentracking verstanden werden. Da der visuelle Kontakt zum Benutzer auch dann gewährleistet sein muß, wenn sich sowohl Kamera/Plattform als auch die Person bewegen, scheiden Tracking-Verfahren, die eine stationäre Kamera voraussetzen, von vornherein aus. Neben der Ausnutzung der impliziten Trackingeigenschaft der nichtlinearen dynamischen Cue-Fusion während der Personenlokalisierung zielen unsere aktuellen Arbeiten wiederum auf die Verwendung eines Multi-Cue-Ansatzes ab, der im wesentlichen Bewegungs- und Farbinformationen im Bild der Omnikamera miteinander kombiniert (siehe auch [10]). Über ein adaptives (da ansichtenabhängiges) Farbmodell der zu trackenden Person soll weiterhin garantiert werden, daß eine sichere Detektion des aktuellen Nutzers auch

dann gesichert wird, wenn zwischenzeitlich bspw. aufgrund von Verdeckung der visuelle Kontakt unterbrochen war. Da sich die Arbeiten zum Personentracking erst am Anfang befinden, können hier noch keine aussagefähigen Ergebnisse präsentiert werden. Ein aktuell am Fachgebiet laufendes Forschungsprojekt (siehe Fußnote im Titel) widmet sich vorrangig dieser Fragestellung.

### 3 Lokale und globale Roboternavigation

Neben scan- bzw. ultraschall-basierten Verfahren wurde für die **lokale Hindernisvermeidung** ebenfalls ein monokularer MULTI-CUE-ANSATZ entwickelt und zur AMS'98 ([9]) vorgestellt. Eine aktuell laufende Arbeit untersucht, inwieweit mittels Omnikamera durch Auswertung von Symmetrien in einem spärlichen Flußfeld eine robuste lokale Navigation möglich ist.



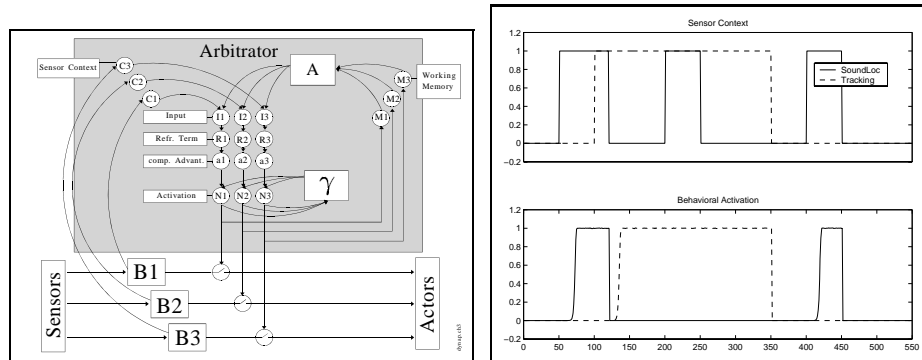
**Abb. 9.** Beispiel für eine in realer Umgebung erlernte Umgebungskarte (MILVA-Plattform mit Laserscanner, schwarz: Hindernis, grau: Freiraum); deutlich erkennbar ist der durch den Odometriefehler verursachte Versatz des Flurbereiches (links oben), nachdem ein geschlossener Kurs durchfahren wurde

Für die **globale Roboternavigation** wird der von THRUN ([14]) und Mitarbeitern entwickelte probabilistische Occupancy-Grid-Ansatz derzeit reimplementiert, der sich während der Tour-Guide-Projekte ([4]), die eine Reihe von Gemeinsamkeiten mit unserem Szenario aufweisen, bereits bewährt hat. Abb. 9 zeigt als erstes vorläufiges Ergebnis eine Umgebungskarte, die von der MILVA-Plattform gelernt wurde. Die Verfahren zur Selbstlokalisierung und der damit verbundenen Korrektur des Odometriefehlers werden zum aktuellen Zeitpunkt implementiert und können noch nicht mit aussagefähigen Ergebnissen hinterlegt werden.

### 4 Steuerarchitektur zur Verhaltensorganisation

Die Notwendigkeit einer effizienten Steuerarchitektur ergibt sich aus der Aufgabe, die verschiedenen „Verhaltensmodule“ zu einem sinnvollen Gesamtverhalten des Service-Systems zu verknüpfen. Aus der Vielzahl der bekannten verhaltensbasierten Ansätze erscheint uns der in ([13]) vorgeschlagene „Dynamic Approach“ am geeignetsten, dessen Architektur in Abb. 10 links zu sehen ist. Das Kernstück dieser Architektur bildet der Arbitrator, der auf einem nichtlinearen DGL-System basiert, welches die Interaktion der verschiedenen Verhaltensmodule modelliert und dessen Attraktoren gewünschte Teilverhalten darstellen. Anhand der in Abb. 10 rechts dargestellten Zeitverläufe soll die Funktionsweise dieses Systems zumindest grob veranschaulicht werden: Zwei Verhalten, das Tracken einer Person (B1) und das Reagieren auf ein akustisches Ereignis (Geräuschlokalisierung, B2), wurden in dem Beispiel realisiert. Im Initialzustand sind beide Verhaltensmodule nicht aktiviert (*Behavioral Activation*  $N_1 = 0$  und  $N_2 = 0$ ).

Beide Module teilen über ihren *Sensor Context* dem Arbitrator mit, inwieweit sie in der aktuellen sensorischen Situation in der Lage sind, ein hinreichend gutes Berechnungsergebnis zu liefern. Nachdem eine Schallquelle lokalisiert wurde, erfolgt ein Schwenken der Kameras in die entsprechende Richtung, und nach erfolgreicher visueller Personenlokalisierung (B3, hier nicht näher betrachtet) beginnt der Trackingprozeß, was eine Inaktivierung der akustischen Aufmerksamkeit nach sich zieht und sich darin äußert, daß  $N2$  vom Zustand 1 in den Zustand 0 übergeht, da bei sicherem Tracking keine „Ablenkung“ durch ein akustisches Ereignis erfolgen soll. Damit wirkt sich ein in dieser Phase auftretendes Geräusch nicht auf das Gesamtverhalten der Plattform aus, da das entsprechende Verhaltensmodul nicht auf die Aktorik zugreifen darf.



**Abb. 10. Links:** Schematische Darstellung der Architektur zur Verhaltenskoordination (adaptiert aus [13]); **Rechts:** Zeitverläufe der Variablen Sensor Context und Verhaltensaktivierung für das im Text erläuterte Beispiel zur Koordination von visueller und akustischer Personenlokalisierung bzw. -tracking

Schlägt das Tracking hingegen fehl ( $N1$  geht aufgrund des fehlenden Sensor Context vom Zustand 1 in den Zustand 0 über), wird die akustische Aufmerksamkeit wieder aktiviert (bzw. nicht mehr vom Trackingmodul inhibiert) und kann zur Unterstützung der Nutzersuche herangezogen werden. Die zur Verhaltenskoordination notwendigen Randbedingungen, die bspw. die gleichzeitige Aktivierung verschiedener Verhaltensmodule verbieten oder definierte zeitliche Randbedingungen (Verhalten X muß vor der Aktivierung von Verhalten Y aktiv sein) festlegen, werden über entsprechende Einträge in die Systemmatrizen  $A$  bzw.  $F$  realisiert. Die vorgestellte Methode gestattet eine sehr elegante Verhaltensbeschreibung und läßt sich problemlos um neue Verhaltensmodule erweitern. Weiterhin können für komplexe Teilverhalten mehrere separate Architekturen mit prinzipiell identischem Aufbau kombiniert werden.

## 5 Roboterartikulation – Sprach-Ein-Ausgabe

Für eine effiziente Interaktion zwischen Benutzer und System ist es wesentlich, daß das System seinen aktuellen „Zustand“ seinem Benutzer mitteilen kann. Zu diesem Zweck sind eine Reihe vordefinierter sprachlicher Äußerungen, wie



bspw. „Ich fahre jetzt zu Ihnen“ oder „Bitte folgen Sie mir“, vorgesehen. Die Ausgabe der dem aktuellen Situations- und Interpretationskontext adäquaten Äußerungen wird von der Steuerarchitektur getriggert, da diese letztendlich entscheidet, welche Teilverhalten in der aktuellen Situation aktiv bzw. inaktiv sind. Ein einfaches, derzeit noch sprecherabhängiges Spracherkennungsmodul zur Erkennung von Kommandoworten wird eingesetzt, um bspw. eine Bestätigung für die vom Roboter beabsichtigte Aktion zu erhalten („Ich würde jetzt zu Ihnen fahren. Bitte bestätigen Sie mit Ja.“).

## 6 Szenariospezifische Service-Angebote

Neben der Kontaktaufnahme und dem kontinuierlichen Kontakthalten zum Benutzer sind eine Reihe szenariospezifischer Teilleistungen auf dem Service-System zu integrieren. Zunächst ist vorgesehen, daß das System lediglich als „Lotse“ fungiert und den Kunden in einen von diesem (per Touchscreen) ausgewählten Marktbereich oder Artikelstandort führt. Zu diesem Zweck wird die Karte der Umgebung, die während der Exploration im Rahmen der globalen Navigation erstellt wurde, entsprechend attribuiert. Perspektivisch läßt sich das Service-Angebot bspw. um die Funktionen „Anforderung eines Fachberaters“, „Präsentation aktueller Sonderangebote“ etc. nahezu unendlich erweitern.

## 7 Zusammenfassung und Ausblick

Das hier vorgestellte Projekt besitzt zum aktuellen Zeitpunkt noch vorrangig Konzeptcharakter und soll im Verlauf der nächsten 3 bis 5 Jahre Stück für Stück umgesetzt werden. Die dargestellten Ergebnisse verdeutlichen den aktuellen Realisierungsstand. Erste umfangreiche Vor-Ort-Untersuchungen in der realen Baumarkt-Umgebung werden ab Oktober diesen Jahres durchgeführt und sicherlich zur Folge haben, daß einige der bereits entwickelten Teilsysteme angepasst oder vollständig neu konzipiert werden müssen. Als unmittelbare Erweiterungen steht die multimodale und damit robustere Hindernisvermeidung an, die durch eine Integration der Ultraschall-Entfernungsdaten und der visuellen Hindernisdetektion nach [9] erreicht werden und sich an das in [6] vorgestellte Verfahren anlehnen soll. Weiterhin ist vorgesehen, die Steuerarchitektur zur Verhaltenskoordination um solche Methoden zu erweitern, die bspw. die Adaption der Systemmatrizen über Reinforcement-Lernprozesse ermöglichen. Die Frage nach der langfristigen Tragfähigkeit der vorgestellten Steuerarchitektur zur Verhaltenskoordination kann zum gegenwärtigen Zeitpunkt noch nicht abschließend beantwortet werden. Gegebenenfalls sind hier Alternativvarianten notwendig, die sich stärker an regelbasierten Ansätzen orientieren.

## Danksagung

Die Autoren danken allen Kollegen und Studenten des Fachgebietes Neuroinformatik für ihre engagierte und konstruktive Mitarbeit.

## Literatur

1. Amari, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.
2. Boehme, H.-J., Braumann, U.-D., Corradini, A., and Gross, H.-M. Person Localization & Posture Recognition for Human-Robot Interaction. In *GW'99 - The 3rd Gesture Workshop, Gif-sur-Yvette, France*. Springer, 1999. in press.
3. Braumann, U.-D., Corradini, A., Böhme, H.-J., and Gross, H.-M. Konturbasierte Personenlokalisierung mittels dreidimensionaler neuronaler Felder und steuerbarer Filter. In *Mustererkennung 1998, 20. DAGM-Symposium*, Informatik aktuell, pages 385–392. Springer Verlag, 1998.
4. Burgard, W., Cremers, A., Fox, D., Lakemeyer, G., Hähnel, D., Schulz, D., Steiner, W., and Thrun, S. The interactive museum tour-guide robot. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, 1998.
5. Darrell, T., Gordon, G., Harville, M., and Woodfill, J. Integrated person tracking using stereo, color, and pattern detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 601–609, 1998.
6. Gaudiano, P., Chang, C., Ecemis, I., Martens, S., Sahin, E., Streilein, W., and Wagner, R. Neurobotics Lab Research: Learning, Vision and Sonar Recognition with Mobile Robots. In *NN'99, 4. International Workshop on „Neural Networks in Applications, Magdeburg*, pages 91–98, 1999.
7. Hunke, M.H. Locating and Tracking of Human Faces with Neural Networks. Technical report, Carnegie Mellon University Pittsburgh, 1994. CMU-CS-94-155.
8. Jähne, B. *Praktikal Handbook on Image Processing for Scientific Applications*. CRC Press LLC, 1997.
9. Krabbes, M., Weber, S., Boehme, H.-J., and Gross, H.-M. Monokulare visuelle Hindernisdetektion auf Basis merkmalsbasierter Bildsegmentierung. In *AMS'98 - 14. Fachgespräch Autonome Mobile Systeme*, Informatik aktuell, pages 85–92. Springer-Verlag, 1998.
10. Munkelt, O., Ridder, C., Hansel, D., and Hafner, W. A model driven 3D image interpretation system applied to person detection in video images. In *ICPR'98: International Conference on on Pattern Recognition*, 1998.
11. Paschke, P. and Schauer, C. A spike-based model of binaural sound localization. In *EWNS'99 - 2nd European Workshop on Neuromorphic Systems, Stirling*, 1999. in press.
12. Rowley, H. A., Baluja, S., and Kanade, T. Human Face Detection in Visual Scenes. Technical report, Carnegie Mellon University, 1995. CMU-CS-95-158R.
13. Steinhage, A. and Bergener, T. Dynamical Systems for the Behavioral Organization of an Abthropomorphic Mobile Robot. In *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior (SAB 98)*. M.I.T. Press, 1998.
14. Thrun, S. Learning Maps for Indoor Mobile Robot Navigation. *Artificial Intelligence*, 99(1):21–71, 1999.
15. Wren, C., Azarbayejani, A. and Darrell, T., and Pentland, A. Pfunder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997. M.I.T. Media Lab Techreport TR 353.
16. Yang, J., Lu, W., and Waibel, A. Skin-Color Modeling and Adaptation. Technical report, Carnegie Mellon University, 1997. CMU-CS-97-146.