

# Navigation mit Schnappschüssen\*

Matthias O. Franz<sup>1</sup>, Bernhard Schölkopf<sup>1</sup>, Hanspeter A. Mallot<sup>1</sup>, Heinrich H. Bülthoff<sup>1</sup> und Andreas Zell<sup>2</sup>

<sup>1</sup> Max-Planck-Institut für biologische Kybernetik, Spemannstr. 38, D-72076 Tübingen

<sup>2</sup> Wilhelm-Schickard-Institut für Informatik, Lehrstuhl Rechnerarchitektur, Köstlinstr. 6, D-72074 Tübingen

**Zusammenfassung** Es wird ein biologisch inspirierter Algorithmus vorgestellt, mit dem sich ein Ort wiederfinden läßt, an dem vorher eine 360°-Ansicht der Umgebung aufgenommen wurde. Die Zielrichtung wird aus der Verschiebung der Bildposition der umgebenden Landmarken im Vergleich zum Schnappschuß berechnet. Die Konvergenzeigenschaften des Algorithmus werden mathematisch untersucht und auf mobilen Robotern getestet.

## 1 Einleitung

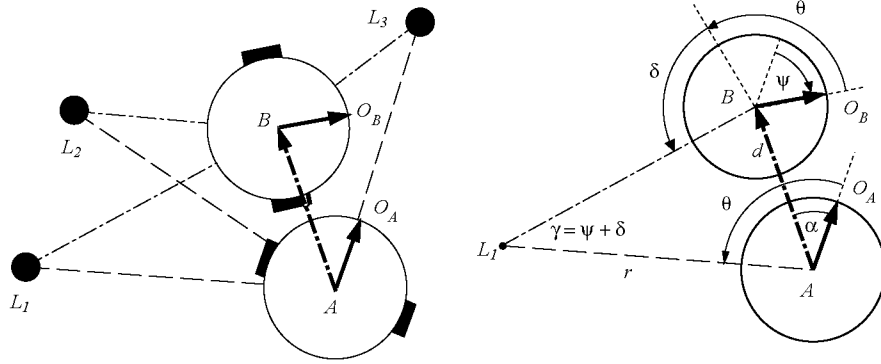
Für viele Navigationsaufgaben ist die Verwendung externer Informationen unumgänglich, da bestimmte Vorgänge - wie z.B. Drift beim Flug oder Schlupf der Antriebsräder beim Fahren - nicht immer intern erkannt werden können. Ein möglicher Ansatz ist die Identifizierung von Orten durch lokale Ansichten. Dabei sind zwei Fälle zu unterscheiden: Zum einen kann die lokale Ansicht den Ort selbst abbilden, zum anderen kann die Ansicht die umgebende Szene vom Ort aus gesehen darstellen. Die erste Möglichkeit bietet sich an, wenn der Ort durch auffällige visuelle Merkmale markiert ist, die man z.B. zur Zielfahrt verwenden kann. Ist dies nicht der Fall, dann muß die Lage des Ortes aus seiner Relation zu den umgebenden Landmarken bestimmt werden (Abb. 1 links).

Eine Vielzahl von Experimenten aus der Verhaltensforschung hat gezeigt, daß bereits Wirbellose wie z.B. Bienen oder Ameisen diese Fähigkeit besitzen (s. Übersicht in Collett, 1992). Die experimentellen Befunde legen nahe, daß diese Tiere ihr Ziel an Orten suchen, an denen der momentane Anblick am besten mit einem dort vorher gespeicherten *Schnappschuß* der Umgebung übereinstimmt. Cartwright & Collett (1983) zeigten anhand eines einfachen Computermodells, daß die beobachteten Suchmuster sich erklären lassen, wenn man annimmt, daß ein Schnappschuß die einzig verfügbare Information ist. Die Richtung des Ziels wurde dabei aus der Veränderung des Sehwinkels und der Größe der umgebenden Landmarken berechnet. Die spätere Implementierung ähnlicher Mechanismen auf mobilen Robotern zeigte, daß sich Schnappschüsse auch für die Roboternavigation eignen (Hong et al., 1992; Röfer, 1997; Franz et al., 1997).

Im folgenden wird ein robuster Algorithmus beschrieben, mit dem sich ein durch einen Schnappschuß der Umgebung markierter Ort anfahren läßt. Dazu

---

\* Gefördert durch das Human Frontier Science Program (HFSP).



**Abbildung 1.** Links: Ist das Ziel nicht durch auffällige visuelle Merkmale definiert, muß die Fahrtrichtung aus der Veränderung der Sehwinkel der umgebenden Landmarken berechnet werden. Rechts: Bewegt sich der Sensorring vom Ausgangspunkt A nach B, so verändert sich der Sehwinkel der Landmarke  $L_1$  um  $\gamma$ . Eine zusätzliche Orientierungsänderung um  $\psi$  führt zu einer Gesamtverschiebung des Bildes von  $L_1$  um  $\delta = \gamma - \psi$ .

werden zunächst die mathematischen Grundlagen der Aufgabenstellung dargestellt, bevor auf die Details des Algorithmus eingegangen wird. Danach beschreiben wir die Implementierung auf mobilen Robotern und schließen mit einer Diskussion der Ergebnisse.

## 2 Mathematische Beschreibung des Problems

Als vereinfachtes Modell eines mobilen Agenten nehmen wir einen omnidirektionalen Sensorring an, dessen Bewegungen auf die Ebene beschränkt sind. Mit Hilfe des Sensors kann der Agent lokale  $360^\circ$ -Ansichten aufnehmen, die die Grauwerteverteilung entlang des Horizontes wiedergeben. Die Position eines Bildpunktes auf dem Sensorring sei durch den Winkel  $\theta$  beschrieben. In diesem Zusammenhang bezeichnen wir alle im Bild identifizierbaren Umgebungspunkte als Landmarken. Dies stimmt nicht immer mit der üblichen Vorstellung einer Landmarke als einem physikalischen Objekt überein, d.h. ein Objekt kann durchaus mehrere Landmarken enthalten.

Der Sensorring bewege sich nun von seiner Ausgangsposition A in Richtung  $\alpha$  zu dem in Entfernung  $d$  liegenden Punkt B und ändere seine Orientierung um den Winkel  $\psi$  (s. Abb. 1 rechts). Durch die Bewegung wandert das an Position  $\theta$  sichtbare Bild der in Entfernung  $r$  liegenden Landmarke  $L_1$  zu einer neuen Position  $\theta + \delta$ . Aus dem Dreieck  $AL_1B$  ergibt sich mit Hilfe des Sinussatzes die Beziehung

$$\frac{r}{d} = \frac{\sin(\pi - \theta + \alpha - \psi - \delta)}{\sin(\psi + \delta)} = \frac{\sin(\theta - \alpha + \psi + \delta)}{\sin(\psi + \delta)}. \quad (1)$$

Sind  $r/d$  und  $\psi$  bekannt, so läßt sich die Bewegungsrichtung  $\alpha$  und damit die

Richtung  $\beta = \alpha - \psi + \pi$  zurück zum Ausgangspunkt aus der Bildverschiebung  $\delta$  berechnen.

Um die Verschiebung  $\delta$  zum Wiederfinden eines durch einen Schnappschuß markierten Ortes verwenden zu können, müssen zwei grundsätzliche Probleme gelöst werden: 1. Dem von uns betrachteten visuell navigierenden Agenten stehen lokale Ansichten als einzige Information zur Verfügung. Der unbekannte Parameter  $r/d$  muß daher über geeignete Annahmen zu typischen Distanzverteilungen in der Umwelt des Agenten geschätzt werden. 2. Für die Berechnung von  $\delta$  müssen Korrespondenzen zwischen dem Schnappschuß und der aktuellen Ansicht gefunden werden.

In dem hier vorgestellten Ansatz wird angenommen, daß alle Landmarken in ungefähr gleicher Entfernung vom Ausgangspunkt liegen. Auf den ersten Blick erscheint dies nicht sehr realistisch, dennoch beeinflussen die resultierenden Fehler die tatsächliche Navigationsfähigkeit des Agenten nur wenig (s. Abschn. 3.2). Darüberhinaus führt diese Annahme zu einem einfachen globalen Verfahren zum Auffinden von Korrespondenzen.

### 3 Berechnung der Zielrichtung

#### 3.1 Parametrisierte Verschiebungsfelder

Unter der Annahme einer typischen Landmarkendistanz  $R$  läßt sich Gl. (1) in

$$\tan(\psi + \delta) = \frac{\frac{d}{R} \sin(\theta - \alpha)}{1 + \frac{r-R}{R} - \frac{d}{R} \cos(\theta - \alpha)} = \frac{\rho \sin(\theta - \alpha)}{1 + \epsilon - \rho \cos(\theta - \alpha)}. \quad (2)$$

überführen, wobei zur Erleichterung der Notation  $\rho = d/R$  und  $\epsilon = (r - R)/R$  gesetzt wird. Die Fälle  $\rho = 0$ ,  $\epsilon = -1$  und  $\rho = 1 + \epsilon$ ,  $\theta = \alpha$  müssen dabei ausgeschlossen werden. Praktisch bedeutet dies, daß der Agent nicht dieselbe Position wie eine Landmarke einnehmen darf.

Wie oben erwähnt, gehen wir von der vereinfachenden Annahme aus, daß alle Landmarken gleich weit vom Ausgangspunkt in Entfernung  $R$  liegen, d.h. wir setzen in Gl. (2)  $\epsilon = 0$ . Die durch

$$\tan(\psi + \delta) = \frac{\rho \sin(\theta - \alpha)}{1 - \rho \cos(\theta - \alpha)} \quad (3)$$

beschriebenen Verschiebungsfelder  $\delta(\theta)|_{\epsilon=0}$  werden durch diese Annahme stark vereinfacht und hängen nur noch von den drei freien Parametern  $\alpha$ ,  $\psi$  und  $\rho$  ab. Die auf diese Weise parametrisierten Verschiebungsfelder können durch Variation von  $\alpha$ ,  $\psi$  und  $\rho$  an das real vorhandene Verschiebungsfeld angepaßt werden. Da die Bewegungsrichtung  $\alpha$  einer der angepaßten Parameter ist, kann der Ausgangspunkt mit Hilfe des folgenden Algorithmus wiedergefunden werden:

1. Für alle Kombinationen von  $\alpha$ ,  $\psi$  und  $\rho$  wird ein Satz von Verschiebungsfeldern nach Gl. (3) berechnet. Mit Hilfe dieser Verschiebungsfelder werden die Bilder der Landmarken im momentanen Anblick verschoben und dadurch

ein Satz neuer Bilder generiert. Wäre die Annahme der gleichen Landmarkendistanz exakt erfüllt, dann würden diese Bilder genau dann entstehen, wenn sich der Agent tatsächlich nach den Parametern bewegt hätte.

2. Der Satz der generierten Bilder wird mit dem am Ausgangspunkt gespeicherten Schnappschuß verglichen, um den Anblick zu finden, der den Schnappschuß am besten rekonstruiert. Als Maß für die Qualität der Rekonstruktion wird das Skalarprodukt zwischen Schnappschuß und Ansicht verwendet.
3. Der Parameter  $\alpha$ , der zur besten Rekonstruktion des Schnappschusses führt, wird zur Schätzung der Fahrtrichtung  $\beta = \alpha + \pi$  zurück zum Ausgangspunkt herangezogen.
4. Der Agent bewegt sich einen Schritt weit in Richtung  $\beta$  und wiederholt den Vorgang, bis die Ausgangsposition wieder erreicht ist. Dies ist der Fall, wenn entweder das Skalarprodukt zwischen dem momentanen Anblick und Schnappschuß eine vordefinierte Schwelle überschreitet oder sich die berechnete Zielrichtung  $\beta$  nach Passage des Ausgangspunktes um  $180^\circ$  dreht.

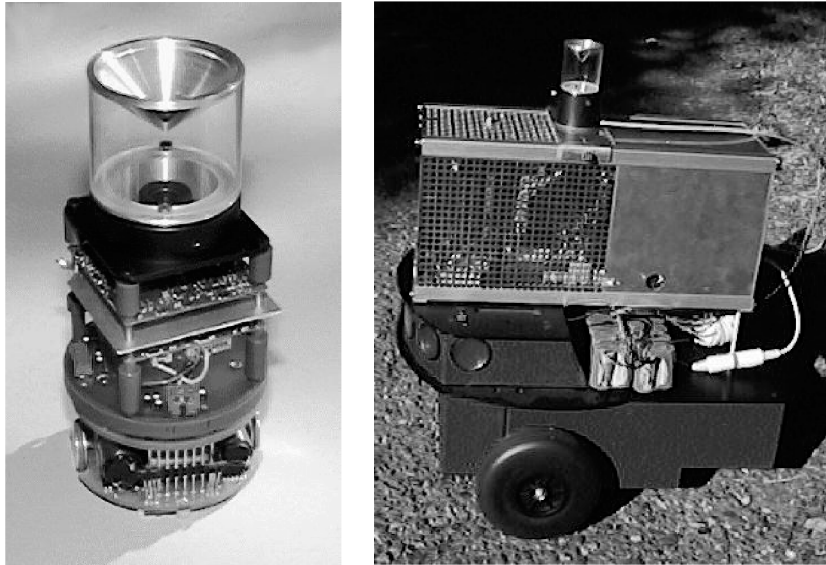
Damit die drei Unbekannten in Gl. (3) bestimmt werden können, müssen mindestens drei Landmarken sichtbar sein. Ist dies nicht der Fall, so kann die Fahrtrichtung nur unter Hinzunahme weiterer Informationen aus Kompassen oder Odometrie berechnet werden.

### 3.2 Fehleranalyse und Konvergenz

Wie eingangs erwähnt basiert der vorgestellte Algorithmus auf der Annahme einer gleichen Landmarkendistanz. Obwohl diese Annahme im Normalfall kaum erfüllt sein wird, behindert der resultierende Fehler das Wiederfinden des Ausgangspunktes nur in geringem Maße. Dies liegt v.a. daran, daß der Fehler um so kleiner wird, je näher der Agent seinem Ziel kommt. Dies folgt direkt aus Gl. (2), wenn wir annehmen, daß der Agent bereits den offenen Bereich um das Ziel erreicht hat (d.h.  $\rho < 1 + \epsilon$ ). In diesem Fall ist die Verschiebung  $\delta$  durch

$$\delta = \arctan \left( \frac{\rho \sin(\theta - \alpha)}{1 + \epsilon - \rho \cos(\theta - \alpha)} \right) - \psi \quad (4)$$

gegeben. Der Fehler durch die Vernachlässigung der individuellen Abweichungen der Landmarkendistanzen von  $R$  ist dann  $E(\epsilon, \rho) := \delta(\epsilon, \rho) - \delta(0, \rho)$ . Sowohl  $\delta$  als auch  $E$  sind stetige Funktionen in  $\rho$  und  $\epsilon$  für  $\epsilon > -1$ . Beide gehen asymptotisch gegen Null für  $\rho \rightarrow 0$ . Daraus folgt, daß für jedes  $\epsilon > -1$ , gegebenes  $\theta$  und  $\alpha$  und eine beliebige Fehlerschranke  $E_0 > 0$  ein  $\rho_0$  existiert, für das  $|E(\epsilon, \rho)| < E_0$  aus  $\rho < \rho_0$  folgt. In anderen Worten bedeutet dies, daß sich das Ziel beliebig genau anfahren läßt, auch wenn die umgebenden Landmarken unterschiedliche Entfernungen vom Ziel haben - vorausgesetzt, der Agent ist hinreichend nahe am Ziel. Jeder Schnappschuß ist also von einem Gebiet umgeben, in dem sich der Ort des Schnappschusses beliebig genau bestimmen läßt. Das gesamte Gebiet, von dem aus sich das Ziel erreichen läßt, wird *Fangbereich* des Schnappschusses genannt. In der praktischen Anwendung ist die theoretisch unbegrenzte Genauigkeit natürlich nicht erreichbar, da hier das Sensorrauschen die wahrnehmbaren



**Abbildung 2.** Links: Khepera mit omnidirektionalem Sensor. Rechts: RWI Pioneer-Plattform mit Industrie-PC zur Bildverarbeitung.

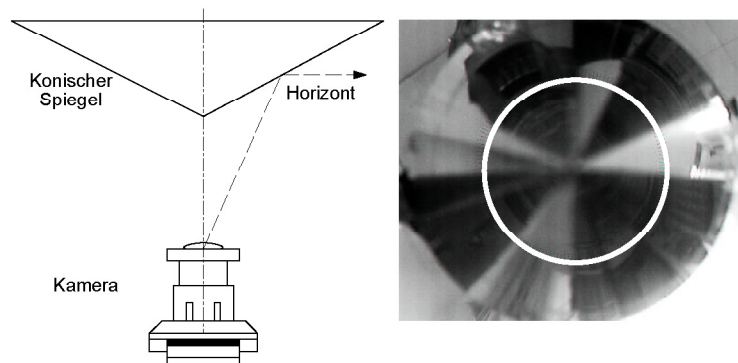
Unterschiede zwischen Schnappschuß und dem momentanen Anblick begrenzt (Franz et al., 1997).

## 4 Experimente mit mobilen Robotern

### 4.1 Experimenteller Aufbau

Die hier gezeigten Daten wurden mit Hilfe eines modifizierten Khepera-Miniaturroboters gewonnen, der über ein Kabel mit einer Workstation verbunden war (Abb. 2 links). Als Landmarken dienten acht Modellhäuser in einer  $118 \times 102$  cm großen Arena. Um den Algorithmus in realitätsnäheren Umgebungen zu testen, wurde ein komplett autonomer Prototyp für Büroumgebungen entwickelt. Hierzu wurde ein Industrie-PC mit Framegrabber-Karte als Host auf eine käufliche RWI Pioneer-Basis montiert (Abb. 2 rechts). Das Kamerasystem und die Kontrollalgorithmen waren auf beiden Plattformen identisch.

Das bildgebende System beider Roboter besteht aus einer Miniatur-Videokamera, die auf einen polierten Metallkegel gerichtet ist (Abb. 3, vgl. Chahl & Srinivasan, 1996). Das ringförmige  $360^\circ$ -Sichtfeld dieser Anordnung erstreckt sich von  $10^\circ$  unter- bis  $10^\circ$  oberhalb des Horizontes. Die Videobilder werden mit 25 Hz in vier Ringen entlang des Horizontes mit einer Winkelauflösung von  $4.6^\circ$  ausgelesen und radial gemittelt, um eventuelle Verkippungen des Roboters auszugleichen. Die aus den so gewonnenen Bildern berechneten Bewegungskommandos werden anschließend seriell zu der Roboterplattform übertragen (12 Hz Khepera, 10 Hz Pioneer).



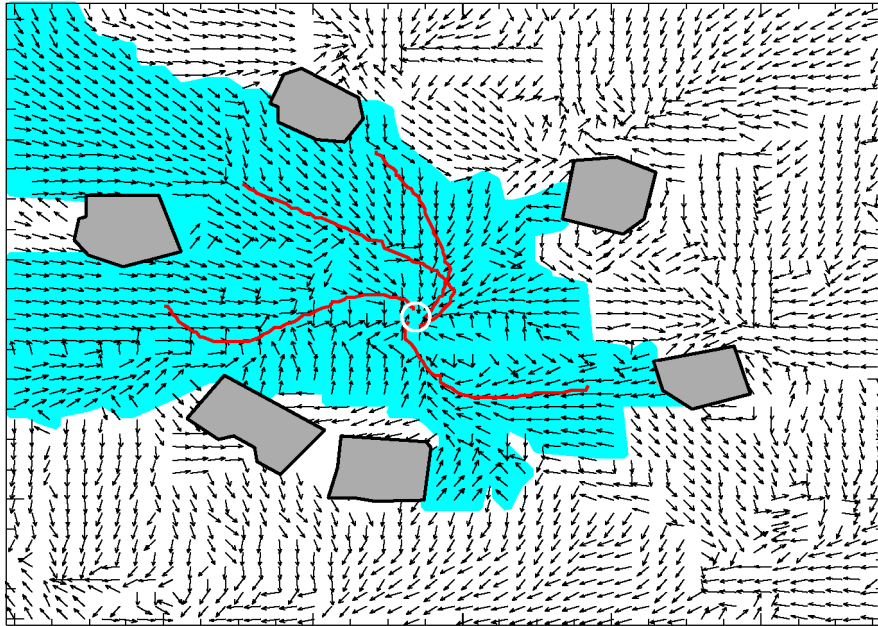
**Abbildung 3.** Die omnidirektionalen Ansichten werden mit einer Kamera aufgenommen, die auf einen polierten Kegel gerichtet ist. Rechts ist der Spiegel von der Kamera aus gesehen. Der weiße Ring markiert den Auslesebereich entlang des Horizontes.

Die Position des Khepera in der Arena wurde mit einer kalibrierten Farbkamera verfolgt. Sowohl Bild- als auch Positionsdaten wurden mit Zeitmarkern aufgenommen und später synchronisiert. Während der Experimente standen beiden Robotern keine absoluten Positionsdaten zur Verfügung.

## 4.2 Ergebnisse

Um die Anwendbarkeit des Algorithmus zu testen, fuhr der Khepera alle zugänglichen Positionen innerhalb der Arena an und nahm dabei 10000 lokale Ansichten auf, die die gesamte Arena relativ gleichmäßig abdecken. Die Größe des Fangbereichs eines Schnappschusses kann mit Hilfe dieser Daten auf folgende Weise dargestellt werden: Für einen gegebenen Schnappschuß wird mit Hilfe des Algorithmus aus jeder lokalen Ansicht ein Vektor in Richtung des Ausgangspunktes berechnet (s. die Vektorkarte in Abb. 4). Ein Raumpunkt wird als Bestandteil des Fangbereiches betrachtet, wenn es einen Pfad entlang der Zielvektoren gibt, der bis zum Ausgangspunkt führt. Abbildung 4 zeigt, daß der Fangbereich in günstigen Fällen den gesamten offenen Raum um das Ziel abdecken kann. Der Fangbereich wird kleiner, wenn die Schnappschüsse näher an Objekten aufgenommen werden, da sich hier die momentane Ansicht bei Bewegungen schneller verändert. Dennoch funktionierte der Algorithmus in allen Gebieten zuverlässig, in denen der Roboter nicht mit Hindernissen kollidierte. Dies konnte anhand des folgenden Experiments gezeigt werden: Zunächst wurden an 20 zufällig gewählten Stellen Schnappschüsse aufgenommen und der Khepera anschließend in Entfernungen von 5 bis 25 cm passiv versetzt. Während 100 Testläufen wurde alle 83 ms ein neuer Zielvektor berechnet und die Fahrtrichtung dementsprechend korrigiert<sup>3</sup> Ein Versuch wurde als Erfolg gewertet, wenn der Roboter kollisions-

<sup>3</sup> Auf einem Standard-PC (133MHz Pentium) kann alle 40 ms ein neuer Vektor berechnet werden. Wegen der langsamen seriellen Kommunikation konnte dies allerdings



**Abbildung 4.** Durch den Algorithmus berechnete Zielvektoren für einen zentralen Schnappschuß in der Mitte der Arena (durch Kreis markiert) mit Trajektorien des Khepera bei der Zielfahrt. Der Fangbereich ist grau dargestellt, die Landmarken sind schwarz umrandet.

frei einen Kreis mit 1 cm Radius um das Ziel innerhalb von 30 s erreichte. Es zeigt sich, daß der Algorithmus bis zu einer Entfernung von 15 cm mit einer Erfolgsrate von 95% funktioniert. Danach sinkt die Erfolgsrate bis auf 50% bei einer Entfernung von 25 cm. Hier befand sich der Roboter oft außerhalb des offenen Gebietes um das Ziel, so daß aufgrund von Verdeckungen in der momentanen Ansicht keine sinnvollen Korrespondenzen zu den Schnappschüssen gefunden werden konnten. Ähnliche Erfolgsraten wurden auch mit dem größeren Prototyp innerhalb der Büroumgebung unseres Institutes erzielt. Die maximale erreichte Distanz, von der aus das Ziel wiedergefunden wurde, lag hier bei ca. 2 m.

## 5 Diskussion

Der hier vorgestellte biologisch inspirierte Algorithmus ermöglicht das Anfahren eines durch einen 360°-Schnappschuß markierten Zielortes. Die Zielrichtung wird aus der Verschiebung der umgebenden Landmarken im Vergleich zum Schnappschuß berechnet. Da die für die Berechnung der Zielrichtung benötigten Entfernungen der Landmarken unbekannt sind, müssen diese über sinnvolle Annahmen geschätzt werden. Der Algorithmus basiert auf der vereinfachenden Annahme,

---

nicht direkt zur Korrektur der Fahrtrichtung benutzt werden.

daß alle Landmarken gleich weit vom Ziel entfernt sind. Diese Annahme hat mehrere Vorteile:

1. Die resultierenden parametrisierten Verschiebungsfelder erlauben die Erstellung von Korrespondenzen durch einen globalen Bildvergleich. Die geometrischen Eigenschaften der Aufgabenstellung gehen bei der Berechnung der Korrespondenzen direkt mit ein. Daher können alle berechneten Verschiebungsfelder in der Realität vorkommen, was bei den allgemeinen korrelations- oder gradientenbasierten Methoden zur Korrespondenzberechnung nicht gewährleistet ist.
2. Der aus der Annahme resultierende Fehler geht bei der Zielfahrt gegen null. Es läßt sich beweisen, daß jeder Schnappschuß von einem Fangbereich umgeben ist, der den gesamten offenen Bereich um das Ziel einnehmen kann.
3. Die niedrigdimensionale Parametrisierung führt auf ein schnell lösbares Optimierungsproblem, so daß die Zielrichtung 25 mal pro Sekunde auf einem 133-MHz-PC berechnet werden kann.
4. Obwohl die Annahme einer gleichen Landmarkendistanz nur selten exakt erfüllt ist, sind die praktisch relevanten Distanzen auf einen ringförmigen Bereich um das Ziel begrenzt: Zum einen verhindern die auf den Robotern verwendeten Hindernisvermeidungssysteme eine zu starke Annäherung an Landmarken, zum anderen wirken sich weit entfernte Landmarken auf die Berechnung der Zielrichtung und auf den Gesamtfehler nur wenig aus.

Natürlich bleibt ein auf einem einzelnen Schnappschuß basierendes Navigationssystem auf die unmittelbare Umgebung des Zieles beschränkt. Durch Kombination mehrerer benachbarter Schnappschüsse zu einer graphenartigen Repräsentation der Umwelt lassen sich aber auch großräumige Navigationsaufgaben mit Schnappschüssen bewältigen (Mallot et al., 1997; Franz et al., 1998).

## Literatur

1. Cartwright, B.A., Collett, T.S.: Landmark learning in bees. *J. Comp. Physiol. A* **151** (1983) 521–543
2. Chahl, J.S., Srinivasan, M.V.: Visual computation of egomotion using an image interpolation technique. *Biol. Cybern.* **57** (1996) 405–411
3. Collett, T.S.: Landmark learning and guidance in insects. *Phil. Trans. R. Soc. Lond. B* **337** (1992) 295–303
4. Franz, M.O., Schölkopf, B., Mallot, H.A., Bühlhoff, H.H.: Learning view graphs for robot navigation. *Autonomous Robots* **5** (1998) 111–125
5. Franz, M.O., Schölkopf, B., Bühlhoff, H.H.: Homing by parameterized scene matching. In *Proc. 4th European Conf. on Artificial Life (1997)* 236–245
6. Hong, J., Tan, X., Pinette, R., Weiss, R., Riseman, E.M.: Image-based homing. In *Proc. IEEE Conf. Robotics and Automation (1991)* 620–625
7. Mallot, H.A., Franz, M.O., Schölkopf, B., Bühlhoff, H.H.: The View-Graph Approach to Visual Navigation and Spatial Memory. In *Proc. 7th Intl. Conf. on Artificial Neural Networks (1997)* 751–756
8. Röfer, T.: Controlling a wheelchair with image-based homing. Dept. Computer Sc., Manchester University, Techn. Report UMCS-97-4-1 (1997)

This article was processed using the L<sup>A</sup>T<sub>E</sub>X macro package with LLNCS style