Handlungsorganisation intentionaler neuronaler Agenten*

in: Proc. SOAVE'97 - Selbstorganisation von adaptivem Verhalten, Ilmenau September 1997, Fortschrittberichte VDI, Reihe 8, pp. 35-44, VDI-Verlag 1997

M. Krabbes, H.–J. Böhme, V. Stephan, H.–M. Gross

Technische Universität Ilmenau
Fachgebiet Neuroinformatik
D-98684 Ilmenau, Postfach 100565
Markus.Krabbes@Informatik.TU-Ilmenau.DE

Zusammenfassung

Es wird ein neuronales Agentensystem vorgestellt, mit dem auf Basis eines lokalen Handlungsrepertoirs in enger Interaktion mit einem Anwender komplexe Verhaltensleistungen erreicht werden können. Grundlage dafür bildet eine geeignete Problemdekomposition in einer heterarchischen Struktur, mit der es gelingt, den gesamten situationsspezifischen Aktionsraum selektierbar zu repräsentieren. Die Realisierung der lokale Handlungsziele verfolgenden Agenten zur Navigation eines mobilen Roboters wird ausführlich vorgestellt. Grundlage dieser intentionalen Agenten bilden überwacht vortrainierte sensomotorische Mappings. Die resultierenden verschiedenen Handlungsvorschläge werden basierend auf ihrem energetischen und zeitlichen Kontext flexibel fusioniert.

1 Einführung und Szenario

Ziel des Projektes GESTIK ist die Entwicklung einer neuronal basierten Steuerarchitektur, mit der sich ein mobiler Roboter unter „Blickkontakt“ zum Anwender entsprechend dessen gestenbasierten Anweisungen teilautonom bewegt. Sich auf der Fahrtroute befindende statische und dynamische Hindernisse sollen dabei geeignet umfahren werden. Dem menschlichen Interaktionspartner kommen in diesem Szenario folgende Funktionen zu:

1. Der Anwender ist der primäre Handlungsantrieb für das System, welches ständig versuchen soll, sich diesem auf eine definierte Entfernung anzunähern, solange er dies nicht durch eine entsprechende (gestenbasierte) Instruktion unterbindet.

2. Situationen, in denen sich dem System mehrere Handlungsalternativen eröffnen, sind vom Anwender durch eine eindeutige Anweisung aufzulösen. Voraussetzung dafür ist, daß entsprechende Konfliktsituationen erkannt werden und sich der Handlungsvorschlag geeignet in den Entscheidungsprozeß einbinden läßt.

3. Wegen der rein visuell basierten Navigation ist nicht auszuschließen, daß das Fahrzeug mit nicht eindeutig interpretierbaren sensorischen Situationen operieren muß. Hier soll der Anwender assistierend eingreifen, um falsche Hypothesen zu korrigieren.

Signifikant für das angestrebte Leistungsvemögen des Agentensystems ist, daß mit diesem lediglich lokales Verhalten erreicht weden kann, da sich die reale Umgebung im visuellen Eindruck der Navigationskamera des Systems niemals eindeutig bezüglich der globalen Position im Einsatzfeld darstellt bzw. darstellen soll (keine eindeutigen Landmarken). Dies stellt im Kontext des Projektziels keine Einschränkung dar, da sich im Zusammenwirken von Anwender und Fahrzeug die erwünschte Systemleistung vollständig und widerspruchsfrei erreichen läßt. Auf diese Weise wird es vielmehr möglich, die lokalen Verhaltensleistungen konsequent zu entwickeln und gleichzeitig demonstrabel

*Die Arbeiten sind Teil des vom Thüringer Ministerium für Wissenschaft, Forschung und Kultur geförderten Projektes GESTIK.

umzusetzen, weil eine globale Bewegungsplanungsebene zunächst völlig entkoppelt ist, sich aber konzeptionell problemlos ergänzen läßt.

Zielsystem ist die Roboterplattform MILVA (`http://cortex.informatik.tu-ilmenau.de/technik.html#robbi`), die mit einem triocularen Visionsystem und on-Board–PC-Rechentechnik ausgerüstet ist. Zwei auf einer horizontal schwenkbaren Traverse montierte Kameras mit je 3 Freiheitsgraden (Pan, Tilt, Zoom) dienen der Bildaufnahme zur gestenbasierten Nutzer-Roboter-Kommunikation, eine zentral montierte Weitwinkel-Kamera liefert den ausschließlich für die Navigation genutzten visuellen Datenstrom. Als Experimentalplattform der hier vorgestellten Untersuchungen dient gegenwärtig der Miniaturroboter KHEPERA (kreisrund, $ø = 5,5cm$; mittige Farbkamera), dessen praktischer Einsatz sich ausgesprochen unproblematisch gestaltet und so Simulationen gut ersetzen kann. Aufgrund einer definierten Labyrinthumgebung ergeben sich keine Probleme in der Untergrund-Hindernis-Trennung, andererseits sind die in der „Real World“ auftretenden Schwankungen und Rauscheinflüsse sowohl für das Netzwerktraining als auch die Validierung der erworbenen Verhaltensleistungen unbedingt erforderlich, um zu robusten Ergebnissen zu gelangen. Diese sollen später auf die Zielplattform MILVA übertragen werden, wobei hierfür noch die zusätzliche Problematik der stark variierenden optischen Erscheinung von Indoor-Umgebungen zu lösen ist.

2 Problemdekomposition und Repräsentation von Handlungsantrieben

Wie bereits angedeutet, sind die folgenden beiden signifikanten Eigenschaften notwendig, um zu einer interaktiven Verhaltensleistung zu gelangen:

- Die sich in einer bestimmten Situation bietenden Handlungsalternativen müssen voneinander separierbar sein. Hierbei ist einerseits die Erkennung von sich gleichzeitig bietenden Handlungsoptionen für das System notwendig, um u.U. eine Anfrage an den Anwender stellen zu können und so Konflikte aufzulösen. Andererseits wird damit erst die Ausschöpfung der gesamten sich bietenden Aktionsvielfalt gewährleistet.
- Jede der einzelnen Handlungsalternativen muß sich in einer Weise unterstützen lassen, in der sowohl die Entscheidung des Anwenders bzw. einer hierarchisch höher liegenden Verarbeitungsebene einbindbar ist, als auch weiterhin das lokale Wissen der Struktur erhalten bleibt und genutzt werden kann.

Zur Umsetzung dieser Spezifika bietet sich eine Dekomposition in prinzipiell identisch arbeitende, aber verschiedene Handlungsziele verfolgende Agenten an, deren Handlungsvorschläge so die gesamte situationsgemäße Aktionsvielfalt repräsentieren.

2.1 Implizite Repräsentation von Handlungsantrieben in intentionalen Agenten

Es bietet sich an, menschliche Problemlösungsstrategien und deren Teilaspekte auf neuronale Agenten abzubilden, die jeweils versuchen, diese Teilaspekte optimal zu bearbeiten. Daher erhält jeder Agent im Rahmen des Gesamtproblems eine Zweckbestimmung, die über einen agentenspezifischen Handlungsantrieb, eine Intention, beschrieben wird. Während in einigen Lernverfahren die Handlungsantriebe über entsprechende Bewertungsfunktionen ausgedrückt werden (z.B. Reinforcement-Lernen), besteht bei überwachten Lernverfahren die Möglichkeit, durch „Vormachen“ einer charakteristischen Problemlösungsstrategie eine implizite Beschreibung der Systemziele zu erzeugen. Dies wird in der Literatur als Experten-Cloning bezeichnet.

Für die Problematik eines lokalen Navigationsverhaltens mit den Teilaspekten gerichtete schnelle Lokomotion, Vermeidung statischer und dynamischer Hindernisse, eingeschränkte Exploration usw. bietet sich an, separate Agenten zu realisieren, die in ihrem Wechselspiel

- Flurbereiche möglichst mittig durchfahren,
- sich bietende Abbiegemöglichkeiten nutzen (Explorationsverhalten, z.B. Nutzersuche),
- Kreuzungsbereiche möglichst geradlinig überfahren (im Sinne eines Beharrungsvermögens) oder
- in Gefahrenbereichen anhalten bzw. umkehren.

In der Realisierung solcher Agentenausrichtungen durch entsprechende Cloning-Strategien zeigt sich ein grundsätzlicher Unterschied zu den Reinforcement-Methoden, bei denen durch entsprechende Bewertungsfunktionen eine Spezialisierung der Agenten auf oben genannte Verhaltensleistungen erreicht wird. Dies ist beim überwachten Agententraining in dieser Weise nicht möglich, weil den Agenten zu **allen** sensorischen Situationen eine situations- und intentionsadäquate Systemantwort präsentiert werden muß. Deshalb manifestiert sich die Realisierung der oben aufgeführten Teilaspekte nicht in einer Trainingsstrukturierung, die das Gesamtproblem in solche Teilaspekte (Situationen) zerlegt, sondern vielmehr in der Definition von Handlungskonzepten, nach denen der Experte beim Vormachen unter mehreren situationsbezogen optionalen Aktionen in konsistenter Weise eine bestimmte auswählt. Das Spektrum dieser festzulegenden Handlungsregeln muß nun für alle auftretenden Teilprobleme adäquates Verhalten ermöglichen. Für die Roboternavigation besteht eine, wenn auch triviale, Möglichkeit der Definition solcher Handlungskonzepte in der Trennung: „Fahre so weit wie möglich...

1. und nutze die nächste sich bietende Gelegenheit, nach rechts abzubiegen!“
2. und nutze die nächste sich bietende Gelegenheit, nach links abzubiegen!“
3. geradeaus und biege nur ab, wenn keine Alternative besteht!“

Diese Einzeldefinitionen sind zwar in jeder Situation anwendbar, ermöglichen jede für sich aber noch keine Bewältigung der Gesamtproblematik. Im Zusammenwirken dieser Handlungsziele wird dagegen ein komplexes Gesamtverhalten erreicht, mit dem nicht nur die verschiedenen Teilaspekte der Roboternavigation abgedeckt werden, sondern das auch die oben genannten Voraussetzungen für eine interaktive Systemleistung erfüllt. So wird beispielsweise bei einem auftretenden Hindernis der Agent „3“ versuchen, bis kurz vor einer Kollision darauf zuzufahren. Die beiden anderen Handlungsziele repräsentieren dagegen ständig alle Möglichkeiten, von diesem Pfad abzuweichen. Spätestens wenn auch der „Geradeaus-Agent“ zu einer Ausweichbewegung gezwungen ist, wird über den „Gewinner“ der beiden „Abbiege-Agenten“ die günstigere Alternative herausgebildet. Dabei existieren ständig Eingriffsmöglichkeiten durch eine höhere Verarbeitungsebene, indem einzelne der optionalen Aktionen unterstützt werden.

2.2 Handlungsorganisation durch Agentenfusion

Wie bereits angedeutet werden die einzelnen Agenten nun in vielen Situationen entsprechend ihrer Intention (trainierten Handlungsziele) kontroverse Handlungsvorschläge generieren, die situationsspezifisch in einer Weise zu fusionieren sind, die die Aspekte des globalen Fahrverhaltens (z.B. Fahrt zum Nutzer oder entsprechend Nutzeranweisung) berücksichtigt. Ein Strukturentwurf zur Zusammenfassung der Handlungsvorschläge solcher Einzelagenten ist in Bild 1 dargestellt: Expertenwissen fließt dabei auf 3 Kanälen in das Gesamtsystem ein:

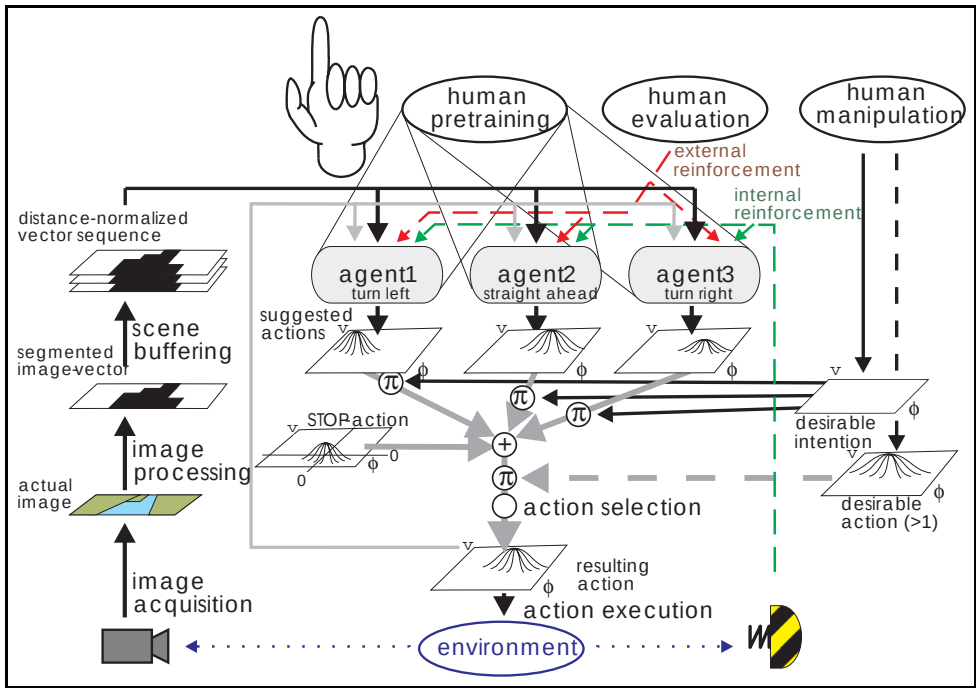

Abbildung 1: *Gesamtübersicht der verteilten Agentenstruktur.*
(Siehe Text für detaillierte Beschreibung.)

1. Die verschiedenen Agenten bilden den visuellen Datenstrom (distance-normalized vector sequence) auf situationsadäquate Aktionen (suggested actions) ab, die in zweidimensionalen Karten durch Geschwindigkeit v und Lenkwinkel ϕ topologisch kodiert sind. Sie erhalten, wie bereits erläutert, ihre intentionale Ausrichtung durch Belehrung mit unterschiedlichen, von einem Experten erstellten Trainingsmengen. (human pretraining, siehe Abschnitt 3.)

2. Die Aktionsvorschläge der über identischen Datenströmen operierenden Agenten werden überlagert (+) und können sowohl intentionsorientiert, also agentenspezifisch (desirable intention), als auch aktionsorientiert (desirable action) durch Anweisungen des Nutzers verstärkt werden (human manipulation). Weil beispielsweise bei Abbiegemanövern u.U. zunächst in die entgegengesetzte Richtung ausgeholt wird, muß dabei der handlungszielorientierte (agentenspezifische) Nutzereingriff die primäre Anweisungseinbindung darstellen. Dieser Eingriff findet zwar bereits auf einer symbolischen Ebene statt, basiert aber dabei auf den vom Experten während des Agententrainings verwendeten Begriffen. Durch den rein modulierenden Eingriff kann sichergestellt werden, daß das Fahrzeug zwar diese externen Anweisungen befolgt, aber dabei immer nur mit dem eigenen situationsgemäßen Verhaltensrepertoire agiert. Die Einkopplung der vom Anwender ausgehenden Anweisungen wird konzeptionell so realisiert, daß über ein Gesten(Posen)-Alphabet, mit dem eine visuelle Nutzer-Roboter-Kommunikation geführt werden soll, eine Interpretation zu einer entsprechenden Agenten- bzw. Aktionsunterstützung möglich wird (siehe hierzu [BBB 97]).

3. Zur weiteren Adaption der Agenten, die im Abschnitt 5 konzeptionell vorgestellt wird, muß die auf Basis des energetischen und zeitlichen Kontextes letztendlich ausgewählte Aktion (action selection) auf die Agenten zurückgeführt werden, um diese dem erzielten Handlungserfolg zuzuordnen. Basis für die Bewertung ist die sensorische Erfahrung von Kollisionen, die als Reinforcement-Signal (internal reinforcement) dazu führen soll, daß das Gesamtsystem in einer späteren vergleichbaren Situation eine andere (bessere) Aktion ausführt. Das Modul human evaluation verkörpert die Handlungsbewertung durch einen Experten über ein external reinforcement. Die Identifikation dieser Kollisionen o.ä. vorwegnehmenden Bewertungssignale aus einem multimodalen Datenstrom (visuell, akustisch) mittels Konditionierung ist Gegenstand weiterer Forschungsarbeiten am Fachgebiet.

Als Ergebnis der agenteninternen Adaptionsvorgänge kann es, wie auch in unbekannten Situationen, dazu kommen, daß ein oder mehrere Agenten keinen Aktionsvorschlag liefern. Für diese Fälle wird vergleichbar einer permanenten Besorgnis ständig eine unterschwellige Stop-Absicht (STOP-action) den Aktionsvorschlägen überlagert, um sich dann mit der Anhalteaktion (v=0, ϕ=0) durchzusetzen.

3 Realisierung der intentionalen Agenten

3.1 Szenariospezifische Umsetzung des ALVINN-Ansatzes auf dem KHEPERA

Als Ausgangspunkt für einen überwacht trainierten Einzelagenten einer derartigen verteilten Steuerarchitektur bietet sich der ALVINN–Ansatz [Pom93] an, dessen szenariospezifische Umsetzung und Erweiterung in diesem Abschnitt betrachtet wird. Die Grundidee von ALVINN besteht in einer direkten Abbildung eines Kamerabildes auf einen Lenkwinkel, um damit ein Straßenfahrzeug auf gewöhnlichen Straßen zu steuern. Dabei dient ein Multi-Layer-Perceptron als Basisstruktur, die das von einem Trainer vorgegebene Verhaltensmuster approximiert und dabei den analogen Wert des einzustellenden Lenkwinkels in Form einer topologischen Ausgabekodierung repräsentiert (siehe Abb. 3).

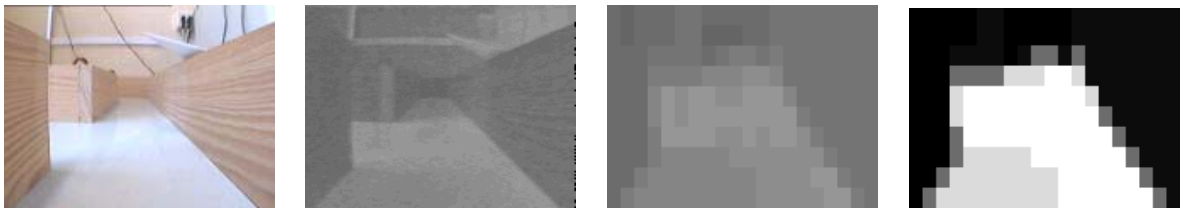

Abbildung 2: *Die Phasen der Bildvorverarbeitung auf dem KHEPERA: das Originalbild (links), die Blau-Gelb-Aktivierungen dieses Bildes nach [PG96] (2. v. links), nach der 3–fachen Unterabtastung des relevanten Bildausschnitts (2. v. rechts) und der Dynamikanpassung (rechts).*

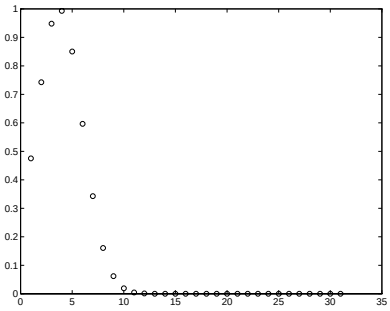

Abbildung 3: *Eindimensionale topologische Kodierung des Lenkwinkels über einem m-dimensionalen Outputvektor: Da es beim KHEPERA mit seinen beiden getriebenen Rädern keine explizite Steuergröße „Lenkwinkel“ gibt, muß diese aus den Geschwindigkeiten der beiden Räder ermittelt werden (LS=LeftSpeed,RS=RightSpeed):*

$$y_i = e^{-\frac{\left(i-\frac{m+1}{2}\cdot(1+\frac{LS-RS}{LS+RS})\right)^2}{M}} \qquad i \in [1\ldots m];$$

Darstellung für $m = 31, M = 10, RS = 5 \cdot LS$

Für die Steuerarchitektur des KHEPERA dient ebenfalls ein zweilagiges Multi-Layer-Perceptron in reiner Feed-Forward-Architektur als Netzwerk, zu dessen Belehrung klassisches Backpropagation eingesetzt wird. Die Vorverarbeitung der Input- und Outputdaten ist in den Abbildungen 2 und 3 dargestellt. Die Trainingsmenge bildet eine 2500 Beispielsituationen umfassende Datenbank, die von einem „Experten“ erstellt wurde und in etwa 50 Trainingsepochen präsentiert wird. An dieser Stelle muß betont werden, daß die Navigation durch den Experten bei der Erstellung der Trainingsdaten ausschließlich mittels der Videobilder der on-Board-Kamera des KHEPERA erfolgte, um konsistente Datenquellen in Trainings- und Recallphase zu garantieren.

Es ist anzumerken, daß ein objektiver Benchmarktest zum Vergleich verschiedener Trainingsdatensätze, Belehrungsdauern, Netzwerkkonfigurationen usw. noch nicht realisiert ist. Die visuelle

Aufzeichnung der KHEPERA-Fahrtrajektorie in den Abbildungen 4 und 7 ist ein erstes Ergebnis dahingehender Arbeiten.

3.2 Probleme / Erweiterungen der ALVINN-Architektur

In dieser ersten Implementierungsform meistert der KHEPERA alle gutartigen Situationen auch in einem unbekannten Labyrinth weitgehend kollisionsfrei und zeigt dabei ein erwartungsgemäßes (dem Training entsprechendes) Verhalten (Abb. 4). Wichtig ist dabei, daß immer Elemente des Untergrunds im Blickbereich der Kamera bleiben. Hierbei macht sich das schmale Blickfeld der Kamera stark einschränkend bemerkbar (Abb. 5).

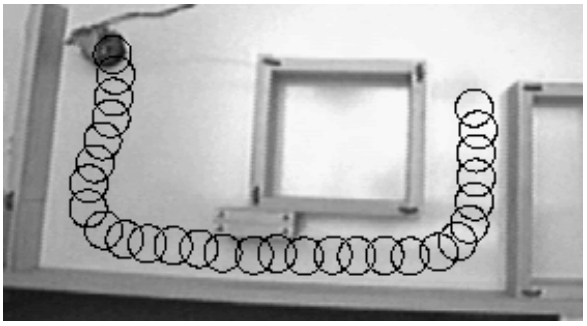

Abbildung 4: *Aufzeichnung einer KHEPERA-Fahrt mittels einer farbigen Markierung (Kreis = KHEPERA-Grundfläche, Bild = Startposition): Das mit der ALVINN-Architektur erreichbare (weil trainierbare) Fahrverhalten: In Gassen richtet sich das Fahrzeug mittig aus, ist nur eine Wand sichtbar, so wird dieser mit einem mittiger Flurfahrt entsprechendem seitlichen Abstand gefolgt.*

3.2.1 Wegfenster

Da dieses Problem bei der MILVA-Roboterplattform mit ihrer 3–Rad-Kinematik auch wegen des Schleppkurvenverhaltens in ähnlicher Weise auftritt (Abb. 5), müssen bereits im KHEPERA-Szenario Strukturen entwickelt werden, die auch hier durch eine sensomotorische Abbildung ein adäquates Verhalten erzeugen. Dabei sollen Architekturen gefunden werden, die die zurückliegende Information bis zu einem festen, den Fahrzeugabmessungen entsprechenden **Weg**-Horizont gleichberechtigt berücksichtigen.

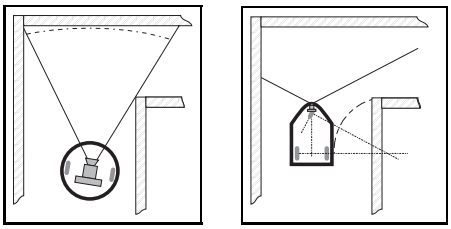

Abbildung 5: *Der schmale Sichtbereich der KHEPERA-Kamera (Öffnungswinkel 60^o) schränkt die erreichbare Verhaltensleistung erheblich ein (links). Vergleichbare Verhältnisse bei MILVA (rechts).*

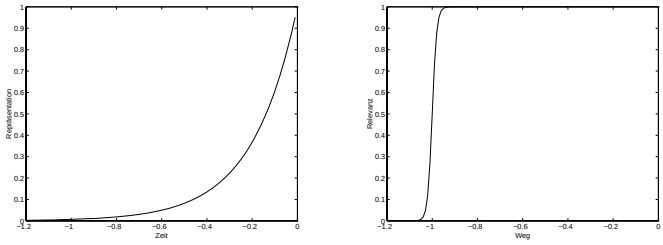

Abbildung 6: *Während rückgekoppelte Netzwerkstrukturen zurückliegende Information vergleichbar mit PT_1-Verhalten repräsentieren (links) bleibt die Relevanz von Bildinformation für ein kollisionsvermeidendes Verhalten über einen den Fahrzeugabmessungen entsprechenden Weg gleichbleibend hoch und bricht erst danach vollständig ein (rechts).*

Wie in Abbildung 6 deutlich wird, bietet sich gegenüber der Einführung von rekurrenten Verbindungen oder dynamischen Neuronen (z.B. mit PT_1–Verhalten) die Einrichtung eines „Wegfensters“

an, in dem zusätzlich die Inputvektoren zurückliegender Situationen mitgeführt werden. Diese Struktur stellt sich als eine vereinfachte Anwendung der „Sliding-Window“-Technik in TDNN-Netzen nach [WH89] dar. Wichtig ist bei einer solchen Erweiterung der Netzwerk-Input-Schicht, daß sich der Versatz der gleichzeitig dargestellten Situationen auf eine konstante Wegbasis bezieht. Um dies auch bei variabler Geschwindigkeit zu realisieren, werden bis zu einem bestimmten Horizont die Inputvektoren gemeinsam mit den zugehörigen Schrittdistanzen zwischengespeichert, von denen die vom aktuellen Standpunkt aus definiert zurückliegenden Wegpunkte ausgewählt werden.

In der Anwendung zeigte diese Erweiterung die erhofften Erfolge nicht nur im Ausrichtungsverhalten zu den Hindernissen bzw. Wegverbreiterungen im erwünschten Abstand (Abb. 7), sondern aufgrund der in der History zusätzlich repräsentierten Information auch in einer Reduktion von Konfliktsituationen und einer Tolerierung kurzzeitigen Verlusts des Untergrund-Sichtkontakts. Bewährt hat sich dabei ein Wegfenster der Tiefe 3 mit einer Wegdifferenz von 3 cm (entspricht halbem KHEPERA-Durchmesser). Zusätzliche Vergangenheitsebenen erhöhen zu stark die Anzahl der zu adaptierenden Wichtungen, weiter auseinanderliegende Ansichten variieren zu stark, um sie durch ein Netzwerk in einem sensomotorischen Zusammenhang nachvollziehen zu können.

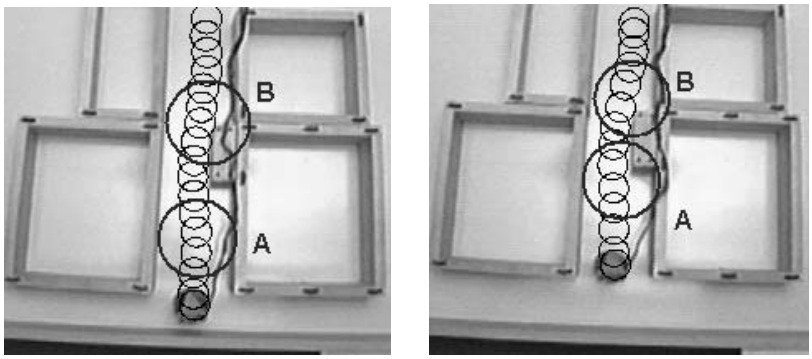

Abbildung 7: *Aufzeichnung der Fahrtrajektorien bei einer Flureinengung: Mit dem Netzwerk ohne History-Input ist ein frühes Einlenken (A) notwendig und die zu frühe Rückkehr zur Flurmitte ist nicht zu vermeiden (B) (links). Mit der zusätzlichen Information des Wegfensters erfolgen sowohl die Ausweich- (A) als auch die Rückkehrbewegung (B) mit dem erwünschten Abstand zum Hindernis (rechts).*

3.2.2 Geschwindigkeitsdimension und Aktivierungsfokus

Parallel zu der oben beschriebenen Input-Erweiterung wurde mit der Einführung der Geschwindigkeit der Aktionsraum um eine Dimension erweitert, um auch das Geschwindigkeitsverhalten des Trainers während der Fahrt nachzubilden. Der bisherige m-dimensionale Output-Vektor wurde deshalb in einen m$\times$n–dimensionalen umgewandelt, auf dem nun in einer zweidimensionalen (ca. M$\times$N breiten) GAUSS-Funktion Lenkwinkel und Geschwindigkeit abgebildet werden (Gl. 1, Abb. 8).

$$y_{ij} = e^{-\frac{\left(i-\frac{m+1}{2}\cdot(1+\frac{LS-RS}{LS+RS})\right)^2}{M}} \cdot e^{-\frac{\left(j-\frac{N}{2}-\frac{LS+RS}{LS_{max}+RS_{max}}\cdot(n-N)\right)^2}{N}} \qquad i \in [1\ldots m]; j \in [1\ldots n] \tag{1}$$

Aus methodischer Sicht bietet die 2-dimensionale Repräsentation gegenüber einer denkbaren 2–fach eindimensionalen Abbildung von Geschwindigkeit und Lenkeinschlag entscheidende Vorteile, weil sich über der Fläche der Output-Neuronen mehrere Handlungsvorschläge **separierbar** darstellen lassen. So läßt sich mit der erhaltenen Ausgabeaktivierung in geeigneter Weise ein dynamisches neuronales Feld nach [Ama77] und [Kop96] anregen, mit dem genau die Effekte erreicht werden können, die eine „intelligente“ Auswertung des Netzoutputs beinhaltet: (I) Durch Mechanismen der

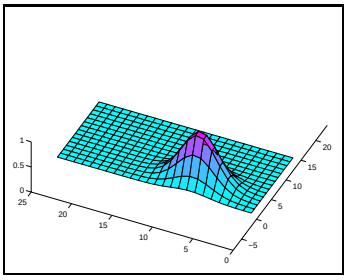

Abbildung 8: *2–dimensionale topologische Aktionskodierung für langsames Rechtsabbiegen als ca. 10×5 breite* GAUSS*-Glocke auf einem 25×15-dimensionalen Outputvektor (in Flächendarstellung).*

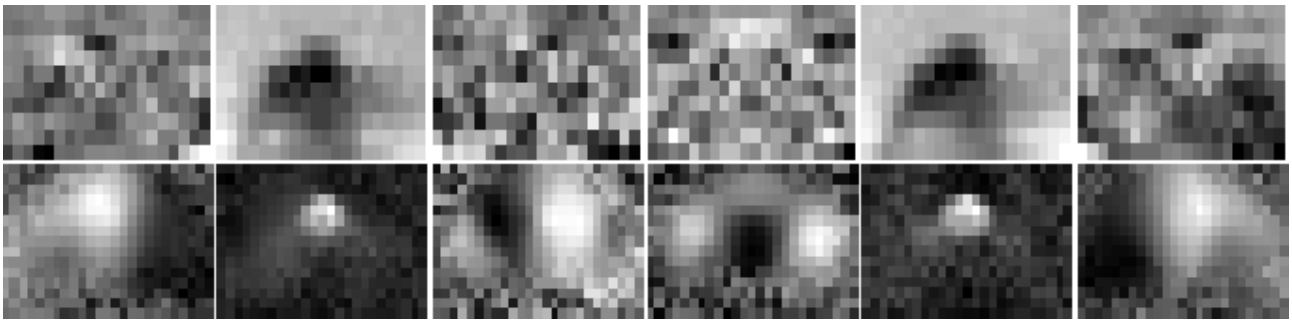

Abbildung 9: *Die afferenten (oben) und efferenten (unten) Wichtungen der 6 Hidden-Neuronen eines trainierten Netzwerks mit 2–dimensionaler Outputkodierung, aber ohne Wegfenster: Während in den Wichtungsmustern der Inputschicht (obere Zeile) sich deutlich die Struktur des Input-Datenstroms wiederspiegelt (vergl. Abb. 2) zeigt die Outputschicht (untere Zeile) starke exzitatorische (hell) und inhibitorische (dunkel) Wichtungen in den typischen Regionen des Aktionsraumes (Stopaktion bei v=0,ϕ=0: in den Aktionskarten mittig unten).*

Maximumsselektion wird auf die Region fokussiert, die sich sowohl in ihrer räumlichen Ausprägung als auch in ihrer Aktivität von ihrer Umgebung am stärksten abhebt. Alle nicht unterstützten Bereiche werden in ihrer Aktivität inhibiert. (II) Nach der Selektion eines lokalen energetischen Maximums wird dieses auch dann weiter fokussiert, wenn dessen Aktivität in einem begrenzten Maße unter die anderer absinkt. Durch die Hystereseeigenschaft ergibt sich ein gewisses Beharrungsvermögen, so daß auch bei kurzzeitigen Einbrüchen der lokalen Aktivierung der Fokus nicht verloren wird. (III) Auch bei einer sich verschiebenden lokalen Aktivierung wird deren Zentrum weiter abgebildet. Damit können die räumlich nie stationären „Blobs“ in der Outputebene des Netzwerks stabil verfolgt werden (Tracking) [KBSG97].

Die besten Resultate in der Auswertung der Ausgabeaktivierung der einzelnen Agenten wurden erzielt, wenn in den durch das nachgeschaltete dynamische neuronale Feld selektierten Regionen die Position des maximal aktivierten Neurons zur Erzeugung des Fahrbefehls herangezogen wurde. Die Fusionierung der Aktionsvorschläge mehrerer Agenten kann in ähnlicher Weise erfolgen, indem der Eintrag deren Ausgabeaktivierungen in eine gemeinsame Karte durch die jeweiligen neuronalen Felder regional gebahnt wird. Über der resultierenden „Gesamterregung“ wird durch ein weiteres dynamisches neuronales Feld wiederum die energiereichste Region selektiert, in der die maximale Aktivierung die letztendlich auszuführende Aktion bestimmt.

4 Ergebnisse

Um eine erste Demonstration des Zusammenwirkens des hier vorgestellten Agentenkollektivs zu realisieren, wurde vergleichbar den Braitenberg-Vehikel[Bra84] eine „Such“-funktionalität implementiert, indem ein farbiges Objekt entsprechend seiner horizontalen Position im Kamerabild in eine entsprechende Agentenmodulation umgesetzt wurde: Proportional zum Abstand des „zu suchenden“ Objekts von der Bildmitte wurde der diese Verschiebung kompensierende Agent (wenn rechts, dann

„turn right“; wenn links, dann „turn left“) in seinem Aktivierungseintrag in die gemeinsame Aktionskarte bis zu 4-fach verstärkt. Zur Verfolgung des Suchobjektes wurde ebenfalls ein (hier eindimensionales) dynamisches neuronales Feld vorteilhaft angewendet. Durch geeignete Parametrierung wird dabei ein weiteres Verhalten nutzbar: die Erregung in den äußeren Neuronen des neuronalen Felds wird bei einem Herauslaufen des Objektes aus dem Bildrand solange erhalten, bis neue Inputerregung den Fokus auf sich zieht. Durch diesen Speichereffekt sucht der Roboter weiter in die Richtung, in der er das Objekt aus dem Blickfeld verliert.

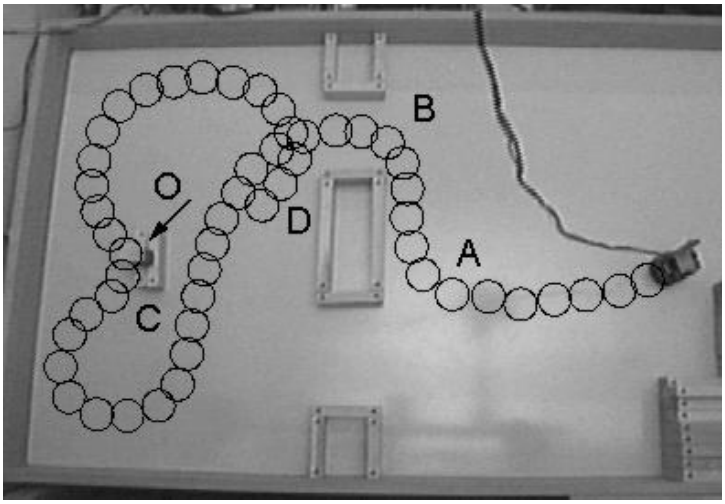

Abbildung 10: *Das „Suchverhalten“ des KHEPERA wird in der Aufzeichnung deutlich sichtbar (Zielobjekt links auf einem Hindernis stehend – O): dem Hindernis (verdeckt nicht das Zielobjekt) wird im letzten Moment ausgewichen, danach wird immer wieder auf das Objekt zugefahren, ohne aber damit zu kollidieren, weil es selbst als Hindernis wahrgenommen wird.*

Abbildung 10 verdeutlicht verschiedene Aspekte der Wechselwirkungen zwischen den Agenten: Die wechselseitige Verstärkung der beiden Abbiegeagenten verursacht ein auf das Objekt gerichtete, aber unstetige Bewegung, wodurch dicht vor dem Hindernis letztendlich der „turn left“-Agent zu einer Ausweichbewegung nach rechts gezwungen ist. Der vermeintlich kürzere Weg wird durch die erzeugte Systemreaktivität nicht zwingend gefunden (A). Da der „turn left“-Agent verstärkt bleibt, wird die Einfahrt nach links genutzt (B). Das Zielobjekt selbst stellt ebenfalls ein Hindernis dar und verursacht so eine Ausweichbewegung, die eine Kollision verhindert (C). Nach der Ausweichbewegung kehrt der Roboter wieder zum Zielobjekt zurück (D). Die großen Kurvenradien auf freien Flächen entsprechen dem Trainingsverhalten in solchen Situationen, weil keine Objekte zur Beobachtung der Eigenbewegung vorhanden sind und bei engeren Kurvenradien mit Objekten im toten Blickwinkel kollidiert werden kann.

5 Ausblick

Hauptziel der weiteren Arbeiten wird eine adaptive Gestaltung der einzelnen Agenten mittels eines vereinfachten Lifelong–Reinforcement–Learning sein. Die Adaptivität der Einzelagenten soll erreicht werden, indem den MLP-Netzwerken, die in der Anwendungsphase festgeschrieben werden müssen, eine parallel operierende, aktiv lernende Struktur zugeordnet wird, die den situationsspezifischen Aktionen des MLP ein Kompetenzmaß (utility) bezüglich der jeweiligen Situation zuordnet. Solch eine Struktur ist dem AHC (adaptive heuristic critic) –Ansatz nach [BSA83] vergleichbar: ein zusätzliches Netzwerk (adaptive critic element, ACE) erlernt eine Wertung (value function) der fixierten Input-Output-Projektion (fixed policy) des MLP (Abb. 11) .

Da das Gesamtsystem letztendlich Aktionen ausführt, die von der vorgeschlagenen Aktion eines einzelnen Agenten verschieden sein können, muß zunächst jeder Agent für sich auf Basis eines Vergleichs von vorgeschlagener und ausgeführter Aktion ein Verantwortlichkeitsmaß bestimmen (responsibility). Darauf basierend lernt das ACE eine Vorhersage des aktionsspezifischen Handlungserfolgs (reward: $0 \ldots 1$), mit dem die Ausgabeaktivierung des MLP moduliert wird. Für das

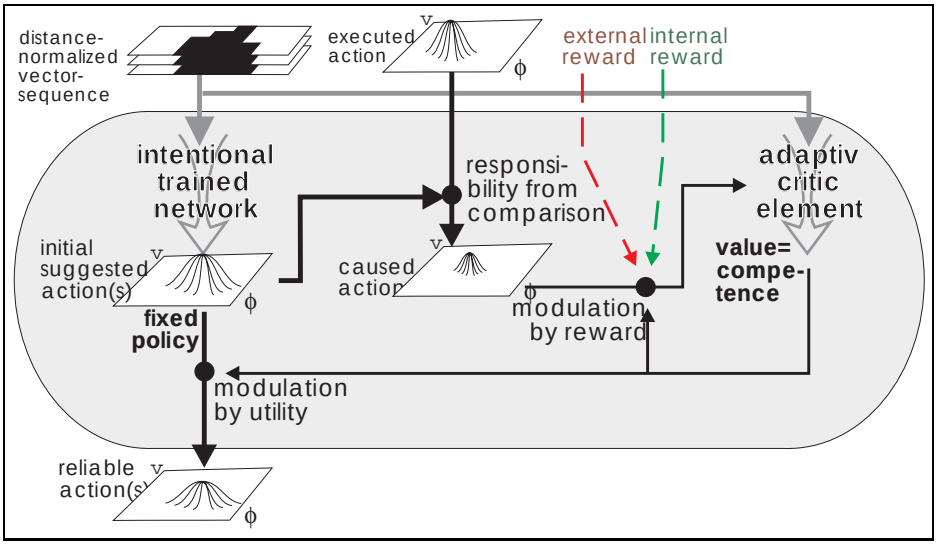

Abbildung 11: *Entwurf einer funktionellen Gesamtverschaltung eines aktiv lernenden Agenten: Ein parallel operierendes, intern überwacht lernendes Netzwerk moduliert situationsspezifisch die Ausgaben des fest vortrainierten Netzwerks entsprechend des erwarteten Handlungserfolgs.*

ACE sind Architekturen notwendig, die im Gegensatz zu dem als Funktionsapproximator wirkenden vortrainierten Netz nicht auf einem statistischen Lernen der Input-Output-Abbildung basieren, sondern auch kritische Einzelfälle durch ein Einschritt-Lernen (One-Shot-Learning) erfassen können. Auf diese Weise wird dem stark generalisierenden, auf Teilmengen nur beschränkt eingehenden, trainierten Netzwerk zur sensomotorischen Projektion eine aktiv lernende Struktur entgegengesetzt, welche diese Projektion verletzende Sonderfälle, sogenannte Negativbeispiele, repräsentiert.

Literatur

[Ama77] AMARI, Shun–Ichi: Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. **In:** *Biological Cybernetics* 27 (1977), S. 77 – 87

[BBB 97] BÖHME, H.–J. ; BRAKENSIEK, A. ; BRAUMANN, U.–D. ; KRABBES, M. ; GROSS, H.–M.: Neural Architecture for Gesture-Based Human-Machine-Interaction. **In:** *Gesture-Workshop Bielefeld*, September 1997

[Bra84] BRAITENBERG, Valentin: *Vehikel — Experimente mit kybernetischen Wesen*. Reinbek: Rowohlt, 1993, 1984

[BSA83] BARTO, A. G. ; SUTTON, R. S. ; ANDERSON, C,W.: Neuronlike adaptive elements that can solve difficult learning control problems. **In:** *IEEE Transactions on Systems, Man, Cybernetics SMC* Bd. 13. Bd. 13, 1983, S. 834 – 846

[KBSG97] KRABBES, M. ; BÖHME, H.–J. ; STEPHAN, V. ; GROSS, H.–M.: Extension of the ALVINN–Architecture for Robust Visual Guidance of a Miniature Robot. **In:** *EUROBOT'97 -2nd EUROMICRO Workshop on Advanced Robile Robots-, Brescia*, Oktober 1997

[Kop96] KOPECZ, K.: Neural Field Dynamics Provide Robust Control of Attentional Resources. **In:** *Aktives Sehen in technischen und biologischen Systemen*, Proceedings of the German Society of Computer Science (GI) Workshop, Hamburg, 1996, S. 137 – 144

[PG96] POMIERSKI, T. ; GROSS, H.–M.: Biological neural architecture for chromatic adaptation resulting in constant color sensations. **In:** *Proc. of the ICNN-96, Washington DC*, IEEE-Press, 1996, S. 734–739

[Pom93] POMERLEAU, D. A.: *Neural Network Perception for Mobile Robot Guidance*. Kluwer Academic Publishers, Boston/Dordrecht/London, 1993

[WH89] WAIBEL, A. ; HAMPSHIRE, J.: Building blocks for Speech. **In:** *Byte* 8 (1989), S. 235 – 242