

Ralf Schug, Matthias Zobel, Joachim Denzler, Heinrich Niemann
Sichtbasierte Personenescortierung mittels einer autonomen mobilen Plattform

erschienen in:

Robotik 2000: Leistungsstand – Anwendungen – Visionen – Trends

VDI-Berichte 1552

Berlin, Germany

2000

S. 459–464

VDI Verlag Düsseldorf

Sichtbasierte Personeneskortierung mittels einer autonomen mobilen Plattform

Ralf Schug, Matthias Zobel*, Joachim Denzler, Heinrich Niemann

Universität Erlangen–Nürnberg

Lehrstuhl für Mustererkennung (Informatik 5)

Martensstr. 3, D–91058 Erlangen

Tel.: +49–9131–85-27775, FAX: +49–9131–303811

Zusammenfassung

In diesem Artikel wird ein System vorgestellt, das einer autonomen mobilen Plattform die Lokalisierung und Eskortierung einer Person mit annähernd konstantem Abstand erlaubt. Dabei wird ausschließlich visuelle Information verarbeitet. Personen werden durch die Kontur des Kopf- und Schulterbereichs modelliert. Nach Einführung eines Maßes für die Bewertung derartiger Konturen können Personen auf relativ einfache Weise lokalisiert werden. Durch die zusätzliche Verwendung von Hautfarbeninformation und einem lokalen Optimieren der Konturhypothesen wird die Lokalisierung verbessert. Aufbauend auf den Lokalisierungsergebnissen erfolgt die dynamische Personenverfolgung mittels des so genannten CONDENSATION-Algorithmus. Abhängig von der geschätzten 3-D Personenposition wird schließlich die Plattformbewegung gesteuert.

Abstract

In this article, we describe a system that enables an autonomous mobile platform to localize and escort a person while keeping approximately a constant distance. In the system, only visual information is used. We model a person by the contour of its head and shoulders. After the introduction of a criterion for measuring the goodness of a contour, persons can be localized in a relatively easy way. By incorporating information about skin color and by a local optimization step, the localization can be further improved. Based on the localization results, the task of dynamic person tracking is done by means of the so called CONDENSATION algorithm. Dependent on the resulting estimation of the 3-D position of the person, the mobile platform is steered to keep track of the moving person.

1 Einleitung

Nachdem in den letzten 15 Jahren Industrieroboter immer stärker an Bedeutung für das produzierende Gewerbe gewonnen haben, ist eine ähnliche Entwicklung auch für den Dienstleistungssektor zu erwarten [5]. So genannte Serviceroboter übernehmen dabei Aufgaben, die vorher nicht durch Maschinen erledigt werden konnten. Beispiele für solche Aufgabengebiete

*Kontaktautor

Diese Arbeit wurde unterstützt durch die Deutsche Forschungsgemeinschaft im Rahmen des SFB 603, Teilprojekt B2.

sind z. B. das selbstständige Reinigen von Böden oder das Führen von Besuchern durch eine Ausstellung. Darüber hinaus wird an der Entwicklung von persönlichen Servicerobotern gearbeitet, die kranke oder behinderte Menschen bei der Lösung alltäglicher Probleme unterstützen sollen [1].

Alle diese genannten Tätigkeiten von Servicerobotern sind dadurch gekennzeichnet, dass sie in engem Kontakt mit Menschen ausgeführt werden. Dazu ist es notwendig, dass der Roboter weiß, wo sich Personen befinden, und dass er auf Änderungen der Personenposition z. B. durch Anhalten oder Ausweichen reagiert. In diesem Artikel wird ein System vorgestellt, mit dessen Hilfe eine Person lokalisiert und durch eine mobile Plattform eskortiert werden kann. Die Steuerung der Plattform erfolgt in einem geschlossenen Regelkreis aus Sensorik und Aktorik, wobei nur rein visuelle Information verwendet wird. Einzig eine rudimentäre Hindernisvermeidung bedient sich der vorhandenen Infrarotsensoren, um beispielsweise das Fahren gegen Wände auszuschließen.

2 Konturrepräsentation

Modell Als charakteristisches Merkmal für eine Person dient der Umriss der Kopf- und Schulterpartie, da dieser bei Bewegungen relativ invariant gegenüber Formänderungen ist. Für die interne Darstellung wird diese Kontur durch einen kubischen B-Spline mit dem Kontrollpunktvektor \mathbf{q} modelliert. Zur Anpassung eines B-Splines an eine im Bild enthaltene Personenkontur könnte der Kontrollpunktvektor direkt manipuliert werden. Auf diese Weise entstehen jedoch sehr leicht Konturen, die keinerlei Ähnlichkeit mehr mit der Kontur einer Person aufweisen. Daher ist es sinnvoll, die erlaubten geometrischen Transformationen durch Verwendung eines so genannten *Shape Spaces* einzuschränken [2]. Ein Shape Space besteht aus einer Basis­kurve, die durch einen Kontrollpunktvektor \mathbf{q}_0 gegeben ist sowie aus einer Transformationsmatrix $\mathbf{W}(\mathbf{q}_0)$, welche die an der Basis­kurve erlaubten Transformationen angibt. Die Elemente eines Shape Spaces werden als *Konfigurationen* bezeichnet und erlauben die kompakte Beschreibung der an der Basis­kurve vorgenommenen geometrischen Transformationen, wie z. B. Translation, Rotation und Skalierung. Mit Hilfe der Transformationsmatrix $\mathbf{W}(\mathbf{q}_0)$ und dem Kontrollpunktvektor \mathbf{q}_0 der Basis­kurve kann eine Konfiguration durch eine lineare Transformation wieder auf den Kontrollpunktvektor \mathbf{q} eines B-Splines abgebildet werden.

Bewertung Zur Bewertung der Güte einer Kontur benötigt man ein Maß, das die Übereinstimmung der zu bewertenden Kontur mit der tatsächlichen Personenkontur widerspiegelt. Hierfür wird eine so genannte *Konturdiskriminante* verwendet, durch die eine Klassifizierung einer Konturhypothese in die beiden Klassen „gesuchte Kontur“ bzw. „nicht die gesuchte Kontur“ ermöglicht wird [4]. Als Merkmale werden dabei Grauwertänderungen über einem bestimmten Schwellwert θ betrachtet, die mit Hilfe eines eindimensionalen Merkmaldetektors entlang von m Normalen zur Kontur bestimmt werden (Abbildung 1, links). Die Normalen fester Länge /



Abbildung 1: (links) Gewinnung der Merkmale zur Berechnung der Konturdiskriminante, (rechts) zufälliges Erzeugen der Konturhypothesen.

werden als *Messlinien* bezeichnet, das Ergebnis der Merkmaldetektion auf der i -ten Messlinie ist die Beobachtung o_i .

Sei A_v das Ereignis, dass der Umriss der Person den Abstand v auf der betrachteten Messlinie besitzt, und \bar{A} das Ereignis, dass die Messlinie die Kontur überhaupt nicht schneidet. Ferner sei B das Ereignis, dass die Personenkontur auf Grund eines guten Kontrasts zum Hintergrund beobachtet werden kann, und \bar{B} das Ereignis, dass dies nicht der Fall ist. Für eine Kontur r und die zugehörige Beobachtung $\mathbf{o} = (o_1, \dots, o_m)^T$ ergibt sich dann die Konturdiskriminante

$$d(\mathbf{r}) = \prod_{i=1}^m \left(\frac{q p(o_i | A_0, \bar{B}) + (1 - q) p(o_i | A_0, B)}{p(o_i | \bar{A})} \right), \quad (1)$$

wobei die Konstante q die für alle Messlinien als gleich angenommene Wahrscheinlichkeit für die Nichtdetektion von Merkmalen der wahren Kontur darstellt.

3 Eskortierung

Initiale Lokalisierung Für die Lokalisierung von Personen werden mittels eines speziellen Shape Spaces zufällige Konturhypothesen generiert und in die zu untersuchende Aufnahme projiziert (Abbildung 1, rechts). Die a priori Wahrscheinlichkeiten für die einzelnen Formparameter, z. B. für die Rotation oder Skalierung, können aus einer Stichprobe geschätzt werden. Um jedoch den Parameterraum für die möglichen Positionen von Konturhypothesen einzuschränken wird die Aufnahme mittels Farbschwellwerten grob nach Hautfarbenregionen segmentiert. Hypothesen werden anschließend nur noch innerhalb dieser Regionen zugelassen. Zur weiteren Verbesserung der Lokalisierung werden die Konturhypothesen nach ihrer Projektion ins Bild lokal an die nächsten Merkmale angepasst [2]. Die anschließende Bewertung der einzelnen Hypothesen durch die Berechnung der Konturdiskriminante kann schließlich für die Initialisierung der Eskortierung verwendet werden.

Verfolgung Zur Schätzung der Position und der Lage einer Person im Bild über die Zeit hinweg wird der CONDENSATION-Algorithmus eingesetzt, bei dem zu jedem Zeitpunkt t die a po-

steriori Dichte $p(\mathbf{x}_t | \mathbf{o}_t, \dots, \mathbf{o}_1)$ über alle Zustände \mathbf{x} durch eine endliche Menge von Samples approxiiert wird [3]. Dazu wird in jedem Zeitschritt die a posteriori Dichte unter Verwendung eines probabilistischen Bewegungsmodells fortgeschaltet und durch ein ebenfalls probabilistisches Beobachtungsmodell anhand von Beobachtungen in einer Aufnahme bewertet (Partikel Filter). Ein Zustand \mathbf{x} besteht aus zwei durch Shape Space Konfigurationen beschriebenen Konturen, nämlich der momentanen und derjenigen des letzten Zeitschritts. Dies ist nötig, um im Bewegungsmodell Änderungen der Formparameter berücksichtigen zu können. Für die Positionskomponente eines Zustandes wird z. B. angenommen, dass sich die Person mit der gleichen Geschwindigkeit im Bild wie im vorherigen Schritt fortbewegt. In ähnlicher Weise geht man davon aus, dass sich die Skalierungsfaktoren in x - und y -Richtung wie im Schritt zuvor verändern. In der Praxis werden kleine Dreh- und Scherwinkel als konstant angesehen. Um auch unvorhersehbare Bewegungsabläufe handhaben zu können, werden die Konfigurationskomponenten zusätzlich mit einem Gauss'schem Rauschen additiv überlagert. Als Beobachtungsmodell kann die in Abschnitt 2 vorgestellte Konturdiskriminante verwendet werden. Experimente zeigen jedoch, dass diese für eine schnelle Verfolgung zu aufwendig zu berechnen ist. Daher wird als Beobachtungsmodell die stark vereinfachte Version der Konturdiskriminante

$$d'(r) = \prod_{i=1}^m \left(q\lambda + \sum_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{v_{i,j}^2}{2\sigma^2}\right) \right) \quad (2)$$

eingesetzt, die jedoch das Auftreten einer gelernten Anzahl von Merkmalen im Inneren der tatsächlichen Kontur nicht berücksichtigt [4]. Die Variable n_i stellt dabei die Anzahl der auf Messlinie i gefundenen Merkmale dar, während $v_{i,j}$ den Abstand des j -ten Merkmals von der Kontur auf dieser Messlinie angibt. Die Konstante λ ist der Poisson-Parameter für die Verteilung von Merkmalen im Hintergrund und die Konstante σ gibt die Streuung des Fehlers bei der Merkmaldetektion an. Die konkrete Schätzung für die Position und die Lage der Person im Bild muss aus der a posteriori Verteilung bestimmt werden, z. B. kann man dazu den Erwartungswert verwenden.

Plattformsteuerung Um den Abstand zwischen der Person und der mobilen Plattform konstant halten zu können, muss dieser geschätzt werden. Dazu wird vereinfachend angenommen, dass alle Köpfe die gleiche Breite besitzen. Ist die Brennweite der Kamera und die Breite eines Bildpunkts in der Bildebene bekannt, kann der Abstand in Abhängigkeit vom Skalierungsfaktor in x -Richtung näherungsweise bestimmt werden. Entsprechend der Abweichung vom gewünschten Abstand wird anschließend die Vorwärts- bzw. Rückwärtsbewegung gesteuert. Befindet sich die geschätzte Position der Person nicht in der Mitte der Aufnahme, so wird versucht, dies durch entsprechende Drehbewegungen der Plattform zu korrigieren. Da sich bei Bewegung des mobilen Systems auch die starr montierte Kamera mitbewegt, muss bei der Fortschaltung der Zustände die Eigenbewegung kompensiert werden. Um die Eskortierung durch das mobile System möglichst autonom zu gestalten, verfügt das System über eine Aufmerksamkeitskomponente, durch die das Scheitern der Eskortierung selbstständig erkannt



Abbildung 2: (links) Die für die Experimente verwendete mobile Plattform MOBSY, (rechts) Beispiel für eine Eskortierung; der zeitlichen Abstand zwischen zwei dargestellten Aufnahmen beträgt 5 Sekunden.

wird. Zur Vermeidung einer Kollision des Roboters mit möglichen Hindernissen wird mittels Infrarotsensoren die Entfernung zu den nächsten Objekten bestimmt. Liegt der Abstand unter einer festgelegten Schwelle, so werden alle Bewegungen der Plattform sofort eingestellt.

4 Experimente

Die Experimente zur Bewertung der Personenlokalisierung und -eskortierung wurden mit dem mobilen System „MOBSY“ (Abbildung 2, links) durchgeführt. Der Roboter verfügt über einen Pentium Pro Rechner mit 200 MHz Taktfrequenz zur Steuerung des Antriebs und zur Abfrage der Sensoren sowie über einen Dual Pentium II Rechner mit 300 MHz Taktfrequenz für Anwendungen. Bei jeweils 1000 für die Lokalisierung generierten Konturhypothesen betrug der Euklid'sche Abstand zwischen dem Schwerpunkt der am besten bewerteten Hypothese und dem Schwerpunkt der tatsächlichen Personenkontur in einem 384×288 Bild durchschnittlich 6,6 Bildpunkte, d. h. die Position der Person im Bild konnte auf etwa 7 Bildpunkte genau bestimmt werden. Durch die Hautfarbensegmentierung und die lokale Anpassung der Hypothesen an die nächsten Merkmale betrug die Rechenzeit für die initiale Lokalisierung ca. 4 Sekunden. Während der nachfolgenden Eskortierung ergab sich bei langsamen Personenbewegungen eine mittlere Abweichung von etwa 10 Bildpunkten zwischen der geschätzten Personenkontur und der tatsächlichen Position. Bei 100 fortgeschalteten Partikeln pro Zeitschritt im CONDENSATION-Algorithmus konnten hierbei 9 Bilder der Größe 384×288 pro Sekunde verarbeitet werden. Die in Abbildung 2 rechts dargestellte Bildfolge zeigt ein Beispiel für eine erfolgreiche Eskortierung, die Schätzung der Personenkontur ist zur Veranschaulichung weiß eingezeichnet. Insgesamt zeigte sich bei den Experimenten, dass durch das beschriebene System Personen zuverlässig lokalisiert und bei genügend langsamen Personenbewegungen eskortiert werden können.

5 Zusammenfassung und Ausblick

Das vorgestellte System erlaubt die Lokalisierung und anschließende Eskortierung einer Person durch ein mobiles System. Die Lokalisierung erfolgt durch die Generierung zufälliger Konturhypothesen, die mit Hilfe der Konturdiskriminante bewertet werden. Mittels der Segmentierung von Hautfarbenregionen und durch eine lokalen Anpassung der Konturhypothesen wird die Lokalisierung verbessert. Die nachfolgende Schätzung der Position und der Lage über die Zeit hinweg geschieht durch den Einsatz des CONDENSATION-Algorithmus mit entsprechenden Bewegungs- und Beobachtungsmodellen. Die Steuerung der Plattformbewegung erfolgt abhängig von der Schätzung der Personenkontur und der Entfernung der Plattform zur Person. Die dadurch bedingte Eigenbewegung der Kamera wird bei der Fortschaltung der Zustände kompensiert. Um die Autonomie des Systems zu steigern, erkennt eine Aufmerksamkeitskomponente selbstständig das Scheitern der Eskortierung z. B. auf Grund zu schneller Bewegungen der Person. Außerdem existiert eine rudimentäre Hindernisvermeidung, die bei zu nahen Objekten die Bewegungen der Plattform einstellt.

Zur Erweiterung des bestehenden Eskortierungssystems kann die zweite auf der mobilen Plattform zur Verfügung stehende Kamera verwendet werden, um mittels Stereosehen den Abstand zur Person genauer zu schätzen. Weiterhin wäre ein automatisches Umfahren von Hindernissen wünschenswert. Zur Erhöhung der Zuverlässigkeit der Eskortierung ist die Untersuchung von weiteren Bewegungs- und Beobachtungsmodellen erforderlich.

Literatur

- [1] U. Ahlrichs, J. Fischer, J. Denzler, C. Drexler, H. Niemann, E. Nöth, D. Paulus: *Knowledge Based Image and Speech Analysis for Service Robots*, in *Proceedings of the International Conference on Computer Vision '99*, Workshop on Integration of Speech and Image Understanding, 1999.
- [2] A. Blake, M. Isard: *Active Contours*, Springer Verlag, Berlin, Heidelberg, New York, 1998.
- [3] M. Isard, A. Blake: *CONDENSATION – conditional density propagation for visual tracking*, *International Journal of Computer Vision*, Bd. 29, Nr. 1, 1998, S. 5–28.
- [4] J. MacCormick, A. Blake: *A probabilistic contour discriminant for object localisation*, in *Proceedings of the 6th International Conference on Computer Vision*, 1998, S. 390–395.
- [5] R. Schraft, H. Volz: *Serviceroboter: Innovative Technik in Dienstleistung und Versorgung*, Springer Verlag, Berlin, Heidelberg, 1996.