

Vorhersage von optischen Flußfeldern zur antizipativen Navigation

V. Stephan, T. Seiler, H.-M. Gross
Technische Universität Ilmenau

Fachgebiet Neuroinformatik
{vstephan, torsten, homi}@informatik.tu-ilmenau.de

Zusammenfassung

Dieser Beitrag stellt einen antizipativen Navigationsansatz vor, welcher auf einer Vorhersage sensorischer Konsequenzen von hypothetisch ausgeführten Aktionen basiert. Auf dieser Prädiktion aufbauend, können ganze Aktionssequenzen mit ihren Konsequenzen für das System intern simuliert und die beste Aktionssequenz real ausgeführt werden. Basierend auf diesem antizipierenden Ansatz stellen wir eine neuronale Architektur vor, welche ausgehend von einem reaktiven Verhalten ein vorausschauendes und planendes Verhalten entwickelt.

1 Einleitung

Die derzeit in der Robotik existierenden Navigationsansätze lassen sich einerseits bzgl. des Einsatzes eines Weltmodells und andererseits hinsichtlich ihres Planungshorizontes einteilen. Typische, aus der Literatur bekannte, modellbasierte Verfahren werden in [12, 17, 16] beschrieben, wobei [12, 16] auch durch interne Simulation auf dem Weltmodell einen entsprechend großen Planungshorizont berücksichtigen. Modellfreie Ansätze, welche im Gegensatz zu reaktiven Verfahren [4] auch zukünftige Folgesituationen mit in die Aktionsauswahl einbeziehen, werden u.a. in [20, 1] beschrieben. Prinzipiell ge-

statten sowohl die Nutzung eines Weltmodells als auch ein möglichst großer Planungshorizont eine Verbesserung des Navigationsverhaltens. Deshalb beschäftigt sich die vorliegende Arbeit mit einer neuronalen Architektur zur modellbasierten antizipierenden Navigation eines mobilen Roboters.

Dieser Beitrag basiert auf Arbeiten von [13], welche die Trennung von Wahrnehmung und Verhaltensgenerierung aufheben und in einem neuronalen Prozeß verschmelzen. Der Kerngedanke dieses Konzeptes besteht in der Wahrnehmung einer sensorischen Situation durch Generierung und Auswertung der sensorischen Konsequenzen von hypothetisch ausgeführten Aktionen. Nach erfolgter interner Simulation wird die Aktion ausgeführt, welche die höchste Bewertung bezüglich des Systemziels (z.B. Hindernisvermeidung) verspricht. Auf diese Weise wird der visuelle Eindruck des Systems nicht direkt, sondern durch mögliche Verhaltensalternativen charakterisiert.

Dieses neuartige Wahrnehmungskonzept wird von verschiedenen Autoren mitgetragen [2, 12, 10]. Die sensorische Prädiktion von Aktionskonsequenzen wurde bereits in [19, 18] vorgestellt, die Tragfähigkeit dieses antizipativen Konzeptes zur Navigation eines realen mobilen Systems wurde bisher jedoch noch nicht nachgewiesen.

Um die Vorteile eines antizipierenden Ansatzes zu nutzen, müssen mehrere Aktions-

sequenzen intern simuliert werden, um dann die beste real auszuführen. Das Problem bei dieser internen Simulation von Aktionen liegt in einer exponentiell wachsenden Anzahl möglicher Aktionsketten im Suchbaum, da in jedem Zustand in der Regel mehrere alternative Handlungen ausführbar sind. Da aus Rechenzeit- und Kapazitätsgründen nicht der gesamte Suchbaum untersucht werden kann, war ein neuer Zugang zu finden, der eine Suche über den für das sensomotorische System relevanten Aktionshypothesen realisiert.

2 Modellarchitektur

Zur Antizipation und Bewertung sensomotorischer Aktionskonsequenzen nutzen wir eine sequentielle Verschaltung von mehreren sogenannten Sequentiellen Sensomotorischen Hypothesenprädiktoren (*SSMHP*) (Abb. 1). Jedes dieser Module kann dabei die sensorischen Konsequenzen genau *einer* hypothetischen Aktion schätzen und bewerten.

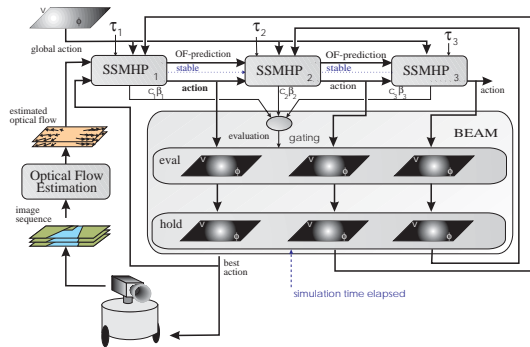


Abbildung 1: Verschaltung von 3 SSMHPs zur internen Simulation von Handlungssequenzen der Länge 3. Das erste Modul arbeitet auf dem realen sensorischen Input, das nachgeschaltete Modul operiert auf dessen hypothetischer Folgesituation usw. Weiterführende Betrachtungen werden in [6, 7] vorgestellt.

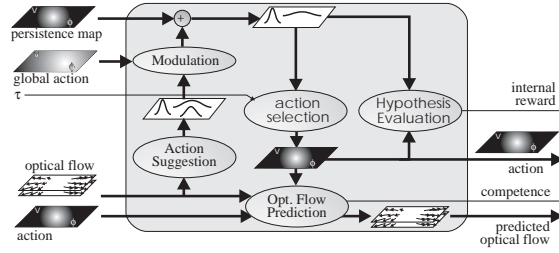


Abbildung 2: Aufbau eines SSMHPs. Eine Beschreibung wird im Text gegeben.

2.1 Sequentieller Sensomotorischer Hypothesenprädiktor

Das Modul *SSMHP* integriert verschiedene Submodule, welche ausgehend von einer sensorischen Situation (real oder hypothetisch) eine Sequenz von alternativen Aktionsvorschlägen generieren und deren Bewertungen sowie die sensorischen Folgesituationen präzisieren (Abb. 2).

Das Modul *Action Suggestion* liefert eine topologisch kodierte Menge an alternativen Handlungsvorschlägen in Form einer Motorkarte. Durch Modulation dieser Karte kann eine im vorherigen Simulationszyklus generierte und am besten bewertete Aktion im Sinne einer Persistenz und der Aktionsvorschlag einer übergeordneten Verhaltensebene auf die Aktionsselektion einwirken und somit die lokale Navigation in ein globales Navigationskonzept integrieren.

Aus der resultierenden Motorkarte werden im Modul *Action Selection*, unter Nutzung eines neuronalen Feldes [15, 5], sequentiell einzelne Aktionsvorschläge in absteigender Reihenfolge bzgl. deren Bewertung extrahiert. Die *Optical Flow Prediction* schätzt die jeweiligen sensorischen Folgesituationen für die selektierten Aktionsvorschläge und abschließend liefert die *Hypothesis Evaluation* das zugehörige Aktionsbewertungssignal. Somit können sequentiell je nach verfügbarer Zeit mehrere alternative hypothetische Aktionen für *einen* Schritt in die Zukunft intern simuliert werden.

Die Lernprozesse in den Modulen *Action Suggestion* und *Optical Flow Prediction* sind vollständig voneinander entkoppelt. Das bedeutet, daß in einem initial untrainierten System beide Lernvorgänge parallel ablaufen können, da einerseits eine schlechte Prädiktion keinen Einfluß auf das Abbilden von Aktionsbewertungen hat und andererseits eine ungenaue Aktionsbewertung das Lernen der Folgezustandsabbildung nicht beeinträchtigt.

2.2 Sensorischer Eingaberaum

Ein zentraler Anspruch dieses Beitrages ist der Einsatz des antizipierenden Navigationssystems in einer Real-World-Umgebung unter Nutzung visueller Informationen. Es ist deshalb eine Sensorik erforderlich, welche eine für das mobile System relevante Situationsbeschreibung ermöglicht und möglichst robust gegenüber Störungen ist. Wir nutzen deshalb den optischen Fluß, welcher durch ein Korrelationsverfahren [14] aus einer monokularen Bildfolge geschätzt wird.

Da zur Navigation ausschließlich Informationen über die Objektanordnung der Umgebung relevant sind, eliminieren wir die Eigenbewegungseinflüsse aus dem geschätzten optischen Fluß (Abb. 3).

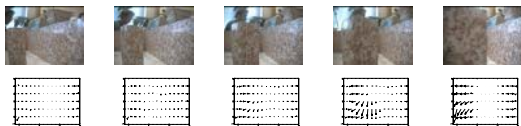


Abbildung 3: Darstellung einer Sequenz von realen Kamerabildern (oben) und den darin während einer Geradeausfahrt geschätzten optischen Flußfeldern (unten).

2.3 Aktionsbewertung

Das Modul *Action Suggestion* liefert, ausgehend von der vorliegenden sensomotorischen Situation, eine Menge von bewerte-

ten Aktionsvorschlägen, die in einer zweidimensionalen Motorkarte (Lenkwinkel und Geschwindigkeit) topologisch kodiert sind (Abb. 4). Diese Motorkarten gestatten einerseits die Bewertung kontinuierlicher Aktionen und haben andererseits den Vorteil, daß mehrere Aktionen unabhängig voneinander repräsentiert werden können. Das ist eine wesentliche Voraussetzung für eine interne Simulation, bei der auch alternative Aktionen berücksichtigt werden sollen.

Die neuronale Architektur zum Erlernen der Aktionsbewertungen besteht aus einem Neural-Gas-Clusterer [11], welcher auf der sensorischen Situation operiert, und einer nachgeschalteten zweidimensionalen Ausgabekarte, welche mit der Clusterschicht vollverschaltet ist und durch normierte Superposition die Motorkarte generiert.

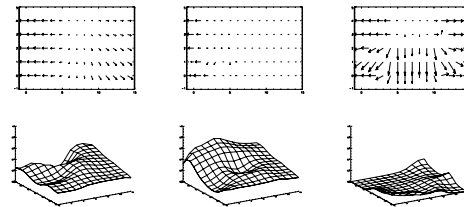


Abbildung 4: Typische, im realen Experimentalfeld gewonnene optische Flußbilder (oben) mit zugehörigen erlernten Aktionsbewertungskarten (unten). Die Bewertung einer Aktion (z -Achse) in einer Situation hängt von deren Geschwindigkeit (y -Achse) und deren Lenkwinkel (x -Achse) ab und repräsentiert ihren Wert (action value) bezüglich der zu realisierenden Hindernisvermeidung. Hohe Werte entsprechen hohen Bewertungen und umgekehrt. Im linken Bildbereich des ersten Flußbildes verursacht ein Hindernis im linken Bereich entsprechend große Verschiebungsvektoren. Das Modul *Action Suggestion* hat gelernt, daß in dieser Situation Linkswendungen zu Kollisionen führen können (erkennbar am Tal in der Bewertungskarte unten links).

Das darin repräsentierte Bewertungssig-

nal ergibt sich aus dem skalaren Bewertungssignal r aus der Umwelt bei realer Ausführung der jeweiligen Aktion \underline{A}^{MM} (immediate reinforcement). Führt eine Aktion zu einer Kollision, so wird ein reward von $r = 0.0$ geliefert, ansonsten beträgt das reward $r = 1.0$. Während des Lernens werden nur die Motorkarten w_i^{MM} der Clusterneuronen i adaptiert, welche an der Generierung des zugrundeliegenden Aktionsvorschlages beteiligt waren, indem die normierte Wichtung der Superposition y_i^{IP} genutzt wird. Die Adaptionvorschrift entspricht einer modifizierten Delta-Lernregel, wie sie bei schwachen Reinforcementverfahren angewendet wird (Gl. 1).

$$\Delta w_{ir}^{MM} = \eta \cdot y_i^{IP} \cdot a_r^{MM} (r - w_{ir}^{MM}) \quad (1)$$

Die normierte Superpositionswichtung y_i^{IP} garantiert situationsspezifisches, und der Faktor a_r^{MM} realisiert aktionsspezifisches Lernen. Die *Action Suggestion* erlernt somit, in welchen Situationen welche Aktionen zu Kollisionen führen, indem es diesen gefährlichen Aktionen kleine Bewertungen zuweist (Abb. 4).

2.4 Aktionsauswahl

In der Motorkarte sind die Bewertungen aller möglichen Aktionen enthalten. Um im Sinne des antizipativen Wahrnehmungskonzeptes auch alternative Aktionen untersuchen zu können, ist es erforderlich, sequentiell mehrere Aktionen aus dieser Motorkarte zu selektieren, wobei mit der am besten bewerteten Aktion begonnen werden soll.

Die in [15, 5] beschriebene neuronale Feld-dynamik selektiert sequentiell einzelne Aktionen aus dem Bewertungsgebirge, beginnend mit der am besten bewerteten. Auf diese Weise werden zuerst die erfolgversprechendsten Aktionen untersucht und, je nach verfügbarer Simulationszeit, im weiteren Verlauf auch weniger gut bewertete, berücksichtigt (Abb. 5).

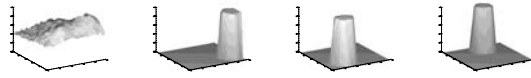


Abbildung 5: *Beispiel der erweiterten Amari-Dynamik. Ausgehend von der stabilen Eingabe (1, v.l.n.r.) werden zeitlich nacheinander die Ausgaben (2) bis (4) generiert. Diese bleiben jeweils eine bestimmte Zeit aktiv und zerfallen anschließend wieder.*

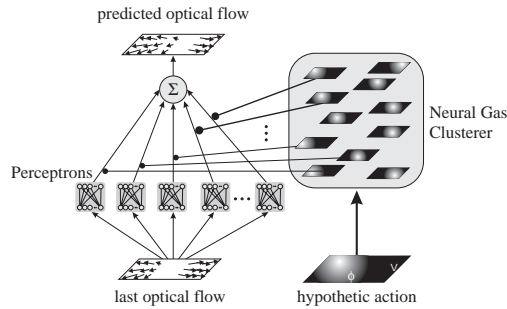


Abbildung 6: *Neuronale Architektur des „Mixture of Experts“-Ansatzes mit Perceptrons und Gating-Netzwerk.*

2.5 Prädiktion sensorischer Konsequenzen

Die selektierte Aktion wird mit dem aktuellen Flußbild im Modul *Optical Flow Prediction* verknüpft, um die daraus resultierende sensorische Folgesituation zu schätzen. Die Performanz dieser Prädiktion ist von entscheidender Bedeutung für die Leistungsfähigkeit des Gesamtsystems, so daß dieses Modul bisher in zwei verschiedenen Architekturen realisiert wurde, welche im folgenden vergleichend untersucht werden sollen.

Modifizierter „Mixture of Experts“-Ansatz mit Perceptrons: Da der Folgefluß primär vom aktuellen Flußfeld abhängt und je nach ausgeführter Aktion verschoben oder skaliert wird, kann eine einfache aber aktionsspezifische Transformation das aktuelle auf das zugehörige folgende Flußbild abbilden. Dazu wird eine Architektur aus lokalen Experten, ähnlich dem „mixture of experts“-Ansatz

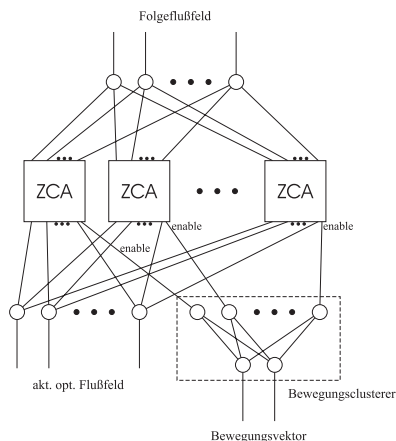


Abbildung 7: Neuronale Architektur des Ansatzes mit hierarchischem Clusterer.

[9], genutzt. Im Gegensatz zu [9] werden die Ausgaben der Experten durch eine Superposition fusioniert, es findet also keine harte Expertenauswahl statt. Jeder lokale Experte wird durch ein Perceptron realisiert, ist auf eine Menge ähnlicher Aktionen spezialisiert und erlernt die zugehörige Abbildungsfunktion. Das Gating der Experten wird von einem Selektor-Netzwerk durchgeführt, welches nach dem Neural-Gas-Ansatz nach [11] arbeitet. Die Ausgaben der lokalen Experten werden durch eine vom Selektor-Netzwerk gewichtete normierte Superposition in die finale Prädiktion überführt (Abb. 6).

Zum Training der Perceptronwichtungen W_i^P werden die sensorischen Eindrücke jeweils vor (\underline{x}^{OF}) und nach einer real ausgeführten Aktion ($\underline{x}_{target}^{OF}$) genutzt (Gl. 2). Die Wichtungsänderung ΔW_i^P wird je nach Wert des Superpositionsgewichtes y_i^{IP} und der aktuellen Lernrate η^P gemäß der Delta-Lernregel berechnet.

$$\Delta W_i^P = \eta^P \cdot y_i^{IP} \cdot (\underline{x}_{target}^{OF} - W_i^P \cdot \underline{x}^{OF}) \cdot \underline{x}^{OF T} \quad (2)$$

Hierarchische Clusterer: Dieser Ansatz nutzt eine hierarchische Verschaltung mehrerer, hinsichtlich online-Lernfähigkeit spezialisierter Clusterer nach [3].

Auch diese Architektur arbeitet ähnlich einem „mixture of experts“-Ansatz, wobei hier zur Generierung der eigentlichen Ausgabe keine Perceptrons, sondern Zustandsclusterer mit einer Ausgabeschicht (ZCA) eingesetzt werden. Diese operieren auf dem vollen Flußfeld und generieren aktionsspezifisch das entsprechende Folgeflußbild. Das aktionsspezifische Gating der Experten wird vom Bewegungsclusterer durch gewichtete normierte Superposition der Expertenausgaben realisiert (Abb. 7).

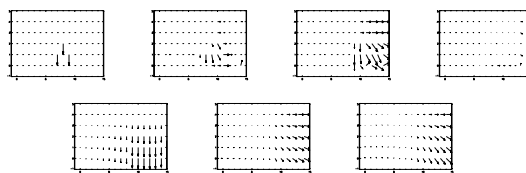


Abbildung 8: Sequenz von realen (oben) und prädizierten Flußsituationen (unten). In der ersten Situation ist im zentralen Bildbereich ein Hindernis erkennbar, welches sich während der Vorwärtsbewegung mit leichter Linksorientierung mehr und mehr in den rechten Bildbereich verschiebt.

Auch bei diesem Ansatz erfolgt das Training anhand des aktuellen und letzten realen Flußfeldes nach realer Ausführung einer Aktion. Ist der Prädiktionsfehler größer als eine definierte Schwelle, so werden entweder neue Clusterzentren eingefügt (incrementeller Charakter) oder bereits vorhandene Clusterzentren umfunktioniert (lifelong-Lernen). Letzteres wird derart realisiert, daß an solchen Positionen im Merkmalsraum mit relativ geringem Prädiktionsfehler Clusterneuronen entfernt werden, um sie an der Position der noch unzureichenden Vorhersage einzufügen. Auf diese Weise wird versucht, mit limitierten Ressourcen die bestmögliche Prädiktion zu realisieren.

Ergebnisse: In Abb. 8 ist eine Beispielsequenz von realen und prädizierten Folgeflußbildern, welche im verwendeten Kheperaszenario erlernt wurden, dargestellt.

2.6 Hypothesenbewertung

Die Bewertung von Aktionshypothesen *Hypothesis Evaluation* basiert auf der Motorkarte der *Action Suggestion* und liefert eine skalare Bewertung für die jeweils selektierte hypothetische Aktion. Dazu wird die mittlere Bewertung in der zugrundeliegenden Motorkarte an der Position des selektierten Aktionsblobs bestimmt.

2.7 Verkettung von *SSMHP*'s

Die Verkopplung mehrerer *SSMHP*s (Abb. 1) gestattet nun die interne Simulation von mehreren sequentiell ausgeführten Aktionen, so daß ein entsprechend vorausschauendes Verhalten realisiert werden kann. Weiterführende Betrachtungen dazu werden in [6, 7] vorgestellt.

3 Ergebnisse

In Abbildung 9 ist ein zeitlicher Ausschnitt aus einem intern simulierten Suchbaum für eine reale sensorische Situation dargestellt. Der Roboter befand sich hierbei in einer Situation vor einem frontalen Hindernis, so daß er noch kollisionsfrei eine Geradeausfahrt durchführen könnte. Diese Geradeausfahrt erfordert allerdings im nächsten Schritt drastische Motorkommandos, um eine Kollision zu vermeiden. Eine Alternative stellt ein frühzeitiges Ausweichmanöver dar, weil auf diese Weise einerseits keine extremen und energiezehrenden Motorkommandos benötigt werden und andererseits ein größerer Sicherheitsabstand zum Hindernis gewahrt und somit die Kollisionswahrscheinlichkeit verringert wird.

Obiges Beispiel zeigt die vorausschauende Arbeitsweise des antizipierenden Navigationsansatzes. Ein antizipierendes System bewertet kurzfristige hohe Bewertungen im Gegensatz zum rein reaktiven Ansatz nicht über, sondern bezieht Bewertungen eines

größeren Zeithorizontes in seine Aktionsauswahl ein. Auf diese Weise wird das Hindernis schon früher **wahrgenommen**, als beispielsweise bei einem reaktiven Navigationsansatz (Abb. 10).



Abbildung 10: Vergleich des reaktiven (gepunktete Linie) mit dem antizipierenden Ansatz (durchgezogene Linie).

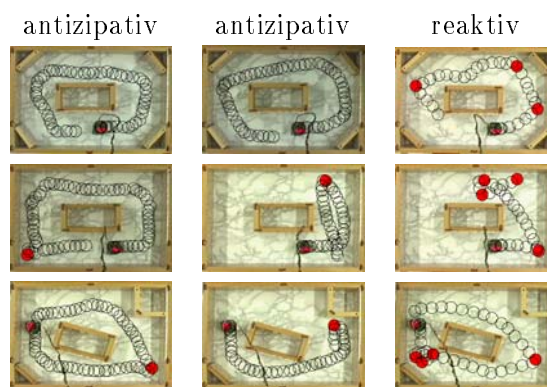


Abbildung 11: Vergleich der Fahrtrajektorien des KHEPERA in realen Umgebungen verschiedener Komplexität (zeilenweise von oben nach unten wachsend) mit verschiedenen Navigationsansätzen. Dazu wurden das antizipative Gesamtsystem mit hierarchischem Clusterer (links) oder dem „Mixture-of-Experts“-Ansatz (mitte) als Prädiktor, bzw. ein rein reaktiver Ansatz (rechts) unter Nutzung identischer Action Suggestion-Module eingesetzt. Kollisionen sind durch rot gefüllte Kreise gekennzeichnet.

In Abbildung 11 werden Fahrtrajektorien des KHEPERA in verschiedenen Umgebungen vorgestellt, die durch drei unterschiedliche Navigationssysteme erzeugt wurden. Auf Verhaltensebene sind kaum Unterschie-

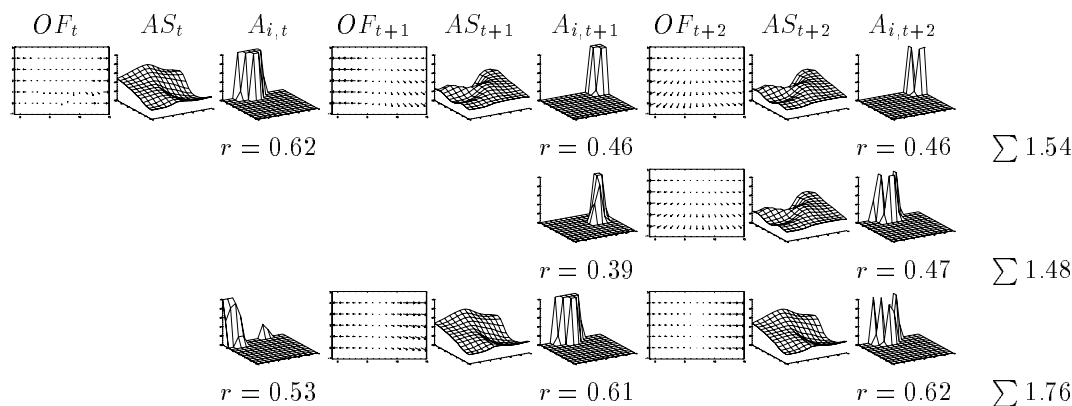


Abbildung 9: Ausschnitt aus einem simulierten Suchbaum mit sensorischen Situationen OF , Aktionsvorschlägen AS und ausgewählten Aktionen A . Die zuerst untersuchte Aktionssequenz (obere Zeile) schlägt die noch mögliche Geradeausfahrt vor, um danach durch extreme Lenkmanöver dem Hindernis auszuweichen. Die dritte Zeile zeigt eine Aktionssequenz, die von einer Linksbewegung schon im Initialzustand ausgeht, deshalb keine extremen Ausweichmanöver benötigt und höher bewertet wird.

de zwischen den zwei vorgestellten Prädiktionsansätzen erkennbar, beide erlauben eine vorausschauende Hindernisvermeidung.

Es ist jedoch klar erkennbar, daß der reaktive Ansatz (jeweils rechts) deutlich mehr Kollisionen verursacht, da ausschließlich die Bewertung der nächsten Aktion berücksichtigt wird. So wird erst ein Ausweichmanöver eingeleitet, wenn eine Kollision unmittelbar bevorsteht. Für reale mobile Systeme existiert dann aber aufgrund physikalischer Gegebenheiten (eingeschränktes Aktionsspektrum) oftmals keine Aktionsfolge, die eine Kollision sicher vermeidet. Der in diesem Beitrag vorgestellte antizipative Navigationsansatz, welcher durch Vorhersage und Bewertung sensorischer Konsequenzen von hypothetisch ausgeführten Aktionen eine vorausschauende Verhaltensweise selbstorganisierend entwickelt, kann in solchen Situationen deutlich sicherer agieren.

4 Ausblick

Ein vorausschauendes Verhalten realisieren neben dem vorgestellten antizipierenden Ansatz auch starke Reinforcementverfah-

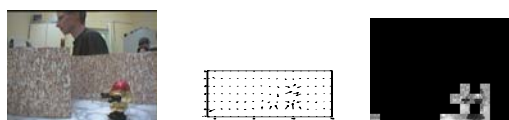


Abbildung 12: In der Situation (links) im Labyrinth wurde die Differenz zwischen Prädiktion und realem Folgeflußfeld (mitte) ermittelt. Die Region großer Prädiktionsfehler korrespondiert mit dem dynamischen Objekt im Bild, welches somit detektiert werden kann (rechts).

ren, wie beispielsweise das Q-Learning [20] oder TD- (λ) -Verfahren [1]. Deshalb sollen in weiteren Untersuchungen systematische Vergleiche mit diesen Ansätzen hinsichtlich der online-Lernfähigkeit, Plastizität, Stabilität usw. durchgeführt werden.

Ein weiterer interessanter Aspekt ist die Nutzung der Prädiktionsfehler zur Detektion dynamischer Hindernisse, weil diese durch ihre Eigenbewegung keinen allgemeingültigen und vom System erlernbaren sensomotorischen Zusammenhang darstellen (Abb. 12).

In nachfolgenden Untersuchungen soll die

Eignung weiterer inkrementeller lifelong-lernfähiger Netzwerke zur Flußprädiktion und Aktionsbewertung untersucht werden. Es ist dazu vorgesehen, ein am Fachgebiet entwickeltes lebenslanges inkrementelles Lernverfahren (LLGNG) [8] auf seine Eignung zu untersuchen.

Literatur

- [1] J. Bruske, I. Ahrns, and G. Sommer. Practicing q-learning. *Proceedings of ESANN'96, Bruges*, pages 25–30, 1996.
- [2] Thomas Christaller. Probehandeln als ein imaginativer Prozeß. *Tagungsband des Workshops Kognitive Robotik*. Zentrum für Kognitionswissenschaften, Universität Bremen, 1995. ZKW bericht Nr. 3/95.
- [3] M. Draber. Vorhersage sensorischer Folgesituationen mit neuronalen Netzen. *Diplomarbeit, Fachgebiet Neuroinformatik, Technische Universität Ilmenau*, 1998.
- [4] A.P. Duchon, W.H. Warren, and L.P. Kaelbling. Ecological robotics: Controlling behavior with optical flow. 1995.
- [5] H.-M. Gross, V. Stephan, and M. Krabbes. A neural field approach to topological reinforcement learning in continuous action spaces. *to appear in Proceedings of WCCI-IJCNN'98, Anchorage, 1998*, 1998.
- [6] H.-M. Gross, V. Stephan, and T. Seiler. Neural architecture for sensorimotor anticipation. *In Proceedings of EMCSR'98, Vienna*, pages 593–598, 1998.
- [7] H.-M. Gross, V. Stephan, and T. Seiler. Perception and Sensorimotor Control - as Internal Simulation and Anticipation Process. *In Proceedings of Dynamische Perzeption, Bielefeld*, 1998.
- [8] F. Hamker and H.-M. Gross. A lifelong learning approach for incremental neural networks. *In Proceedings of EMCSR'98, Wien, 1998*, pages 599–604, 1998.
- [9] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton. Adaptive mixtures of local experts. *Neural Computation* 3, 1991, pages 79–87, 1991.
- [10] S. M. Kosslyn. Aspects of a cognitive neuroscience of mental imagery. In Stephen M. Kosslyn and Richard A. Andersen, editors, *Frontiers in Cognitive Neuroscience*, chapter 47, pages 603–608. A Bradford Book, The MIT Press, 1995.
- [11] Thomas Martinetz and Klaus Schulten. A “neural gas” network learns topologies. In Teuvo Kohonen, Kai Mäkisara, Olli Simula, and Jari Kangas, editors, *Artificial Neural Networks*, pages 397–402. Elsevier Amsterdam, 1991.
- [12] R.C. Miall and D.M. Wolpert. Forward models for physiological motor control. *Neural Networks, Vol. 9, No. 8, Elsevier Science Ltd, 1996*, pages 1265–1279, 1996.
- [13] 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.
- [14] Werner von Seelen et al. *A Neural Architecture for Autonomous Visually Guided Robots — Results of the NAMOS Project —*. Düsseldorf: VDI Verlag, 1995.
- [15] T. Seiler, V. Stephan, A. Heinze, and H.-M. Gross. Handlungsauswahl durch Antizipation sensomotorischer Konsequenzen. *Proceedings of SOAVE'97, Ilmenau, 1997*, 1997.
- [16] Richard S. Sutton. Td models: Modeling the world at a mixture of time scales. *In Proceedings of the 12th Int. Conf. on Machine Learning*. Morgan Kaufmann, 1995.
- [17] R.S. Sutton. Integrated architectures for learning, planning and reacting based on approximating dynamic programming. *Proceedings of the 7. International Conference on Machine Learning*, pages 216–224, 1990.
- [18] J. Tani. Visual attention and learning of a cognitive robot. *Proceedings of ICANN'97, Lausanne, October 1997, Switzerland, Springer Verlag*, pages 697–702, 1997.
- [19] J. Tani, J. Yamamoto, and H. Nishi. Dynamical interactions between learning, visual attention, and behavior: An experiment with a vision-based mobile robot. *Proceedings of ECAL'97, MIT-Press, 1997*, pages 309–317, 1997.
- [20] Ch. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8:279–292, 1992.