in: Proc. SOAVE'97 - Selbstorganisation von adaptivem Verhalten, Ilmenau September 1997, Fortschrittberichte VDI, Reihe 8, pp. 109-118, VDI-Verlag 1997

Handlungsauswahl durch Antizipation sensomotorischer Konsequenzen*

Torsten Seiler, Volker Stephan, Andrea Heinze, H.-M. Gross
{torsten, vstephan, heinze, homi}@informatik.tu-ilmenau.de
Technische Universität Ilmenau
Fakultät Informatik und Automatisierung
Fachgebiet Neuroinformatik
11. Juli 1997

Zusammenfassung

Diese Arbeit stellt eine Architektur vor, die ein autonom agierendes System in die Lage versetzen soll, eigenes Handeln nicht nur herauszubilden, sondern auf der Grundlage von hypothetisch ausgeführten Aktionen und den daraus resultierenden sensorischen Konsequenzen Handlungssequenzen vor ihrer Ausführung zu bewerten. Durch diese sensomotorisch gesteuerte interne Simulation kann das System aus der Menge der simulierten hypothetischen Aktionssequenzen die im Sinne der Systemaufgabe beste Aktionsfolge ausführen. Für das dabei auftretende Problem der Suche im sensomotorischen Raum wird eine Lösung vorgeschlagen. Zur Repräsentation von Handlungsalternativen und deren Selektion werden neuronale Felder vom Amari–Typ verwendet, die um einige wesentliche Funktionen erweitert wurden.

1 Einleitung und Motivation der Architektur

In der Robotik existieren derzeit zwei große Hauptströmungen. Die eine basiert auf *reaktiven* Steuerungskonzepten (siehe Brooks [2, 3, 4]), die andere auf Konzepten mit Planung und Vorausschau, teilweise hervorgegangen aus der klassischen AI (siehe Chatila [5]).

Hauptkritikpunkt am Brooksschen Ansatz ist, daß eine Entscheidung über die auszuführende Handlung nur anhand der vergangenen und der aktuellen sensorischen Situation möglich ist. Eine Vorhersage über die Entwicklung des sensorischen Inputs in Abhängigkeit von den Systemhandlungen ist nicht möglich. Um dennoch die Auswahl der Handlungen des Systems auch über einen größeren zeitlichen Horizont zu verbessern, wurden Verfahren entwickelt, die Bewertungen vollständiger Sequenzen zukünftiger Aktionen auf einen Skalar abbilden [14, 13]. Der Nachteil dieser Ansätze liegt in dem aufwendigen Lernregime, da prinzipiell jede denkbare Aktionssequenz in jedem Zustand wenigstens einmal erlebt werden muß, um das jeweilige skalare Bewertungsmaß zu erhalten. Es ist somit nicht möglich, eine noch nicht in ihrer Gesamtheit erlebte Folge von Aktionen zu bewerten, obwohl alle Teilfolgen schon erlebt wurden.

Der antizipatorische Ansatz [9, 10] beruht auf der Vorhersage der sensorischen Konsequenzen von Aktionen. Aufbauend auf der sensorischen Prädiktion und Bewertung von Einzelaktionen werden Hypothesenbildung und interne Simulation von Aktionsfolgen möglich. Durch Verkettung bisher erlebter, kurzer Aktionsfolgen können die Konsequenzen längerer Aktionssequenzen

*Die Arbeiten sind Teil der DFG–Projekte SEMRINT (Gr 1378/2-1) und SEMINT (Gr 1378/1-1)

abgeschätzt und bewertet werden, selbst wenn diese Sequenzen noch nie zuvor in ihrer Gesamtheit aufgetreten sind.

Davon ausgehend stellen wir in dieser Arbeit ein Modell zur lokalen Navigation eines mobilen Roboters basierend auf dem optischen Fluss vor, dessen Leistungsumfang von einem einfachen reaktiven Systemverhalten bis hin zu einem System reicht, welches intern vor der Handlungsausführung die sensorischen Konsequenzen hypothetischer Handlungen simuliert. Durch Online-Lernen der Vorhersage sensorischer Konsequenzen von ausgeführten Aktionen einerseits und deren Bewertungen andererseits, ist das System zunehmend in der Lage, das initial reaktive zu einem vorausschauenden[1] Verhalten auszubauen.

2 Modellarchitektur

Wahrnehmung wird verstanden als Generierung von Hypothesen basierend auf bisher erlebten sensomotorischen Zusammenhängen. Zur Generierung von Hypothesen müssen alternative Handlungen repräsentiert und ausgewählt werden können. Sollen im Falle diskreter Handlungsräume Sequenzen von Handlungen betrachtet werden, so entsteht eine Baumstruktur, wobei die Knoten die resultierenden sensorischen Informationen darstellen und die Kanten die verursachenden Aktionen. Bei einem quasikontinuierlichen Aktionsraum nimmt die Breite dieses Baumes rasch zu, wobei viele nahe beieinanderliegende Äste ähnliche Konsequenzen besitzen. Wird das zeitliche Raster zur Auswahl der Aktionen immer enger, so muß der Baum bei gleichbleibendem zeitlichen Planungshorizont entsprechend tiefer untersucht werden, weil eine längere Aktionsfolge erforderlich ist. Beide Fakten laufen der begrenzten Rechenzeit und den begrenzten Ressourcen entgegen. Deshalb begnügen wir uns mit der Untersuchung ausgewählter Hypothesen und betrachten nur einen beschränkten zeitlichen Horizont.

Im folgenden Abschnitt wird das Basissystem unserer Modellarchitektur beschrieben, welches ausgehend von einer sensorischen Situation einen Aktionsvorschlag generiert und dessen Bewertung sowie die zugehörige sensorische Folgesituation abschätzt.

2.1 Sequentieller sensomotorischer Hypothesenprädiktor

In Abb. 1 wird der Aufbau eines sequentiellen sensomotorischen Hypothesenprädiktors veranschaulicht. Ausgehend von einem sensorischen Input[2] und der verursachenden Aktion teilen sich zwei Datenströme. Das Modul *action suggestion* liefert eine topologisch kodierte Menge an Handlungsvorschlägen in Form einer Aktivierungskarte. Durch Modulation dieser Karte kann eine übergeordnete Instanz auf die Vorschlagsgestaltung einwirken und somit die lokale Navigation in ein globales Navigationskonzept integrieren. Für die nachfolgenden Betrachtungen ist diese Modulation zunächst nicht von Bedeutung. Aus dieser Aktivitätskarte werden im Modul *action selection*, unter Nutzung eines neuronalen Feldes mit der in Abschnitt 3 beschriebenen Dynamik, sequentiell einzelne Aktionsvorschläge extrahiert. Die *optical flow prediction* schätzt die sensorische Folgesituation des Aktionsvorschlages. Abschließend liefert die *hypothesis evaluation* das zugehörige Aktionsbewertungssignal. Somit wurde *eine* hypothetische Aktion intern simuliert.

[1] Vorausschauend ist in diesem Zusammenhang als „interne Simulation unter Berücksichtigung zugehöriger Bewertung“ zu verstehen.

[2] Wir benutzen den optischen Fluß als sensorischen Input.

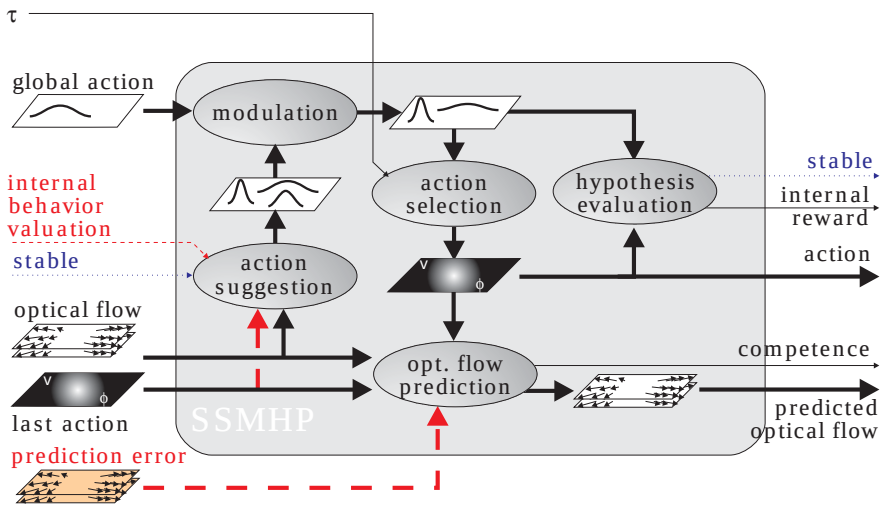

Abbildung 1: *Aufbau des sequentiellen sensomotorischen Hypothesenprädiktors (SSMHP). Er besteht aus einem Modul zur Generierung einer topologisch kodierten Menge von Handlungsalternativen (action suggestion), einem Modul zur Selektion von Handlungsalternativen (action selection), einem Modul zur Vorhersage der sensorischen Konsequenzen der ausgewählten Aktion (optical flow prediction) und einem Modul zur Bewertung der hypothetischen Aktion (hypothesis evaluation).*

2.1.1 Aktionsvorschlag

Das Modul *action suggestion* liefert, ausgehend von der vorliegenden sensomotorischen Situation, eine Menge von Aktionsvorschlägen, die in einer Aktionskarte kodiert sind. In einer Dimension der Aktivitätskarte werden die Geschwindigkeiten, in der anderen die Lenkwinkel kodiert. Durch topologische Kodierung kann jeder beliebige Lenkwinkel und jede Geschwindigkeit innerhalb der zulässigen Intervalle realisiert werden. Somit repräsentiert jeder Punkt in der Karte einen Punkt im kontinuierlichen Aktionsraum und dessen Aktivierung die bisher erlernte Bewertung bezüglich dieser Aktion.

Das Modul *actions suggestion* ist als einfaches Feed–forward Netzwerk realisiert, wobei perspektivisch auch eine inkrementelle Architektur eingesetzt werden kann. Die Ausgabe des Moduls wird durch eine einfache Deltaregel adaptiert, wobei sich die Änderung der Ausgabe an der Stelle $\vec{r}$ aus der vorgeschlagenen Aktionsmenge $y(\vec{r}, t-1)$ und der tatsächlich ausgeführten Aktion $a(\vec{r}, t)$ ergibt (Gleichung 1).

$$\Delta y(\vec{r}) = \eta \cdot a(\vec{r}, t)\,[b - y(\vec{r}, t-1)] \tag{1}$$

η ist die Lernrate und $b \in [0..1]$ die Bewertung der ausgeführten Aktion.

2.1.2 Aktionsauswahl

Die sequentielle Auswahl der vorgeschlagenen Aktionen aus der Aktivitätskarte erfolgt im Modul *action selection*. Die im Abschnitt 3 beschriebene Felddynamik selektiert sequentiell einzelne Regionen (und damit konkrete Aktionen) aus dem Bewertungsgebirge und stabilisiert diese für eine bestimmte Zeit, wobei deren Reihenfolge von der Bewertung bzw. Energie[3] der einzelnen Regionen abhängt. Auf diese Weise werden zuerst die erfolgversprechendsten Aktionen untersucht und, je nach verfügbarer Simulationszeit, im weiteren Verlauf auch die weniger gut bewerteten, berücksichtigt.

[3]Unter Energie verstehen wir das vom Inputgebirge überdeckte Volumen, welches mit der Bewertung der jeweiligen Aktion korreliert.

2.1.3 Prädiktion der sensorischen Konsequenz

Nach erfolgreicher Selektion einer Aktion wird diese mit der aktuellen sensorischen Situation im Modul *optical flow prediction* verknüpft und die daraus resultierende sensorische Folgesituation geschätzt. Die Adaption dieses Moduls erfolgt jeweils nach realer Ausführung einer hypothetischen Aktion, indem die prädizierte und die reale Folgesituation auf Sensorebene verglichen werden. Die Güte der Vorhersagen wird in einem situations- und aktionsspezifischen Kompetenzsignal $c_{\vec{r}}$ gespeichert, welches erst bei kleinen Prädiktionsfehlern e^{Pred} hohe Werte liefert (Gleichung 2). Unter Nutzung dieses Kompetenzsignales ist die Abschätzung einer sinnvollen Simulationstiefe möglich. Auf diese Weise können die knappen Simulationsressourcen auf realistische Planungszweige konzentriert und effektiv genutzt werden.

$$c_{\vec{r}}(t+1) = (1-\eta^c)\cdot c_{\vec{r}}(t) + \eta^c \frac{1}{1+e^{Pred}_{\vec{r}}} \tag{2}$$

2.1.4 Hypothesenbewertung

Die Hypothesenbewertung basiert auf dem Modul *action suggestion* und liefert eine skalare Bewertung für vorgeschlagene hypothetische Aktionen.

2.2 Prädiktion und Bewertung von Sequenzen

Der bisher beschiebene *Sequentielle Sensomotorische Hypothesenprädiktor* ist in der Lage, ausgehend von einer sensorischen Situation sequentiell konkrete Aktionen vorzuschlagen und die damit verbundenen Bewertungen sowie die sensorischen Folgesituationen abzuschätzen. Durch Verkettung mehrerer solcher Hypothesenprädiktoren können Handlungssequenzen intern simuliert werden, indem der erste Prädiktor über der realen Situation operiert, der zweite dagegen auf dem Prädiktionsergebnis des ersten, usw. (Abb. 2). Die Anzahl der auf diese Weise verwendeten replikativen Teilsysteme bestimmt die maximale Simulationstiefe im Suchbaum.

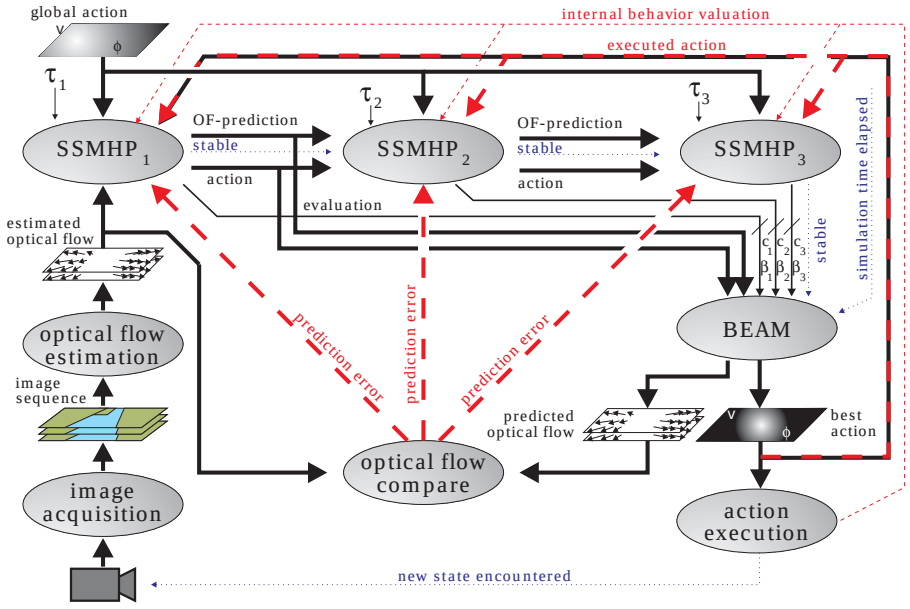

Abbildung 2: *Kopplung von drei sequentiellen sensomotorischen Hypothesenprädiktoren (SSMHP) mit unterschiedlichen Dynamikzeitkonstanten τ zur Simulation von drei aufeinanderfolgenden Aktionen. Die Startaktion wird zusammen mit den geschätzten Aktionsbewertungen β und Prädiktionskompetenzen c im Best Evaluated Action Memory (BEAM) gespeichert.*

Da in einer Situation unter Umständen auch mehrere hoch bewertete Aktionen vorgeschlagen

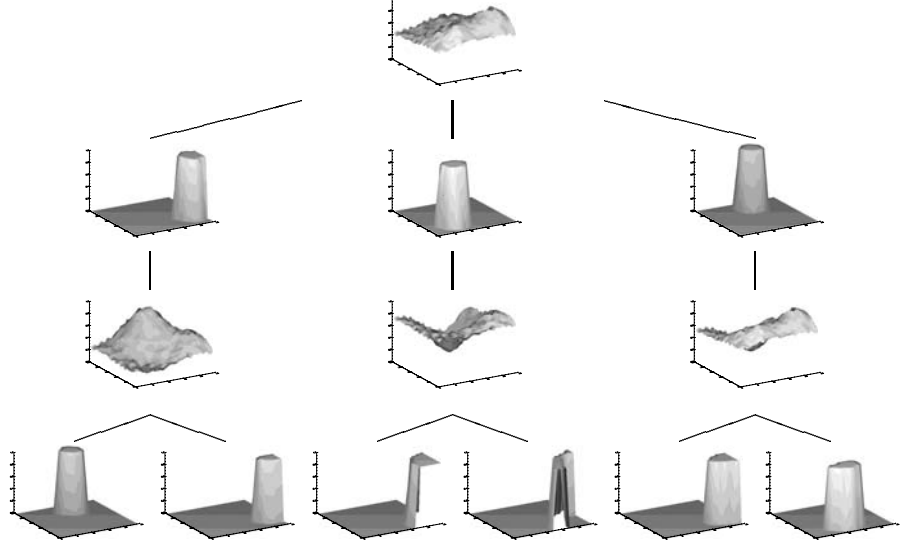

Abbildung 3: *Prinzip des zeitlichen Wechselspiels von zwei sequentiellen sensomotorischen Hypothesenprädiktor-Einheiten bei der Simulation unterschiedlicher Motorsequenzen (Baumsuche der Tiefe 2). Im obersten Bild ist der generierte Aktionsvorschlag für das erste der zwei Module zu erkennen. Aus diesem bildet sich ein spezifischer Blob heraus, der als Eingabe (zusammen mit der prädizierten sensorischen Situation) für den nachfolgenden Hypothesenprädiktor fungiert. Daraufhin wird ein Aktionsvorschlag generiert, aus dem wiederum eine spezifische Aktion ausgewählt wird, usw. Zerfällt die Eingabe eines Prädiktors durch die temporale Inhibition, so ist die Simulation für diesen Bereich des Suchbaumes beendet. Die Stabilisierungsdauer der ausgewählten Aktion nimmt mit zunehmender Tiefe ab. Insgesamt wurden in dieser Darstellung 6 verschiedene Handlungssequenzen analysiert.*

werden, ist jeder *Hypothesenprädiktor* in der Lage, sequentiell mehrere verschiedene Handlungen hypothetisch auszuführen (Breitensuche). Die Wahl der Dynamik der neuronalen Felder legt die Breite der Baumsuche fest, also wieviele Aktionen bezüglich einer sensorischen Situation sequentiell analysiert werden. Die einzelnen *Hypothesenprädiktoren* arbeiten auf gestaffelten zeitlichen Skalen, so daß der jeweilige Nachfolger nur solange Zeit hat, eigene Hypothesen zu testen, wie die Ausgaben des Vorgängers stabil bleiben. Auf diese Weise werden in der verfügbaren Simulationszeit mehrere Aktionssequenzen durchsimuliert, wobei die erste Aktion der am besten bewerteten Sequenz der bisher intern simulierten Aktionsfolgen im *Best Evaluated Action Memory* (BEAM) gespeichert wird (Abb. 2).

In Abb. 3 ist exemplarisch das zeitliche Wechselspiel von zwei *Hypothesenprädiktoren* dargestellt, wobei das erste Modul sequentiell drei Aktionen aus der Aktionsbewertungskarte seiner *action suggestion* selektiert und das nachgeschaltete System auf der neuen Situation aufbauend wiederum eine Aktionsbewertungskarte mittels *action suggestion* generiert und daraus jeweils zwei Einzelaktionen auswählt.

3 Sequentielle Hypothesenbildung mittels erweiterter Amari–Dynamik in neuronalen Feldern

3.1 Neuronale Felder

Die dynamischen Felder nutzen wir, um aus der Aktionskarte des *action suggestion* sequentiell einzelne konkrete Aktionen zu selektieren, eine bestimmte Zeit stabil zu halten und danach wieder zerfallen zu lassen, um nachfolgenden Modulen die Möglichkeit zu geben, eigene Akti-

onsalternativen abzutesten.

Sie basieren auf den von Amari [1] beschriebenen nichtlinearen rückgekoppelten neuronalen Feldern mit einer topologisch geordneten Eingabe. Die Dynamik solcher Felder läßt sich nach [8] durch Gleichung (3) beschreiben, wobei $\vec{r} \in R^2$ die Position eines Neurons im Feld ist.

$$\tau \frac{d}{dt} u(\vec{r},t) = -u(\vec{r},t) - h + \int_R w(\vec{r}-\vec{r'}) S[u(\vec{r'},t)] d^2\ r' + x(\vec{r},t) \tag{3}$$

$S[u(\vec{r'},t)]$ ist eine positive sigmoide Funktion, τ eine Zeitkonstante, h die globale Inhibition, $x(\vec{r},t)$ der Input und $w(\vec{r}-\vec{r'})$ die laterale Kopplung zwischen den einzelnen Neuronen. Die laterale Kopplung enstpricht einer Mexican-Hat Funktion, d.h. einer lokal exzitatorischen und global inhibitorischen Funktion.

Die Felddynamik selektiert aus dem Inputmuster die Region mit der maximalen Energie, hebt diese durch hohe Neuronenaktivierungen heraus und unterdrückt alle anderen Neuronen. Dieses aktivierte Neuronenensemble kodiert durch seine Lokalisation im neuronalen Feld die Aktionshypothese mit der höchsten Bewertung in der aktuellen Situation. Um sequentiell mehrere Aktionshypothesen zu generieren, benötigen wir zusätzlich eine Eigenschaft, die als temporäre Inhibition bezeichnet werden kann. Nach einer gewissen Zeit soll eine einmal selektierte Region unterdrückt werden, damit auch noch weitere Hypothesen ausgewählt werden können.

3.2 Erweiterung der Felddynamik

Aufbauend auf den Arbeiten von Gross [7, 6] wurde die Amari–Felddynamik bezüglich der gewünschten zeitlichen Verhaltensweise modifiziert [11, 12]. Mit Hilfe dieser Erweiterungen ist es nun möglich, die energiereichsten (am besten bewerteten) Regionen gesondert im Eingaberaum sequentiell zu selektieren. Die gewünschte temporale Inhibition wird durch eine zweite Neuronenschicht aus sogenannten Chandelierzellen realisiert.

In Gleichung (4) wird die Dynamik eines Neurons der unteren Schicht des neuronalen Feldes an der Position $\vec{r}$ beschrieben. Die Neuronen in dieser Schicht werden als Pyramidenzellen bezeichnet.

$$\begin{aligned} \tau \frac{d}{dt} z_{Py}(\vec{r},t) &= -z_{Py}(\vec{r},t) + \alpha_{NB} \int_R w(\vec{r}-\vec{r'}) y_{Py}(\vec{r'},t) d^2\ r' + \alpha_I \cdot x(\vec{r},t) \\ &\quad -h \cdot (1 - y_{Py}(\vec{r},t)) \cdot S\left[\theta_{NB} - \int_R w(\vec{r}-\vec{r'}) y_{Py}(\vec{r'},t) d^2\ r'\right] \end{aligned} \tag{4}$$

$$\text{mit:} \quad y_{Py}(\vec{r},t) = \begin{cases} S[z_{Py}(\vec{r},t)] & : \ y_{Ch}(\vec{r},t) < \theta_{Ch} \\ 0 & : \ \text{sonst} \end{cases} \tag{5}$$

z_{Py} ist die Aktivierung der Pyramidenzelle, y_{Py} deren Ausgabe, y_{Ch} die Ausgabe der korrespondierenden Chandelierzelle (aus der darüberliegenden, zweiten Schicht zur temporalen Inhibition), $S[\cdot]$ eine Sigmoidfunktion, θ_{Ch}, θ_{NB} sind Schwellwerte, $w(\vec{r}-\vec{r'})$ die laterale Kopplung zwischen den einzelnen Neuronen, α_{NB}, α_I sind Faktoren, h die globale Inhibition und x der Input des neuronalen Feldes.

Die Chandelierzellen dienen als Akkumulator der korrespondierenden Pyramidenzellenausgaben und inhibieren oberhalb einer Schwelle diese Pyramidenzellenausgaben. Die Folge der Inhibition ist ein Wegfall des Inputs der Chandelierzelle, der wiederum nach einer bestimmten Zeit zu einer Deaktivierung dieser Zelle führt.

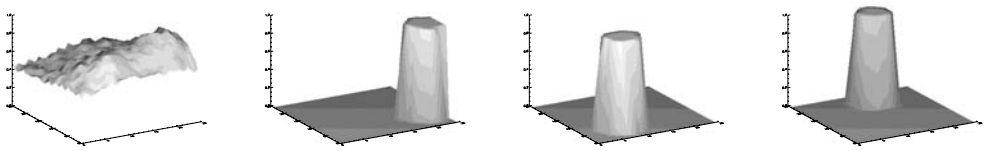

Abbildung 4: *Beispiel der erweiterten Amari–Dynamik. Ausgehend von der stabilen Eingabe (1, v.l.n.r.) werden zeitlich nacheinander die Ausgaben (2) bis (4) generiert. Diese bleiben jeweils eine bestimmte Zeit aktiv und zerfallen anschließend wieder. Klingt die temporale Inhibition ab, so beginnt der Zyklus von vorn.*

Der Vergleich von Gleichung (4) mit Gleichung (3) zeigt folgende Modifikationen:

- Wichtung des Inputs $x(\vec{r},t)$ mit α_I zur Steuerung des Einflusses des Inputs,
- die Modulation der mit α_{NB} gewichteten Umgebungsaktivität durch den Input $x(\vec{r},t)$ stellt sicher, daß sich die Aktivierungen der Pyramidenzellen nur in solchen Regionen ausbreiten können, die von einem realen Input unterstützt werden,
- die Unterdrückung der globalen Inhibition h durch die Ausgabe $y_{P_y}(\vec{r},t)$ stellt sicher, daß sich unterschiedlich große Pyramidenzellregionen unabhängig von der globalen Inhibition aktivieren können,
- die Unterdrückung der globalen Inhibition durch die mittlere Umgebungsaktivität oberhalb der Schwelle θ_{NB} stellt sicher, daß sich das Aktivierungsmuster auch über schwache Inputregionen ausbreiten kann, ohne explosionsartig zu diffundieren.

In Abb. 4 wird der Selektionsmechanismus dargestellt. Ausgehend von einem stabilen Input in Form einer Aktivitätskarte werden nacheinander verschiedene Gebiete herausgehoben. Nach einer gewissen Zeit beginnt der Zyklus wieder von vorn. In [11, 12] wird eine ausführliche mathematische Beschreibung der modifizierten Felddynamik gegeben.

4 Ergebnisse

Alle Untersuchungen zur Leistungsfähigkeit der vorgestellten Architektur wurden zunächst auf einem Simulator durchgeführt. Der simulierte Roboter bewegt sich in einem unbekannten Labyrinth und besitzt ebenso wie die reale Zielplattform Sensorik zur Erfassung des optischen Flusses und zur Kollisionsdetektion. Da der optische Fluß in hohem Maße redundante Eigenbewegungsinformationen beinhaltet, wurde eine Normierung der Flußbilder auf eine rein translatorische Bewegung mit fester Geschwindigkeit vorgenommen. Die Motorik des Roboters gestattet analoge Geschwindigkeiten in einem Intervall zwischen $0\frac{cm}{s}$ und $5\frac{cm}{s}$ (entspricht Roboterlänge) bei einem analogen Lenkwinkelbereich zwischen $-45°$ und $+45°$.

Während das System in seiner Umwelt navigiert, erhält es ein skalares Bewertungssignal, welches die Navigationsleistungen hinsichtlich der a priori definierten Zielstellung (kollisionsfreie, schnelle Fahrt) bewertet (Abb. 5). Das System soll versuchen, ein möglichst hohes Bewertungssignal von der Umwelt zu erhalten (keinen Schmerz und effektive Lokomotion).

Eine Grundvoraussetzung für jegliche Art von Navigation ist die Fähigkeit, die Konsequenzen von Aktionen in verschiedenen Situationen abschätzen zu können. Diese Leistung erbringt das Teilsystem (*action suggestion*), welches während einer Erkundungsphase in seiner Umwelt *allgemeine* sensomotorische Zusammenhänge selbstorganisierend erlernt.

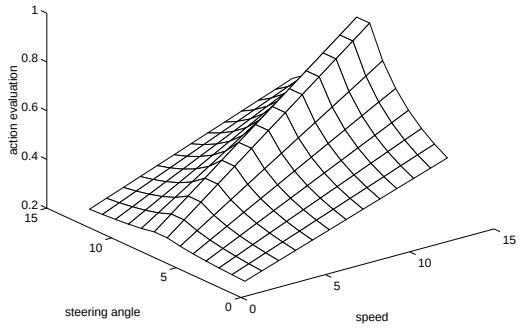

Abbildung 5: *Darstellung einer ersten einfachen Funktion zur Berechnung eines skalaren Bewertungsmaßes (Reinforcement) in Abhängigkeit von Geschwindigkeit und Lenkwinkel. Wenn durch die ausgeführte Aktion keine Kollision verursacht wurde (Bewertung = 0.0), ist das Reinforcement proportional der Geschwindigkeit und nimmt mit größeren Lenkwinkeln ab.*

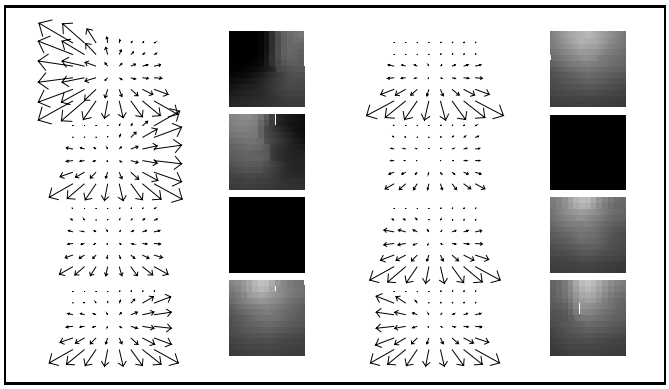

Abbildung 6: *Darstellung von 8 der verwendeten 50 NG–Neuronen mit trainierten Input– (optischer Fluß jeweils links) und Outputwichtungen (Grauwertverteilung jeweils rechts), wobei dunkle Regionen schlechte und helle gute Bewertungen kodieren. Innerhalb der Karte repräsentiert oben „schnell vorwärts“, unten „stop“, links „nach links“ und rechts „nach rechts“ Bewegung.*

In einer ersten Implementierung nutzen wir einen adaptiven neuronalen Vektorquantisierer (Neural Gas) zur Abbildung des optischen Flusses auf sensorische Zustände. An diese Zustände werden durch eine nachgeschaltete Neuronenschicht situationsabhängige Aktionsbewertungen geknüpft. Da der Roboter über ein kontinuierliches Aktionsspektrum verfügt, werden die Aktionsbewertungen in einer zweidimensionalen Neuronenkarte mit 13×13 Neuronen quasikontinuierlich repräsentiert.

In Abb. 6 sind exemplarisch ausgewählte Neuronen des Vektorquantisierers mit ihren Input– und Outputwichtungen dargestellt. Es ist deutlich erkennbar, daß sich einzelne Neuronen auf verschiedene sensorische Situationen einstellen und auch entsprechende Aktionsbewertungskarten erlernt haben. Das Neuron 1 (links oben) realisiert in Situationen mit Hindernissen im linken Bildbereich eine nach rechts ausweichende Fahrweise, Neuron 2 (links, 2. Zeile) in entgegengesetzter Hinderniskonstellation eine Ausweichbewegung nach links. In Situationen mit Freiraum wird dagegen eine schnelle Geradeausfahrt bevorzugt (z.B. Neuron 5 rechts oben).

Die Fähigkeit zur internen Simulation hypothetischer Aktionen erfordert die Vorhersage deren Konsequenzen. In unserer Architektur erfolgt die Antizipation direkt auf sensorischer Ebene unter Nutzung bewegungsspezifischer Perzeptrons, welche eine lineare Abbildung des alten auf das neue Flußbild realisieren. In Abb. 7 ist ein Ausschnitt aus einer gefahrenen Sequenz mit den jeweiligen prädizierten und realen Flußbildern dargestellt. Es ist erkennbar, daß das Prädiktionsnetzwerk in der Lage ist, die sensorischen Konsequenzen dieser hypothetischen Aktionen mit ausreichender Qualität vorherzusagen.

Zur Demonstration der Leistungsfähigkeit wurden ein Roboter mit zufälliger Aktionsauswahl, ein rein reaktiv agierender Roboter, welcher unter Nutzung der em action suggestion in der aktuellen Situation die am besten bewertete Aktion ausführt, und unser intern simulierendes System unter identischen Randbedingungen im Simulator hinsichtlich ihres Navigationsverhalten verglichen. Abb. 8 zeigt die zeitlichen Verläufe des Reinforcementsignales, der Geschwindigkeit und des Lenkwinkels der drei Navigationsansätze. Das Reinforcementsignal ist ein direktes Maß für die jeweilige Navigationsleistung. Der Vergleich zwischen den drei Systemen zeigt, daß der

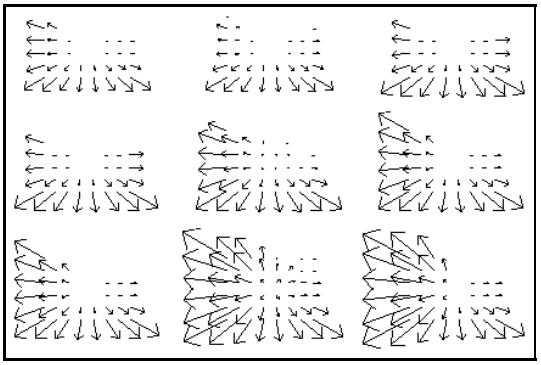

Abbildung 7: *Flußbilder aus drei aufeinanderfolgenden Situationen, wobei sich der Roboter durch schnelle Fahrt mit leichter Linksdrehung einer Wand nähert. Die erste Spalte zeigt den alten realen optischen Fluß, die zweite die aktionspezifische Prädiktion und die dritte Spalte das reale Folgeflußbild.*

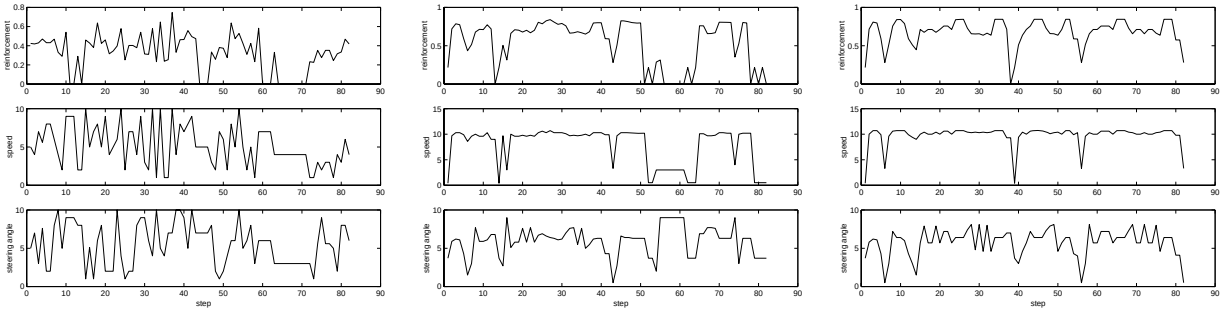

Abbildung 8: *Vergleich des zufällig navigierenden (links), reaktiven (mitte) und intern planenden Roboters (rechts) anhand des Reinforcementsignales (oben), der gefahrenen Geschwindigkeit (mitte) und des Lenkwinkels (unten). Der intern simulierende Roboter fährt schneller und sicherer als die anderen und erhält deshalb ein höheres Bewertungssignal (Reinforcement).*

intern simulierende gegenüber dem reaktiven Ansatz Vorteile besitzt, weil er nicht nur eine, sondern bis zu drei aufeinanderfolgende hypothetische Aktionen bewerten kann. Auf diese Weise kann er Gefahrensituationen schon deutlich früher erfassen und rechtzeitig Ausweichmanöver einleiten. Der reaktive Roboter bezieht in seine Aktionsauswahl lediglich die Aktionsbewertungen der aktuellen Situation ein, beachtet nicht deren Folgen und gerät deshalb häufiger in kritische Situationen. Der Vergleich anhand numerischer Kennwerte des Navigationsverhaltens untermauert den Vorteil des intern simulierenden Systems:

Typ	Simulationsschritte	Kollisionen	mittleres Reinforcement
zufällig fahrend	82	17	0.32
reaktiv	82	13	0.54
intern simulierend	82	1	0.69

Die Kollisionen des intern simulierenden Systems sind auf ungenaue Prädiktionen bei stärkeren Richtungsänderungen zurückzuführen, welche insbesondere an Gefahrenstellen notwendig werden. Durch die Rotation erscheinen neue Objekte im Kamerabild, deren Auswirkungen auf den optischen Fluß vom Antizipationsmodul im gegenwärtigen Entwicklungsstand nicht vorhergesagt werden können.

5 Zusammenfassung und Ausblick

Wir stellten ein System vor, welches basierend auf topologisch kodierten Mengen hypothetischer Aktionen eines kontinuierlichen Aktionsspektrums sequentiell einzelne Handlungen selektieren, eine Bewertung durchführen und die sensorischen Konsequenzen abschätzen kann.

Dieses Teilsystem realisiert bereits einen reaktiven Navigationsansatz. Durch Verkettung mehrerer Hypothesenprädiktoren ist es möglich, längere Aktionssequenzen intern zu simulieren, um anschließend die beste gefundene Sequenz zur realen Ausführung zu bringen. Die Modellarchitektur ist im Gegensatz zu starken Reinforcementverfahren konzeptionell sogar in der Lage Sequenzen zu analysieren, die bisher noch nie in ihrer Gesamtheit durchlebt wurden.

Die Problematik der Baumsuche in kontinuierlichen Aktionsräumen wurde in ein replikatives System von Hypothesenprädiktoren auf der Basis dynamischer neuronaler Felder verlagert. Durch Wahl geeigneter Dynamikparameter auf den verschiedenen Hypothesenbildungsebenen konnte ein anwendbarer Kompromiß zwischen konventioneller Breiten– und Tiefensuche gefunden werden.

Die Vorhersage typischer sensorischer Konsequenzen von Aktionen erlaubt es perspektivisch, die Abweichungen nicht mehr zwangsläufig als Fehler der Prädiktion zu betrachten, sondern Aufmerksamkeitsprozesse auf die Abweichungen zwischen prädiziertem und realem sensorischen Input nach Handlungsausführung zu lenken. Voraussetzung dafür ist ein Zuverlässigkeits- oder Kompetenzmaß der Prädiktion. Dazu werden gerade Konzepte entwickelt und untersucht.

Den Untersuchungen im Simulator folgen Tests auf realen Systemen, wie dem Miniaturroboter KHEPERA und der Roboterplattform MILVA. Weiterhin werden ausführliche Vergleiche mit starken Reinforcementverfahren, welche ebenfalls Folgesituationen in die Aktionsauswahl einbeziehen, vorgenommen.

Literatur

[1] Shun–Ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] R.A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, (6):3–15, 1990.

[3] R.A. Brooks. Intelligence without representation. *Artificial Intelligence*, (47):139–159, 1991.

[4] R.A. Brooks. New approaches to robotics. *Science*, (253):1227–1232, 1991.

[5] R Chatila. Control architectures for autonomous mobile robots. In *Proc. of PerAc'94*, pages 254–265. IEEE Computer Society Press, 1994.

[6] H.-M. Gross et al. A hierarchical neural network for data driven and knowledge controlled selective visual attention. In *Proc. 14. DAGM-Symposium Mustererkennung*, pages 341–346. Informatik-Aktuell, Springer, 1992.

[7] H.-M. Gross. *Simulation eines Arrays corticaler Prozessoren zur inhaltsgesteuerten parallelen Informationsverarbeitung nach dem Vorbild des primären visuellen Cortex.* Dissertation, TH Ilmenau, 1989.

[8] K. Kopecz. Neural field dynamics provide robust control of attentional resources. In *Proc. Workshop: Aktives Sehen in technischen und biologischen Systemen, Hamburg*, 137–144, infix 1996.

[9] R. Möller and H.-M. Gross:. Perception through anticipation. In P. Gaussier and J.-D. Nicoud, editors, *From Perception to Action Conference PerAc'94, Los Alamitos.*, pages 408–411. IEEE Computer Society Press, 1994.

[10] Ralf Möller. *Wahrnehmung durch Vorhersage – Eine Konzeption der handlungsorientierten Wahrnehmung.* Dissertation, TU-Ilmenau, Juni 1996.

[11] V. Stephan. Vision-basierte Ansätze für ein selbstorganisierendes Explorationsverhalten eines mobilen Miniaturroboters in einer Labyrinthwelt. Diplomarbeit, TU Ilmenau, 1996.

[12] V. Stephan and H.-M. Gross. Formerhaltende sequentielle visuelle Aufmerksamkeit in columnar organisierten neuronalen Feldern. In Proc of 19. DAGM-Symposium, September 1997.

[13] R.S. Sutton. Learning to predict by the methods of temporal differences. Machine Learning Vol. 3, 9-44, 1988.

[14] Ch. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8:279–292, 1992.