Proc. Conference Robotics 2002, pp 575-580, Ludwigsburg, Germany, VDI-Berichte 1697, VDI Verlag 2002

Wechselseitige Unterstützung von sonarbasiertem und visuellem Personentracking *

Mutual Support of Sonar-based and Visual Person Tracking

T. Wilhelm, H.-J. Böhme und H.-M. Groß

Technische Universität Ilmenau, Institut für Neuroinformatik
89684 Ilmenau
{wilhelm, hans, homi}@informatik.tu-ilmenau.de

Zusammenfassung

Für Serviceroboter, die mit Menschen interagieren müssen, stellt das Auffinden und kontinuierliche Verfolgen einer Person eine wesentliche Grundfertigkeit dar. Es wird ein Verfahren vorgestellt, welches visuelle Informationen und Sonardaten kombiniert, um Hypothesen über die Position eines Interaktionspartners über die Zeit zu verfolgen.

Servicerobots intended to interact with people must be able to localize and continuously track their users. A method is described which integrates information from visual and sonar-based tracking methods while updating hypotheses about the position of the robots human user. Each tracking method uses information from the other to generate a more robust measure of the user's position and thus a more robust behaviour generation is achieved.

1 Einleitung

Der Einsatz von Servicerobotern hat in den letzten Jahren wesentlich an Bedeutung gewonnen. Obwohl diese Forschungsrichtung noch in den Kinderschuhen steckt und Anwendungen bis jetzt hauptsächlich demonstrativen Charakter haben, konnte der Einsatz von einigen Modellen auch schon in realistischen Szenarien gezeigt werden [2]. Entscheidende Bedeutung beim Entwurf eines Serviceroboters, der im Umfeld von menschlichen Nutzern arbeiten und Aufgaben für diese erledigen soll, kommt der Entwicklung einer dem Aufgabengebiet des Roboters entsprechenden Mensch-Maschine-Schnittstelle zu. In dem von den Autoren verfolgten Szenario übernimmt der Roboter die Aufgabe eines Einkaufsassistenten in einem Baumarkt. Dieser soll in der Lage sein, mit einem potentiellen Nutzer Kontakt aufzunehmen, dessen Einkaufswünsche zu erfragen und den Kunden zu den entsprechenden Artikeln zu lotsen. Neben den Problemen der Navigation, Selbstlokalisation und Pfadplanung in einer stark unstrukturierten Umgebung ergibt sich aus dem Einsatzszenario die Notwendigkeit für intuitive und

*gefördert durch die Projekte PERSES und SERROKON des Thüringer Ministeriums für Wissenschaft, Forschung und Kunst

natürliche Kommunikationsmöglichkeiten mit Nutzern. Das Hauptaugenmerk dieses Beitrags liegt dabei auf der Fähigkeit des Roboters, die Position eines Kunden in geeigneter Weise zu verfolgen.

2 Tracking

Es liegt auf der Hand, warum ein Serviceroboter, der direkten Umgang mit Menschen pflegen soll, in der Lage sein muß, seinen Interaktionspartner kontinuierlich zu verfolgen. Wendet sich der Nutzer während der Kommunikation ab, macht es für den Roboter keinen Sinn, weiter Ausgaben zu produzieren. Ebenso stupide würde das sture Weiterfahren des Roboters wirken, wenn sich der Nutzer während einer Lotsenfahrt in eine andere Richtung bewegt oder zurückbleibt. Um diese Aufgabe zu lösen, stehen einem Roboter verschiedene Sensorsysteme zur Verfügung. Die Entfernung zu Objekten läßt sich beispielsweise über Laser- oder Sonarsensoren ermitteln. Diese Systeme liefern allerdings nur begrenzt Hinweise über die Natur des Objektes. Obwohl aus der Literatur Verfahren bekannt sind, die aus dem Profil eines Laser-Scans Hypothesen für menschliche Nutzer erzeugen [5], kann doch nicht garantiert werden, daß es sich tatsächlich um Menschen handelt. Die Autoren verfolgen aus diesem Grund zusätzlich zur Sonarmessung einen visuellen Ansatz zum Auffinden und Verfolgen von Menschen, da dieser sichere Hypothesen über das Vorhandensein eines Menschen in der Szene liefert.

2.1 Sonarbasiertes Tracking

Bei dem verwendeten Roboter handelt es sich um einen B21 der Firma RWI IS Robotics. Dieser ist mit zwei Ebenen von Sonarsensoren ausgestattet, die die Entfernung zu einem Objekt bis maximal $22.5m$ bestimmen. Die Meßwerte dieser Sensoren sind starken Schwankungen unterworfen, die sowohl vom Material des Objektes, dem Winkel, unter dem der Ultraschall auf das Objekt auftrifft und der Objektentfernung abhängen. Sichere Messungen kann man bis maximal $2m$ erwarten. Aus diesem Grund werden die Rohdaten der Sensoren auf folgende Weise vorverarbeitet:

1 Eliminieren von fehlerhaften Messungen: Entfernungen größer $22,5m$ sind ungültig und werden durch den vorhergehenden Wert ersetzt
2 räumliche Glättung: Minimum aus drei benachbarten Entfernungswerten
3 zeitliche Glättung: Tiefpaßfilter
4 Berechnung der Bewertung: $W^{(c)}_{Sonar} = 1 - d^{(c)}_{sonar}/d_{max}$, wobei $d^{(c)}_{sonar}$ der bis dahin vorverarbeitete Sonarmeßwert an der Position c im Scan und d_{max} die maximal zu berücksichtigende Entfernung $(1,5m)$ ist; Werte größer d_{max} werden mit 0 gewichtet

In dem resultierenden Vektor wird das Maximum gesucht, also das dem Roboter nächstgelegene Objekt. Aus dessen Position kann ein in der speziellen Situation gewünschtes Verhalten generiert werden. Der Roboter verfügt hierfür über verschiedene Operationsmodi:

1 *Kommunikation:* versuche das Eingabegerät (Touchscreen) immer am Nutzer zu halten
2 *Lotsenfahrt:* halte Mindestabstand zum Nutzer und diesen immer hinter dem Roboter

3 *Folgefahrt:* halte Mindestabstand zum Nutzer und diesen immer vor dem Roboter

Der wesentliche Vorteil des sonarbasierten Trackings ist seine geringe Berechnungskomplexität, die ein schritthaltendes Verfolgen eines Nutzers erlaubt. Es erzeugt aber nur das korrekte Verhalten, solange der Nutzer auch tatsächlich das nächste Objekt ist. Ansonsten wendet sich der Roboter jedem beliebigen Objekt in seinem Umfeld zu. Um diesem Umstand Rechnung zu tragen, wird ein visuelles Tracking-Verfahren integriert.

2.2 Visuelles Tracking

Grundlage des visuellen Tracking-Verfahrens bildet der Condensation-Algorithmus [6]. Die eigentliche Funktion des Tracking-Verfahrens, nämlich für jeden Bildpunkt die Wahrscheinlichkeit des Vorhandenseins einer Person zu ermitteln und die entstehende Dichtefunktion über die Zeit zu verfolgen, wird hierbei mit einer relativ kleinen Anzahl von Samples approximiert. Der Vorteil von Condensation liegt also in einer drastischen Reduzierung des Rechenaufwandes, da die Wahrscheinlichkeit nicht für jedes Pixel, sondern nur für die n Stützstellen, an denen sich Samples befinden, berechnet werden muß. Des weiteren ist das Verfahren in der Lage, multimodale Verteilungen zu tracken, bzw. mit mehreren Hypothesen über die aktuelle Position des Nutzers zu arbeiten. Das visuelle Tracking arbeitet auf einem ins kartesische Koordinatensystem transformierten Bilderstrom einer omnidirektionalen Kamera und deckt somit den gesamten Bereich um den Roboter ab. Im folgenden werden die Merkmale, die für die Berechnung der Sample-Wahrscheinlichkeiten herangezogen werden und deren Verknüpfung beschrieben.

Farbe Ein gebräuchlicher Ansatz zum Auffinden von Gesichtern ist die Suche nach Hautfarbe. Um eine weitgehende Beleuchtungsinvarianz zu erreichen, wird im dichromatischen r-g-Farbraum gearbeitet ($r = R/(R+G+B)$, $g = G/(R+G+B)$). Für die Bildung des Modells werden Bildbereiche manuell als Hautfarbe klassifiziert und in den r-g-Farbraum transformiert. Da nicht gewährleistet werden kann, daß das entstehende Histogramm hinreichend dicht besetzt ist (Abbildung 1 links), wird die Punktewolke mit einer unimodalen bivariaten Gaußverteilung approximiert [7] (Abbildung 1 rechts).

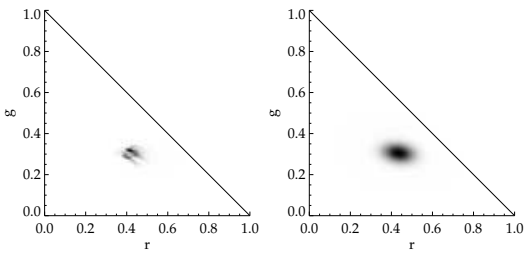

Abbildung 1: Links: Punktewolke der als Hautfarbe klassifizierten Bildregionen im r-g-Farbraum. Rechts: Bivariater Gauß-Klassifikator. Die Kovarianzmatrix wurde mit dem Faktor 2 skaliert, um Fluktuationen bei variabler Beleuchtung besser auszugleichen.

Kopf-Schulter-Kontur Einen zweiten Hinweis für das Vorhandensein eines Nutzers bietet dessen Kopf-Schulter-Partie (KSP) [1]. Dazu wurde aus Frontalansichten von Personen ein Modell bestimmt, das den Konturverlauf einer mittleren Kopf-Schulter-Partie beschreibt (Abbildung 2). Es wird nun zunächst mit einem Strukturtensor [4] die Orientierung in einer lokalen Umgebung jedes Samples bestimmt und für jedes Sample ein Template-Matching mit dem

KSP-Modell durchgeführt. Das beschriebene Verfahren wird auf einer Multiskalenrepräsentation mit drei Ebenen und Halboktavenabstand angewendet, so daß Kopf-Schulter-Konturen in einem Entfernungsbereich von $0,5 - 1,5m$ detektiert werden können.

Abbildung 2: Links: Zwei Beispiele aus dem Datensatz, der zur Bestimmung des Konturmodells verwendet wurde. Rechts: Das Konturmodell einer mittleren Kopf-Schulter-Partie. Die Grauwerte kodieren lokale Orientierungen zwischen $0°$ und $180°$.

Verknüpfung von KSP und Farbe Obwohl beide Merkmale sehr personenspezifisch sind, kann es vorkommen, daß sie Nutzer nicht detektieren bzw. dort Nutzer detektieren, wo keine sind. Das heißt, es wird eine geeignete Verknüpfungsmöglichkeit benötigt, die zwischen pessimistisch und optimistisch parametrisiert werden kann. Aus diesem Grund wurde der MinMax-Operator ($minmax(a,b) = \gamma min(a,b) + (1-\gamma)max(a,b)$) verwendet, dessen Verhalten durch den Parameter γ kontinuierlich zwischen den Extrema festgelegt werden kann. Maximum ($\gamma = 0$) bedeutet, daß alle falsch-positiven Detektionen voll in den Condensation-Algorithmus eingehen, während das Minimum ($\gamma = 1$) dazu führt, daß ein von einem der beiden Cues nicht erkannter Nutzer gänzlich unberücksichtigt bleibt. Untersuchungen zur geeigneten Wahl des Parameters sind Gegenstand aktueller Arbeiten. Abbildung 3 zeigt Ergebnisse der beiden Cues und deren Verknüpfung.

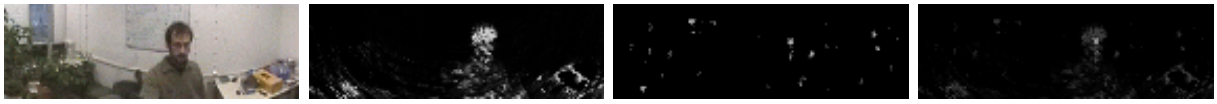

Abbildung 3: Darstellung einzelner Cues des visuellen Tracking v.l.n.r.: Originalbild; Ergebnis der Farbdetektion; Ergebnis der KSP-Detektion; $MinMax$ Verknüpfung ($\gamma = 0.3$). Nur an der Position des Nutzers liefern beide Cues einen Beitrag.

3 Gegenseitige Unterstützung der Trackingverfahren

Es liegt nun nahe, die beiden Verfahren, sonarbasiertes und visuelles Tracking, zu kombinieren, um so die Stärken beider Verfahren auszunutzen. Zum einen kann das visuelle Tracking den Sonar-Scan verwenden und in die Berechnung der Sample-Wahrscheinlichkeit einbeziehen. Zum anderen soll das visuelle Tracking vermeiden, daß die Verhaltensgenerierung des sonarbasierten Trackings unbelebte Objekte ansteuert.

3.1 Unterstützung des visuellen Trackings durch das Sonartracking

Da sowohl das Kamerabild als auch der Sonar-Scan eine $360°$-Beschreibung der Umgebung des Roboters repräsentieren, ist es möglich, jeder Position $\mathbf{x}$ im Bild eine korrespondierende Position c im Scan zuzuordnen (Abbildung 4). Die Bewertung des Sonar-Trackings geht modulierend in die Berechnung der Bewertung für jedes Sample i ein (Gleichung 1). Jedem Sample wird nach Gleichung 2 eine Wahrscheinlichkeit zugewiesen.

$$W^{(i)}_{Sample}(\mathbf{x}) = minmax\left(W^{(i)}_{Farbe}(\mathbf{x}), W^{(i)}_{KSP}(\mathbf{x})\right) W_{Sonar}(c) \quad (1)$$

$$P^{(i)}_{Sample}(\mathbf{x}) = \frac{W^{(i)}_{Sample}(\mathbf{x})}{\sum_i W^{(i)}_{Sample}(\mathbf{x})} \quad (2)$$

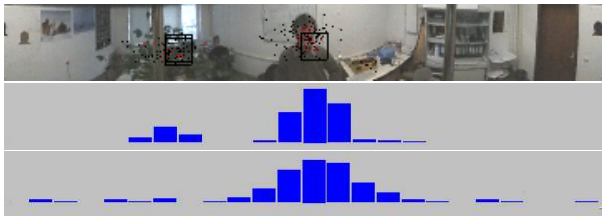

Abbildung 4: Oben: Entzerrtes Bild der Omnikamera mit eingezeichneten Condensation-Samples und Nutzer-Hypothesen. Mitte: Summe der Sample-Wahrscheinlichkeiten für jeweils ein Bildspalte C. Unten: Vektor der vorverarbeiteten Sonarmessungen.

Die Einbeziehung der Sonar-Messungen bewirkt, daß eine Richtung, in der tatsächlich ein Nutzer anzutreffen ist, durch eine korrespondierende hohe Bewertung im Sonar-Scan begünstigt wird, während Samples, die nicht durch eine Sonarmessung Bestätigung erfahren, gänzlich aussterben. Mit dieser Modifikation wird das visuelle Tracking wesentlich robuster, da entfernte Objekte, die zwar eine hohe Bewertung aus den visuellen Modulen zugesprochen bekommen, nicht vom Sonartracking unterstützt werden und somit keine Samples anziehen können (siehe Abbildung 5).

Abbildung 5: Gegenüberstellung von rein visuellem Tracking (links) und sonar-unterstütztem Tracking (rechts). Es wird nur jedes zehnte Bild der Sequenz dargestellt (die Zeitachse verläuft von oben nach unten); der Nutzer bewegt sich ein mal um den Roboter. Der Roboter dreht sich sonarbasiert mit. Während beim ausschließlich visuellen Tracking viele Samples an der Tür hängenbleiben (hautfarbänhnlich), verliert das Verfahren mit Sonar-Unterstützung den Nutzer nicht.

3.2 Unterstützung des Sonartrackings durch das visuelle Tracking

Tatsächlich ist der umgekehrte Fall, also die Unterstützung des sonarbasierten Trackings durch das visuelle Tracking, der interessantere, da nur das Sonartracking direkten Einfluß auf die Verhaltensgenerierung hat. Dabei wird das Kamerabild in C Spalten unterteilt, die mit je einer Sonarmessung korrespondieren. Nun wird in jeder Spalte c die Summe der Sample-Wahrscheinlichkeiten berechnet (Abbildung 4 Mitte). Das Resultat ist ein Vektor mit hohen Werten an den Stellen, an denen wahrscheinlich eine Person zu finden ist. Für die Verhaltensgenerierung wird die Position des Maximums im visuellen und im sonarbasierten Scan

verglichen. Stimmen diese überein, kann das Sonartracking wie beschrieben die Steuerkommandos generieren, andernfalls wird jede Aktion unterdrückt. Auf diese Weise können weitere Nutzer an den Roboter herantreten, ohne daß Gefahr besteht, daß sich der Roboter von seinem aktuellen Interaktionspartner ab- und einer anderen Person zuwendet.

4 Zusammenfassung und Ausblick

Das vorgestellte Verfahren ist in der Lage, die Position eines Interaktionspartners über die Zeit zu verfolgen. Dabei resultiert die Kombination von sonarbasiertem und visuellem Tracking in einer deutlichen Steigerung der Robustheit gegenüber den Einzelverfahren. Weitere Arbeiten zielen auf den Einsatz während der Fahrt. Im Moment scheitert dies am visuellen Tracking, welches in stark strukturierten Umgebungen relativ leicht vom Nutzer auf andere Personen oder Objekte umspringt.
Z.B. soll ein alternatives Farbmodell untersucht werden, welches aus einer Matrix besteht, in der Hautfarbpixel entsprechend der Häufigkeit ihres Auftretens in einer Datenbank eingetragen und mit einer Dilatations-Erosions-Operation nachbearbeitet werden [3]. Dieses Vorgehen verspricht durch die genauere Approximation der Datenverteilung eine Reduzierung der falschpositiven Klassifikationen. Um ein Verwechseln mit anderen Personen auszuschließen, sollte weiterhin ein Modell des aktuellen Nutzers erstellt und für das visuelle Tracking verwendet werden.

Literatur

[1] Braumann, U.-D. *Multi-Cue-Ansatz für ein dynamisches Auffälligkeitssystem zur visuellen Personenlokalisation*. PhD thesis, Technische Universität Ilmenau, 2001.

[2] 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.

[3] Feyrer, S. *Detektion, Lokalisierung und Verfolgung von Personen mit einem mobilen Serviceroboter*. PhD thesis, Eberhard-Karls-Universität Tübingen, 2000.

[4] B. Jähne. *Practical Handbook on Image Processing for Scientific Applications*. CRC Press LLC, 1997.

[5] Schulz, D. and Burgard, W. Probabilistic state estimation of dynamic objects with a moving mobile robot. *Robotics and Autonomous Systems*, 34:107–115, 2001.

[6] Schulz, D., Burgard, W., Fox, D., and Cremers, A.B. Tracking Multiple Moving Targets with a Mobile Robot using Particle Filters and Statistical Data Association. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2001.

[7] Yang, J. and Waibel, A. A Real-Time Face Tracker. In *Proceedings of WACV'96*, 1996.