SZENENHAFTE MODELLE FÜR ZEITABHÄNGIGE EREIGNISSE

M. MOHNHAUPT, B. NEUMANN

FBI-HH-B-127/87

FEBRUAR 1987

FACHBEREICH INFORMATIK
UNIVERSITÄT HAMBURG
BODENSTEDTSTRASSE 16
D-2000 HAMBURG 13

Kurzfassung

Im Kontext einer natürlichsprachlich gesteuerten Szenenanalyse soll eine Äußerung wie z.B. "Ist ein Auto in die Schlüterstrasse abgebogen ?" für eine 'Top-down' Steuerung von niederen Ebenen des Bildverstehens ausgenutzt werden. Die für eine effektive Kontrolle des 'Sehprozesses' notwendigen räumlichen und zeitlichen Randbedingungen müssen in diesem Fall u.a. aus dem Verb der Anfrage (hier 'abbiegen') abgeleitet werden. Die in früheren Arbeiten (NEUMANN und NOVAK 83, NEUMANN 84, NOVAK und NEUMANN 86) vorgeschlagenen propositionalen Ereignismodelle sind gut geeignet für die Generierung natürlichsprachlicher Äußerungen ('Bottom-up'), zeigen sich aber als weniger gut verwendbar für eine 'Top-down' Steuerung. Hierfür erweist sich eine eher szenenhafte, also der Szene analoge Repräsentation als günstig, welche räumliche und zeitliche Beziehungen explizit macht, und mit der z.B. typisches Verhalten, Unsicherheit und Unschärfe adäquat behandelt werden kann. Eine solche Repräsentation, sowie die auf ihr arbeitenden Prozesse, sind Gegenstand dieses Beitrages. Außerdem wird gezeigt, wie diese Repräsentation auf einfache Weise durch Anhäufung von Erfahrung (Sammlung von Einzelbeispielen) gebildet werden kann.

Abstract

In this report trajectory accumulation frames (TAFs) are proposed as a means for recording trajectories of moving objects and representing the accumulated experience of many observations. It is shown that a TAF can be used to generate expectations about new motion situations. Depending on the degree of generalization and abstraction, predicted behavior may follow an overall trent or resemble individual experiences. In addition TAFs allow to predict typical behavior in situations where obstacles constrain the possible movements. It is intended to use TAFs for natural-language guided motion analysis in a vision system.

SZENENHAFTE MODELLE FÜR ZEITABHÄNGIGE EREIGNISSE

1. Einleitung

Abbildung 1

Dem uninformierten Betrachter wird es vermutlich nicht gelingen, Abbildung 1 zu erkennen. Hat man aber die Information, daß die Zeichnung eine Kuh in Seitenansicht darstellt, welche den Kopf in Richtung Betrachter gedreht hat, wird die Interpretation des Bildes möglich.

Dies Beispiel macht den Einfluß von 'Top-down' Information bei der visuellen Wahrnehmung deutlich. Offensichtlich kann 'Top-down' Information die Wahrnehmung vereinfachen und beschleunigen. Darüber hinaus demonstriert Abbildung 1, daß in manchen Situationen ohne 'Top-down' Information überhaupt keine sinnvolle Interpretation vom Sehsystem geleistet werden kann. Die 'Top-down' Information schränkt den Suchraum drastisch ein, wobei weitgehend ungeklärt ist, wie dies im Einzelnen passiert. Ohne 'Top-down' Information scheint eine kombinatorische Explosion bei der Hypothesenbildung des visuellen Systems die Erkennung der Kuh zu vereiteln.

Wir befassen uns in unseren Untersuchungen mit der Bedeutung und den Möglichkeiten von 'Top-down' Information für maschinelle Sehsysteme. Insbesondere sind wir an einer 'Top-down' Steuerung von niederen Bilddeutungsprozessen für die Entdeckung und Verfolgung von Bewegungen interessiert. Am Beispiel einer natürlichsprachlich gesteuerten Szenenanalyse

von Strassenverkehrsszenen soll u.a. eine Äußerung wie "Ist ein Auto in die Schlüterstrasse eingebogen ?" für eine 'Top-down' Steuerung genutzt werden. Information über die Art der Bewegung ('abbiegen'), den Ort der Bewegung ('Schlüterstrasse') und das Objekt, das sich bewegt ('Auto') , sollen eine gezielte Bewegungsanalyse ermöglichen. Nach BINFORD 82 dürfte bei dem gegenwärtigen Stand der Kunst auch nur ein 'Top-down' Verfahren zu akzeptablen Ergebnissen bei Bildmaterial aus der natürlichen Umwelt führen.

In diesem Report diskutieren wir die Ableitung von geometrischen und zeitlichen Randbedingungen für die 'low-level' Analyse aus der Information, die in einer natürlichsprachlichen Äußerung enthalten ist. Dies ist ein wichtiges Teilproblem einer 'Top-down' Steuerung. Die Überführung der natürlichsprachlichen Äußerung in eine Tiefenstruktur (Parsing) und eine effektive Bewegungsanalyse unter den abgeleiteten Randbedingungen werden hier nicht behandelt.

Wir nehmen an, daß der Informationsgehalt der natürlichsprachlichen Äußerung in Form von Kasusrahmen der Verben vorliegt (z.B.: Verb: abbiegen, Agent: Auto1, Lokativ: Schlüterstrasse, ...). Als Ergebnis der vorgeschlagenen Ereignismodelle und Verfahren sollen die räumlichen und zeitlichen Randbedingungen in einer Weise explizit gemacht werden, daß sie direkt von Komponenten der niederen Bilddeutung verwendet werden können. Mithilfe von Ereignismodellen (z.B. für 'abbiegen') und aktuellen Daten (ein angefangener Abbiegevorgang) soll eine Vorhersage darüber berechnet werden, was (Auto) in naher Zukunft wo zu erwarten ist, um dann später eine eingeschränkte Bewegungsanalyse durchzuführen.

Mit den vorgeschlagenen Repräsentationen und Prozessen erheben wir nicht den Anspruch, psychologische Ergebnisse zu erklären. Dennoch dienen uns Erkenntnisse über die Kognition bei Menschen als Vorbild oder mindestens als Ideenlieferant. BLOCK 81 und KOSSLYN 80 geben einen Überblick über die Repräsentation bildhafter Vorstellungen ('Imagery' Debatte). Leider existiert wenig Literatur über die Repräsentation zeitabhängiger Zusammenhänge im kognitiven System des Menschen. Für die Steuerung der Szenenanalyse sind u.a. Untersuchungen über Wechselwirkungen zwischen bildhaften Vorstellungen und visuellen Wahrnehmungen relevant.

Im Kontext eines 'Top-down' gesteuerten Bildverstehens stellen wir an Ereignismodelle die folgenden funktionalen Anforderungen :

(a) Erlernbarkeit

Die Repräsentation soll erlernbar sein aus gemachten Erfahrungen. Wir wollen Wissen über typische Objektbewegungen ansehen als angehäuft und abstrahiert aus konkreten Beobachtungen (Ansammlung von Einzelbeispielen). Die Benutzung derselben Datenstruktur für konkrete Beobachtungen und für daraus abgeleitetes Wissen erlaubt eine sinnvolle Interpretation des Modellwissens. Insbesondere können Visualisierungen generiert werden unter Benutzung derselben Datenstruktur, wie für visuelle Daten. FINKE 80 und FINKE 85 präsentieren Evidenz dafür, daß beim visuellen System des Menschen Vorstellungen und Wahrnehmungen z.T. auf denselben Datenstrukturen operieren.

(b) Unschärfe

Vorerwartungen über typische Objektbewegungen sind oft unscharf, d.h. sie können Objektbewegungen nicht exakt spezifizieren, sondern nur innerhalb einer gewissen Bandbreite. Dabei können einige Abbiegevorgänge typischer als andere sein. Andererseits können bestimmte Eigenschaften, in manchen Fällen z.B. die Bewegungsrichtung, mit Sicherheit repräsentiert werden. Die Repräsentation soll verschiedenen Graden der Sicherheit und Typizität gerecht werden, und zwar in einer natürlichen (erfahrungsbasierten) Weise.

(c) Generalisierung

Beim Lernen aus Beobachtungen ist die Generalisierungsfähigkeit von Einzelfällen eine wichtige Anforderung an die Repräsentationsform. Die Repräsentation soll auch solche Fälle abdecken können, die nicht in jedem Detail erfahren wurden. Wenn z.B. ein Auto geringfügig anders abbiegt als die bisherigen Beobachtungen, soll trotzdem eine Vorhersage seines ungefähren Kurses möglich sein. Außerdem soll eine adäquate Behandlung von in gewissem Sinne ungewöhnlichen Situationen möglich sein. Wenn z.B. in bestimmten Situationen ein Hindernis (z.B. ein haltendes Auto) auf einer Kreuzung steht, muß das Modell eine der Situation angepaßte

Vorhersage erlauben. Ein weiterer Aspekt von Generalisierung betrifft einen Wechsel der stationären Umgebung. Wir wollen z.B. 'Abbiegeerfahrung' generalisieren von einer bestimmten Kreuzung und es auf andere Kreuzungen anwenden können, auch wenn diese eine andere geometrische Form haben.

(d) Gruppierung

Beim Anhäufen von Erfahrungen können verschiedene Verhaltensmuster auftauchen, z.B. 'Rechtsabbiegevorgänge' und 'Linksabbiegevorgänge'. Solche 'Makromuster' können als Gruppierung bzw. Abstraktion von Einzeltrajektorien angesehen werden. Gruppierung ist sinnvoll, um zwischen konzeptuellen Einheiten zu unterscheiden, bzw. um eine 'Grobsicht' auf das Modell zu ermöglichen.

Wir schlagen als Repräsentation einen 'Trajectory Accumulation Frame' (TAF) vor. TAFs sind 4-dimensionale Zählerfelder für Zustandsvektoren, welche Bewegung beschreiben. Jeder Eintrag korrespondiert mit der Anzahl von Beobachtungen, die für einen bestimmten Zustand gemacht wurden. In Abschnitt 2 wird dieses Ereignismodell im Detail entwickelt und im Zusammenhang mit der Speicherung und dem Abruf von Einzeltrajektorien diskutiert. Abschnitt 3 behandelt Situationen, in denen Einzeltrajektorien nicht mehr in der gespeicherten Erfahrung unterschieden werden und die TAFs benutzt werden, um 'typisches' Verhalten vorherzusagen. Es wird gezeigt, daß das Verfahren zum Abruf von Einzelbeispielen direkt übertragen werden kann auf den für uns interessanteren Fall, wahrscheinliches bzw. typisches Verhalten zu prognostizieren. Neben dem Vorhersageverfahren werden Operationen diskutiert, die der Aufbereitung von Erfahrungen in der Lernphase dienen. Dabei wird u.a. von Details abstrahiert, und fehlende Erfahrungen werden ergänzt. In Abschnitt 4 wird auf Situationen eingegangen, in denen eine erfahrungsbasierte Vorhersage auf besondere, so nicht erfahrene Verhältnisse angepaßt werden muß (Hindernisproblem). Dazu wird eine Inhibitionsoperation vorgeschlagen, welche im Einzelfall bestimmte Erfahrungen unterdrückt. Abschnitt 5 dient der Zusammenfassung und Diskussion der Ergebnisse, sowie dem Ausblick auf zukünftige Arbeiten.

Die vorgeschlagenen Ereignismodelle und Verfahren wurden auf einer Symbolics 3640 implementiert. Die gezeigten experimentellen Ergebnisse sollen die

theoretischen Überlegungen untermauern.

2. Trajectory Accumulation Frames (TAFs)

In diesem Abschnitt diskutieren wir 'Trajectory Accumulation Frames' als Datenstrukturen für das Speichern und Abrufen von Objekt-Trajektorien. Unsere Beispieldomäne sind Strassenverkehrsszenen, in denen Ereignisse wie 'abbiegen', 'überholen', etc. modelliert werden sollen. Wir beschränken uns daher auf Bewegungen in der Ebene. Die Ergebnisse sind auf Bewegungen im Raum erweiterbar.

In früheren Arbeiten (NEUMANN und NOVAK 83, NEUMANN 84, NOVAK und NEUMANN 86) wurden propositionale Ereignismodelle für die Repräsentation der Bedeutung von Bewegungsverben dieser Art vorgeschlagen, um aus 4-dimensionalen Szenenbeschreibungen natürlichsprachliche Äußerungen zu generieren. Im Kontext von 'Top-down' gesteuertem Bildverstehen erweisen sich propositionale Modelle insbesondere deshalb als weniger geeignet, weil geometrische Detailinformationen, welche für eine Vorhersage benötigt werden, bei propositionalen Beschreibungen verloren gehen. Eine der Szene analoge ('szenenhafte') Beschreibung, die räumliche und zeitliche Beziehungen explizit macht, kann die von dem Modell geforderten funktionalen Eigenschaften besser erfüllen.

Ein TAF ist ein 4-dimensionales Zählerfeld C (x,y,r,b), das ein bestimmtes Gebiet der xy-Ebene abdeckt (z.B. eine bestimmte Kreuzung). Jeder Punkt in diesem 4-dimensionalen Raum hat einen Zähler und repräsentiert eine bestimmte Kombination von xy (Ort), r (Geschwindigkeitsrichtung) und b (Geschwindigkeitsbetrag). Der Vektor $\underline{S}$ = (x,y,r,b) beschreibt den Zustand eines Objekts (z.B. eines Autos) zu einer bestimmten Zeit. Jede Objekttrajektorie (z.B. bei einem Abbiegevorgang) hinterläßt eine zusammenhängende Spur im 4-dimensionalen Feld, indem sie die Zähler der 'getroffenen' (x,y,r,b) Zellen inkrementiert. Wenn mehrere Objekte nacheinander beobachtet werden, werden mehr (oder möglicherweise dieselben) Zellen inkrementiert, ohne daß zwischen verschiedenen Objekten unterschieden wird.

Im Folgenden werden wir den Abruf von einzelnen Trajektorien aus einem TAF diskutieren. Bei einer gegebenen Startzelle liegt es nahe, in der Nachbarschaft nach einer Zelle zu suchen, deren Zähler grösser als Null ist. Diesem Vorgehen liegt die einfache Annahme zugrunde, daß die wahrscheinlichste Fortsetzung einer angefangenen Trajektorie in ihrer Nachbarschaft zu suchen ist. Wir bezeichnen den Bereich, der von einem Startpunkt aus nach Fortsetzungen für eine Trajektorie abgesucht wird, als qualifizierte 4D-Nachbarschaft eines Punktes.

In der Projektion des 4-dimensionalen Modells auf die xy-Ebene sind nicht alle Nachbarn auch qualifizierte 4D-Nachbarn. Denn Punkte, welche im 4-dimensionalen Modell getrennt sind, weil sich z.B. ihre Geschwindigkeitsrichtungen drastisch unterscheiden, können in der Projektion auf die xy-Ebene benachbart sein. Es kommen also nur solche Punkte als Nachfolger in Frage, die in keiner der vier Koordinaten sprunghaft abweichen.

Wir verwenden folgende Diskretisierungen :

- die xy-Ebene wird durch ein quadratisches Gitter repräsentiert (in unseren Experimenten 40x40)
- Geschwindigkeitsrichtungen werden in 45-Grad Schritten diskretisiert (Kettenkode)
- Geschwindigkeitsbeträge werden in einem sinnvollen Teilbereich diskretisiert (z.B. in 6km/h Schritten von 0-60km/h)

Die qualifizierte 4D-Nachbarschaft N(S) einer Zelle S enthält alle Zellen S'

- die in einem 3x3x3x3 Würfel um S liegen,
- deren Geschwindigkeitsrichtung r' ihrer Lage relativ zu S entspricht (von S aus eine kontinuierliche Fortsetzung erlaubt)
- und von der Richtung r um nicht mehr als 45 Grad abweicht.

Also enthält N(S) insgesamt 9 Elemente (an 3 verschiedenen Orten mit je 3 verschiedenen Geschwindigkeitsbeträgen). Jede Zelle in N(S), deren Zähler grösser als Null ist, ist mögliche Nachfolgezelle von S. Im Allgemeinen gibt es mehr als einen möglichen Nachfolger, deshalb definieren wir den Nachfolger von S als diejenige Zelle in N(S) mit dem höchsten Zähler.

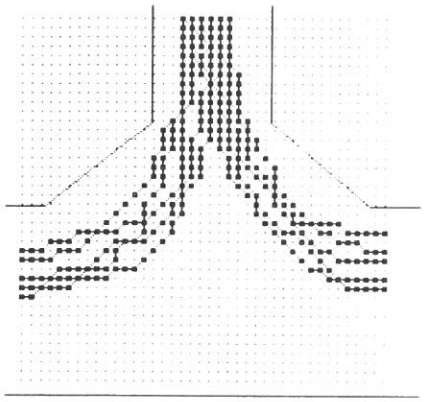

Abb. 2a: TAF mit 10 Trajektorien

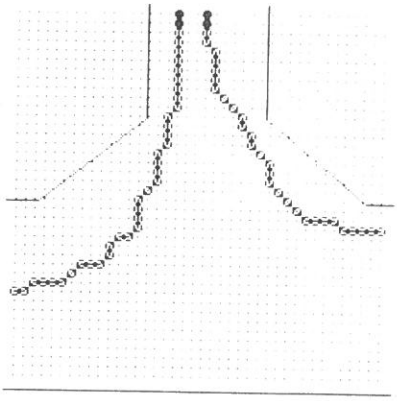

Abb. 2b: Zwei abgerufene Trajektorien

Abbildung 2a zeigt die Projektion eines TAF mit 10 Trajektorien auf die xy-Ebene. Die wegen r oder b im Modell unterscheidbaren Zellen sind in der Projektion auf xy nicht mehr unterscheidbar. Erst wenn zwei Trajektorien an einem Punkt in allen vier Koordinaten übereinstimmen, inkrementieren sie den selben Zähler. Bei einem Trajektorien-Abruf bedeutet dies, daß von einem beobachteten Beispiel auf ein anderes 'übergewechselt' werden kann. Eine denkbare Erweiterung des Modells auf Informationen 2-ten Grades (Beschleunigungen) könnte einen Teil dieser Überschneidungen auflösen, scheint uns zur Zeit aber nicht gerechtfertigt für die Befriedigung der anfangs formulierten funktionalen Anforderungen. Wenn eine Vorhersage wegen lokal übereinstimmender Zustandsbeschreibungen auf der 'falschen' Trajektorie landet, werden unsere funktionalen Anforderungen nicht unterlaufen, da wir an der Vorhersage von typischem Verhalten und weniger an dem Nachvollziehen von Einzelbeispielen interessiert sind.

In der Abbildung 2b sind zwei unterschiedliche Vorhersagen (offene Kreise) von unterschiedlichen Startpunkten (geschlossene Kreise) zu sehen. In den oben gezeigten und allen weiteren Beispielen wurden die Geschwindigkeitsbeträge konstant gehalten. Dies vereinfacht und veranschaulicht die graphische Darstellung des 4-dimensionalen Aktivitätenfeldes, mindert aber die Aussagekraft der Beispiele nicht.

3. Typische Trajektorien und Prototypen

Im letzten Kapitel ist deutlich gemacht worden, wie TAFs definiert sind und wie aus TAFs einzelne Trajektorien abgerufen werden können. Im Folgenden diskutieren wir die Repräsentation von prototypischem Verhalten in TAFs und beschreiben, wie eine Prognose über das wahrscheinliche Verhalten einer konkreten Trajektorie berechnet werden kann. Wir gehen davon aus, daß sprachliche Äußerungen mit prototypischem Verhalten verknüpft werden können. Daher wollen wir im Kontext einer natürlichsprachlich gesteuerten Szenenanalyse die aus der Anfrage resultierenden räumlichen und zeitlichen Randbedingungen explizit machen. Denn diese Ergebnisse sollen als Filter in Prozesse der niederen Bilddeutung eingehen.

Wir diskutieren Situationen, in denen eine Vorhersage nicht mehr von einem Einzelbeispiel abhängt, sondern von der Erfahrung aus sehr vielen sich überlagernden Trajektorien. Wir führen in diesem Zusammenhang eine Verwaschungsoperation ein, welche die Zählerfunktion in eine nahe Nachbarschaft propagiert und beschreiben eine Konvergenzoperation, mit der von weniger wahrscheinlichen Trajektorien abstrahiert wird und mit der Maxima verstärkt werden. Ein auf vielen Trajektorien basierender TAF, auf dem Verwaschungs- und Konvergenzoperationen angewandt wurden, stellt die für die 'Top-down' Steuerung der Bildanalyse vorgesehene Wissensquelle dar.

Eine besondere Rolle im 4-dimensionalen Zählerfeld spielen Ketten von lokalen Zählermaxima (Zellen, die 'typischer' sind als ihre Nachbarn). Sie werden dazu ausgenutzt, das in einem TAF kodierte charakteristische Verhalten zu explizieren.

Im Folgenden konkretisieren wir diese Ideen und zeigen experimentelle Ergebnisse. Wir zeigen insbesondere, wie konkrete räumliche und zeitliche Angaben über das Verhalten in naher Zukunft, also ein Richtungs- und Geschwindigkeitsfilter für eine Bewegungsanalyse, berechnet werden kann.

Vorhersagen

In Kapitel 2 wurde beim Abruf von Trajektorien, ausgehend von einem Startpunkt,

nach der Zelle in der qualifizierten 4D-Nachbarschaft gesucht, die den höchsten Zählerstand hat. Für die Berechnung von prototypischem Verhalten interessieren uns drei verschiedene Arten der Vorhersage, die zu Varianten des bisherigen Verfahrens führen.

- Die Vorhersage der wahrscheinlich zu erwartenden Trajektorie, also eine Angabe über prototypisches Verhalten.

- Die Vorhersage aller Trajektorien, welche oberhalb einer bestimmten Wahrscheinlichkeit liegen, also die Angabe eines Bereiches, der z.B. für eine 'low-level' Bewegungsanalyse als Filter fungieren kann.

- Die Vorhersage mehrerer wahrscheinlicher Trajektorien. Dies liefert Informationen über mögliche Alternativen, welche jeweils in ihrer lokalen Umgebung am wahrscheinlichsten sind.

Die erste Art der Vorhersage kann mit dem in Kapitel 2 beschriebenen Algorithmus realisiert werden. Ausgehend von einem Startpunkt wird die Zelle in der qualifizierten 4D-Nachbarschaft mit der höchsten Aktivierung als Nachfolgezelle genommen. Für die zweite Art der Vorhersage wird der Algorithmus geändert. Nachfolgezellen eines Startpunktes sind hierbei alle Zellen in der qualifizierten 4D-Nachbarschaft, welche ein relatives Maximum sind, oder eine Aktivierung oberhalb eines bestimmten Schwellwertes (einer bestimmten Wahrscheinlichkeit) haben. Eine Zelle kann hierbei also mehrere Nachfolgezellen haben, so daß ein Vorhersagegebiet entsteht. Bei der dritten Art der Vorhersage wird in der qualifizierten 4D-Nachbarschaft nur nach Zellen gesucht, die ein relatives Maximum an Aktivität haben (wahrscheinlicher sind als ihre Nachbarn). Auch hierbei kann es zu einer Zelle mehrere Nachfolgezellen geben.

Wir werden die verschiedenen Arten der Vorhersagen anhand experimenteller Ergebnisse jeweils darstellen, nachdem wir weitere Operationen auf TAFs (Verwaschung und Konvergenz), sowie ein Verfahren zur Darstellung von lokal typischem Verhalten (Skelette) definiert haben.

Verwaschung

Verwaschung ist eine Generalisierungsoperation mit dem Effekt, daß Erfahrungen, die durch eine Zählerzelle repräsentiert werden, zu den Nachbarn dieser Zelle propagiert werden. Dies wird dadurch erreicht, daß zu jeder Zelle der gewichtete Mittelwert ihrer Nachbarzellen orthogonal zur Bewegungsrichtung addiert wird. Nachbarzellen in Bewegungsrichtung tragen entsprechend ihrer positiven Differenz zu der Zelle bei. Der Betrag der Aktivierung, der auf die Nachbarn verteilt wird, ergibt sich aus der eigenen Aktivität, gewichtet mit einem Faktor BLUR.

Dieser geschilderten Art der Verwaschung liegt die Annahme zugrunde, daß durch die Beobachtung einer Trajektorie, welche durch eine Menge von Zellen S(i) repräsentiert wird, auch leicht abweichende Trajektorien denkbar sind und mitrepräsentiert werden sollten. Dies ist plausibel, falls keine widersprechenden Informationen aus der Szene bekannt sind (es ist z.B. wenig sinnvoll, über Strassenbegrenzungen hinaus zu verwaschen).

Wir demonstrieren im folgenden einige Effekte der Verwaschung anhand von experimentellen Ergebnissen. Wegen unserer Quantisierung verwaschen wir zunächst nur Ort und Geschwindigkeitsbetrag, da die relativ grobe Richtungsquantisierung (45 Grad Schritte) ein vorsichtiges Vorgehen verlangt (nach zweimaligem Verwaschen der Richtungen wären um 180 Grad abweichende Trajektorien erlaubt, was unerwünscht ist). Die implementationsbedingten Einschränkungen mindern dennoch nicht die Aussagefähigkeit der Ergebnisse.

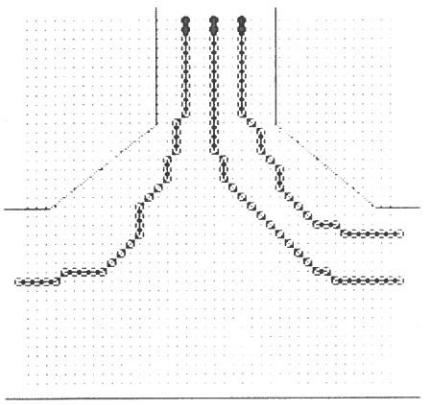

Abb. 3a: Vorhersagen nach einmaliger Verwaschung

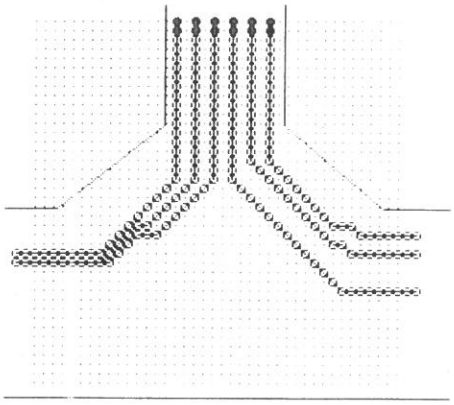

Abb. 3b: Vorhersagen nach dreimaliger Verwaschung

Die Abbildungen 3a und 3b zeigen Vorhersagen der wahrscheinlichsten Trajektorien nach einmaliger bzw. dreimaliger Verwaschung (BLUR = 0.7), ausgehend von verschiedenen Startpunkten (geschlossene Kreise). Das zugrunde liegende Modell (TAF) ist dasselbe wie in Abbildung 2a. Man kann erkennen, daß Vorhersagen in Bereichen möglich sind, in denen keine direkte Erfahrung gemacht wurde. Außerdem ist zu sehen, daß die Vorhersagen nicht unbedingt einzelnen Erfahrungen folgen, sondern den Überlagerungen mehrerer Erfahrungen. Die Propagierung in die Nachbarschaften hat bewirkt, daß Trajektorien parallel zu den gemachten Erfahrungen auch vorhersagbar sind. Dies ist ein erwünschter Generalisierungseffekt.

Man kann an den Vorhersagen auch ersehen, daß mit der Verwaschungsoperation kaum eine Abstraktion geleistet wird. Denn es bleiben beispielsweise nebeneinanderliegende relative Maxima bestehen. Sie addieren sich im Allgemeinen nicht zu einem neuen Hauptmaximum. Außerdem unterdrückt ein Maximum hoher Aktivität i.A. nicht ein weniger starkes Maximum. Beides äußert sich in den Vorhersagen darin, daß ausgehend von einem Startpunkt diejenigen Pfade favorisiert werden, die parallel zu den Pfaden stärkster Aktivierung laufen.

Interessieren uns aber z.B. nur die wesentlichen (häufigsten) Verläufe (z.B. von Abbiegevorgängen), muß eine dahingehende Abstraktion erst noch geleistet werden. Deshalb diskutieren wir eine Abstraktionsoperation 'Konvergenz'.

Konvergenz

Mit der Verwaschungsoperation wurde Information in die nahe Nachbarschaft propagiert. Wir haben demonstriert, daß dadurch auch in Bereichen, wo keine direkten Erfahrung vorliegen, plausible Vorhersagen möglich sind. Bei der Konvergenzoperation werden dicht nebeneinanderliegende Maxima zu einem neuen Maximum abstrahiert und es werden Nebenmaxima zugunsten eines Hauptmaximums unterdrückt. Nach Anwendung dieser Operation können die in einem TAF enthaltenen Verläufe (konzeptuelle Einheiten) explizit gemacht werden. Wir demonstrieren die Konvergenzoperation anhand von Beispielen.

Anschaulich beschrieben, verteilt eine Zelle bei der Konvergenzoperation Aktivität auf diejenigen Zellen, von denen sie bei einer Prädiktion zu erreichen ist. Dadurch verstärken Zellen mit hoher Aktivität Pfade, die zu ihnen hinführen. Abhängig von der Wahl des Parameters CONV werden dadurch parallel laufende Nebenmaxima zugunsten eines Pfades in Richtung des Hauptmaximums aufgelöst.

Eine Zelle $\underline{S}$ = (x,y,r,b) addiert bei der lokalen Konvergenzoperation Aktivität auf alle diejenigen Nachbarzellen $\underline{S}(i)$, von denen aus man die Zelle $\underline{S}$ erreichen kann, abhängig von der eigenen Aktivität gewichtet mit einem Faktor CONV.

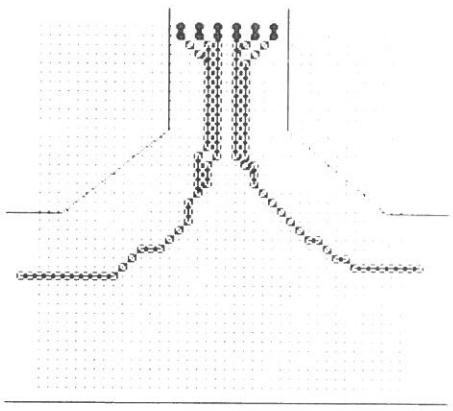

Abb. 4: Vorhersagen nach dreimaligem Verwaschen und einmaliger Anwendung der Konvergenzoperation

Die Abbildung 4 wurde mithilfe der in 2a gezeigten Trajektorien berechnet. Der TAF wurde dreimal verwaschen (BLUR = 0.7) und danach einmalig mit der Konvergenzoperation (CONV = 0.8) bearbeitet. Es werden einige Vorhersagen ausgehend von verschiedenen Startpunkten gezeigt. In der Zeichnung wird deutlich (vergleiche mit 3a und 3b), daß die Konvergenzoperation eine Abstraktion von den Nebenmaxima bewirkt. Die Vorhersagepfade konvergieren zu einem Hauptmaximum, folgen also nach kurzer Zeit den gleichen Pfaden. Sie nähern sich dem Hauptmaximum, sofern es die Änderung der Geschwindigkeitsrichtung zuläßt.

Skelette

Vorhersagen des wahrscheinlichen Verhaltens werden berechnet, indem von einer Zelle ausgehend die Nachfolgezelle mit dem höchsten Zählerstand ausgewählt wird. Die Wege entlang von lokalen Maxima formen ein Muster typischen Verhaltens, daß denjenigen Trajektorien entspricht, die am meisten durch Erfahrung gestützt sind. Wir nennen die Summe dieser Wege das Skelett des TAF. Vorhersagen folgen im Allgemeinen den Skeletten. Skelette sind u.a. deshalb nützlich, weil man unabhängig von einer bestimmten Startzelle mit Skeletten das in einem TAF kodierte typische Verhalten visualisieren kann.

Eine Zelle $\underline{S}$ = (x,y,r,b) ist dann relatives Maximum, wenn sie grösser oder gleich groß, im Vergleich zu allen ihren Nachbarzellen (x+,y,r,b), (x-,y,r,b), (x,y+,r,b), (x,y-,r,b), (x,y,r+,b), (x,y,r-,b), (x,y,r,b+), (x,y,r,b-) ist. Hierbei stehen '+' bzw. '-' für die jeweiligen Nachbarzellen in einer bestimmten Dimension.

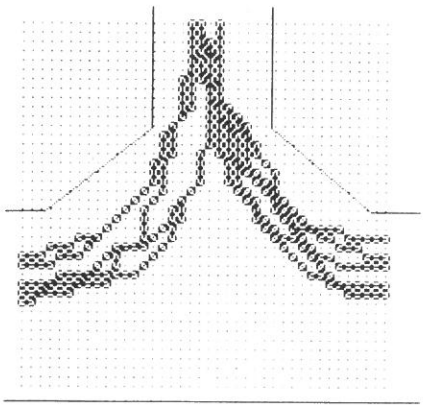

Abb. 5a: Skelett für TAF aus 2a

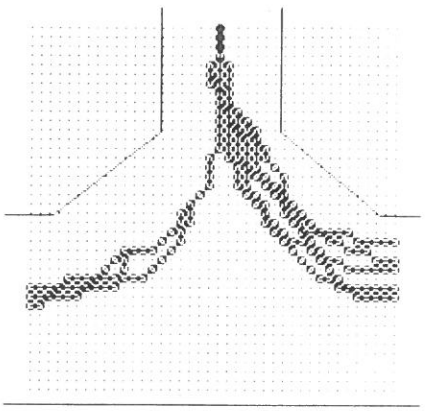

Abb. 5b: Vorhersagen (relative Maxima)

Abbildung 5a zeigt das Skelett für den TAF, welcher mit den in 2a gezeigten Trajektorien gebildet wurde. In Abbildung 5b sind Vorhersagen (relative Maxima) für denselben TAF zu sehen, ausgehend von einem vorgegebenen Startpunkt. 5b stellt eine Untermenge von 5a dar.

Skelette haben die folgenden Eigenschaften:

(a) Der Nachfolger einer Skelettzelle ist wiederum eine Skelettzelle

Dies folgt aus der Definition eines Skeletts. Ein Skelett kann angesehen werden als ein System typischer Beispiele mit Verzweigungen und Vereinigungen an Punkten, wo Alternativen zusammentreffen oder sich verzweigen.

(b) Ein Skelett definiert eine endliche Menge von Trajektorien. Wir nennen die Menge Trajektorien-Prototypen.

Wenn wir, ausgehend von einem Startpunkt, einen Pfad im Skelett verfolgen, gibt es Punkte, an denen mehrere Alternativen möglich sind. Das größte Maximum ist die wahrscheinlichste Fortsetzung, andere relative Maxima führen auf weniger wahrscheinliche Pfade. Die Menge aller dieser Pfade ist die Menge der Prototypen, welche durch den TAF definiert sind. Prototypen können für Gruppierungen nützlich sein, z.B. zum Zerlegen der

Erfahrung in konzeptuelle Einheiten. Nach einer Vielzahl von Beispieltrajektorien und lokalen Abstraktionsoperationen ist es z.B. denkbar, daß ein Skelett nur noch aus zwei wesentlichen Pfaden besteht, einem prototypischen 'Rechtsabbiegen' und einem prototypischen 'Linksabbiegen'. Beispiele dafür werden weiter unten (Abb. 7) gezeigt.

(c) Vorhersagen in der Nachbarschaft des Skeletts tendieren dazu, dem Skelett zu folgen.

Dies folgt ebenfalls aus der Definition von Skeletten. Falls keine anderen Informationen vorliegen, folgt eine Vorhersage also typischen (am meisten durch Erfahrung gestützten) Pfaden.

Weitere experimentelle Ergebnisse

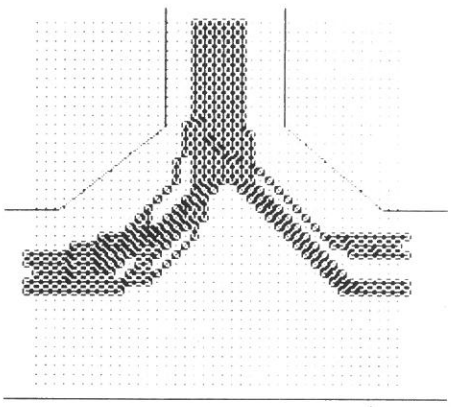

Abb. 6a: Skelett für TAF in 3a

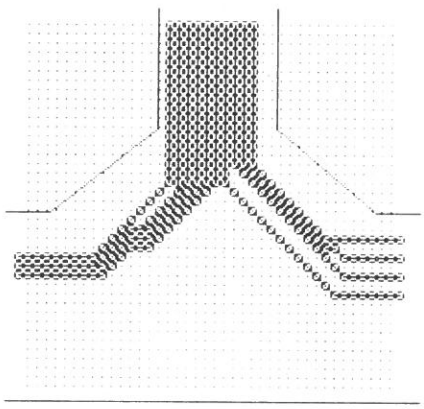

Abb. 6b: Skelett für TAF in 3b

Auch an den Skeletten der TAFs in den Abbildungen 3a (einmal verwaschen) und 3b (dreimal verwaschen) werden die Verwaschungseffekte deutlich. Man sieht die durch Verwaschung verursachte Generalisierung in Bereiche, in denen keine direkten Erfahrungen vorliegen. Viele nebeneinanderliegende Maxima (besonders in 6b) verdeutlichen, daß die Verwaschung kaum eine Abstraktion bewirkt hat.

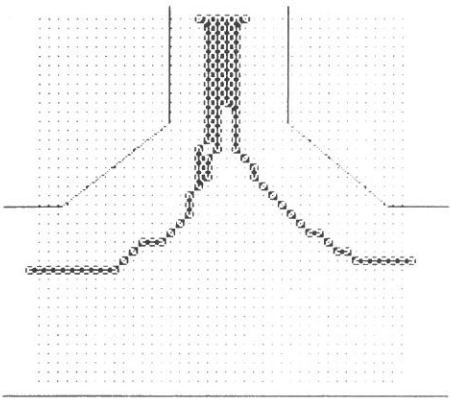

Abb. 7: Skelett für TAF in Abbildung 4

Abbildung 7 zeigt das Skelett zu dem TAF aus Abbildung 4. Der TAF wurde dreimal verwaschen und einmal mit der Konvergenzoperation bearbeitet. Die abstrahierende Wirkung der Konvergenzoperation wird deutlich (vergleiche mit 6b). Es sind nur die wesentlichen Verläufe übriggeblieben ('Rechtsabbiegen' und 'Linksabbiegen').

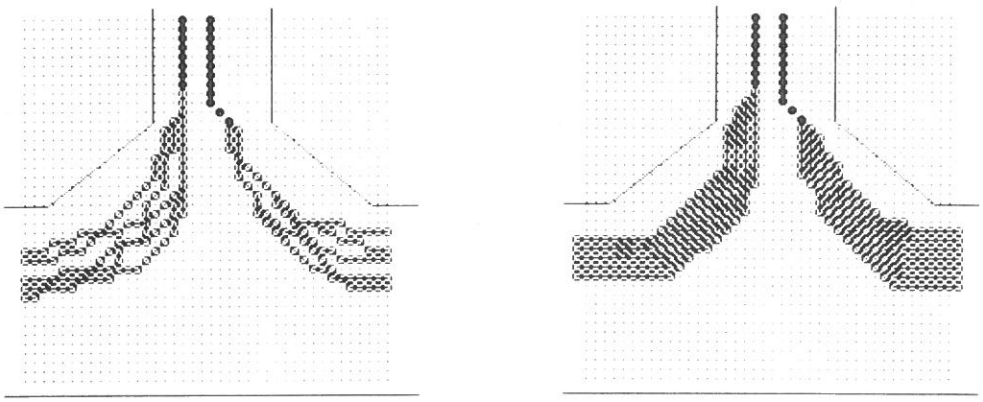

Abb. 8a und 8b: Vorhersagen der wahrscheinlichsten zu erwartenden Trajektorien ausgehend von verschiedenen Startpunkten

Die Abbildungen 8a und 8b zeigen wesentliche Ergebnisse dieses Kapitels. In 8a wurde der gleiche TAF, wie in Abbildung 2a benutzt, während in 8b zunächst

dreimal verwaschen wurde. Ausgehend von verschiedenen Startpunkten wurden diejenigen Bereiche berechnet, die für den weiteren Verlauf der Trajektorien am wahrscheinlichsten sind. Die Abbildungen geben die örtlichen Beschränkungen an. Es wurden ebenfalls Beschränkungen für das Geschwindigkeitsverhalten berechnet, welche in dieser Darstellung allerdings nicht sichtbar sind. TAFs erlauben also einen Richtungs- und Geschwindigkeitsfilter für die Bewegungsanalyse in Bildfolgen zu berechnen.

Zusammenfassung

Wir haben das Problem der Repräsentation von beobachteten Trajektorien untersucht mit dem Ziel, typisches Verhalten explizit zu machen und Vorhersagen über Verhalten aufgrund von Erfahrungen zu ermöglichen. Wir haben TAFs als Lösung vorgeschlagen und haben gezeigt, daß diese Repräsentation einige interessante Eigenschaften hat.

- In TAFs kann zwischen mehr oder weniger typischem Verhalten unterschieden werden; eine wichtige Dimension der Erfahrung wird damit adäquat behandelt. Im Kontext 'Top-down' gesteuerter Bildanalyse können z.B. TAF-Bereiche mit hoher Wahrscheinlichkeit dazu benutzt werden ein Suchgebiet für die Ebenen der niederen Bilddeutung zu definieren. Wir haben gezeigt, daß sich ein Richtungs- und Geschwindigkeitsfilter für eine Steuerung von niedrigen Ebenen der Bewegungsanalyse mithilfe der TAFs effektiv berechnen läßt (Abbildungen 8a und 8b stellen die Suchbereiche für zwei konkrete Beispiele graphisch dar).

- TAFs liefern einen natürlichen Übergang von einzelnen zu akkumulierten Erfahrungen. Dadurch ergibt sich eine natürliche Repräsentation von prototypischem Wissen.

- Ein TAF kann für die Generalisierung und für die Abstraktion von Beobachtungen genutzt werden. Dies ist ein notwendiger Schritt, um Erfahrung auf neue aber ähnliche Situationen anwendbar zu machen.

- Ein TAF kann durch sein Skelett und die dadurch definierten

Prototypen charakterisiert werden. Dies liefert u.a. die Möglichkeit, über prototypisches Verhalten Schlüsse zu ziehen.

In Abschnitt 4 wollen wir uns der Frage zuwenden, wie man die in den TAFs enthaltene Information an leicht veränderte Situationen anpassen kann, um z.B. in einer Abbiegesituation auftretende Hindernisse adäquat behandeln zu können.

4. Anpassung der TAFs an veränderte Situationen

Generalisierungsfähigkeit und Abstraktionsfähigkeit sind zwei wichtige - wie bereits ausgeführt - funktionale Anforderungen an das Modell zur Repräsentation beobachteter Trajektorien. Im vorherigen Abschnitt wurde gezeigt, wie Informationen in den TAFs generalisiert werden, um auch Aussagen in solchen Gebieten berechnen zu können, in denen keine direkten Erfahrungen vorliegen, und wie wesentliche Informationen aus den angehäuften Einzelbeobachtungen abstrahiert werden können.

An dieser Stelle wird diskutiert, wie die TAFs an leicht veränderte Situationen angepaßt werden können. Wenn z.B. an einer Kreuzung, für die Erfahrung in Form eines TAF vorliegt, ein Hindernis steht, sollen nur Trajektorien vorhergesagt werden, die auf diese Situation passen. Vorhersagen, welche auf das Hindernis zusteuern, wären unplausibel und sollten verhindert werden.

Das Anpassen von bestehenden Konzepten an leicht veränderte Situationen scheint auch im kognitiven System von Menschen eine Rolle zu spielen. Bei drastischeren Änderungen des Kontextes sollten aber eher andere konzeptuelle Einheiten berücksichtigt werden, die dem geänderten Kontext gerechter werden.

TAF mit Randbedingungen

Wir passen TAFs an veränderte Situationen (Hindernisse) mithilfe von Randbedingungen an. Wir verstehen in diesem Zusammenhang unter Hindernissen alle 4-dimensionalen Gebiete eines TAF, für die einschränkende Bedingungen

gelten. Dabei wird angenommen, daß diese Einschränkungen in der Phase der Modellbildung (Lernphase) nicht gegolten haben.

Randbedingungen sind also Gebiete des 4-dimensionalen TAF, in denen keine Aktivitäten zulässig sind und durch die konsequenterweise keine Trajektorien passieren können. Hindernisse können rein örtlich auftreten (sämtliche Geschwindigkeiten sind in einem bestimmten xy-Gebiet unzulässig, weil ein LKW auf der Strasse parkt), als auch geschwindigkeitsabhängig sein (bestimmte Geschwindigkeiten sind z.B. bei Glatteis in der Kurve unplausibel, unabhängig vom Ort).

Ein Hindernis wird in einen TAF eingeführt, indem im Hindernisgebiet alle Zähler auf Null gesetzt werden. Es ist aber unzureichend, erst direkt am Hindernis über das Hindernis zu erfahren (z.B. bei einer Vorhersage). Die Information muß zumindest soweit durch das Aktivierungsfeld propagiert werden, daß auf das Hindernis zuführende Trajektorien unmöglich werden. Wir schlagen eine Inhibitionsoperation vor, welche die Hindernisinformation geeignet durch das Netzwerk propagiert.

Inhibition

Die Inhibitionsoperation ist eine Operation, die Informationen von einem Hindernisgebiet eines TAF zu den Nachbarn inhibitorisch propagiert. Anschaulich beschrieben werden bei der Inhibition diejenigen Trajektorien aus dem TAF herausgenommen, die durch oder gegen das Hindernis führen würden.

Eine Zelle S wird durch Nullsetzten ihres Zählers unter folgenden Bedingungen inhibiert :

- Die Zähler aller Zellen in ihrer qualifizierten 4D-Nachbarschaft (s.o) sind Null,

- oder die Zähler aller Zellen, aus denen S erreicht werden kann sind Null.

Dieses Verfahren wird auf alle Zellen des TAF angewandt, bis keine Veränderung mehr stattfindet. Der Algorithmus sichert, daß sukzessive alle Zellen auf Null gesetzt werden, die nicht erreicht werden können (keine aktiven Vorgänger haben), oder von denen aus keine Trajektorien fortgeführt werden können (keine aktiven Nachfolger). Wir demonstrieren die Inhibition anhand folgender Beispiele.

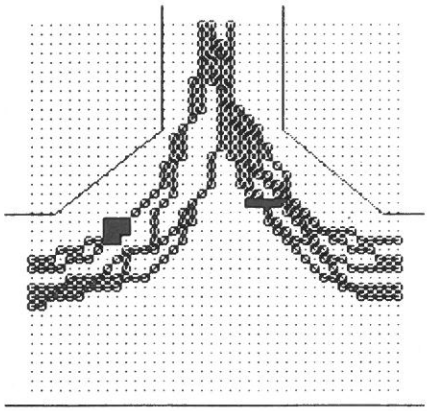

Abb. 9a: Skelett zu 2a mit Hindernis

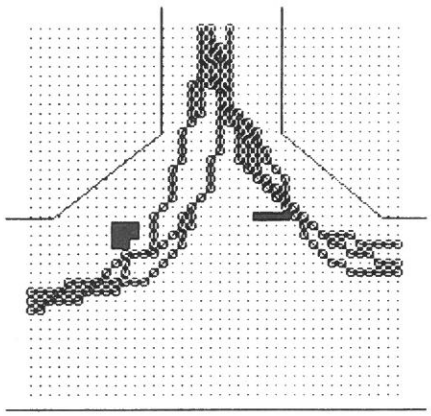

Abb. 9b: Skelett nach Inhibition

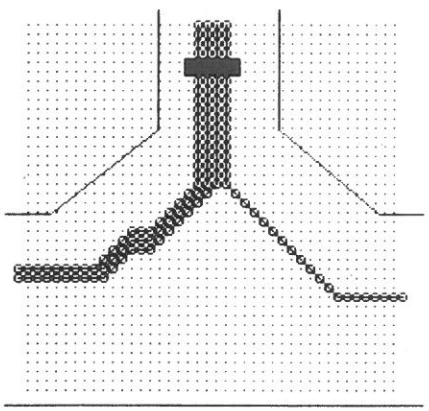

Abb. 9c: Skelett zu 2a nach dreimaligem Verwaschen

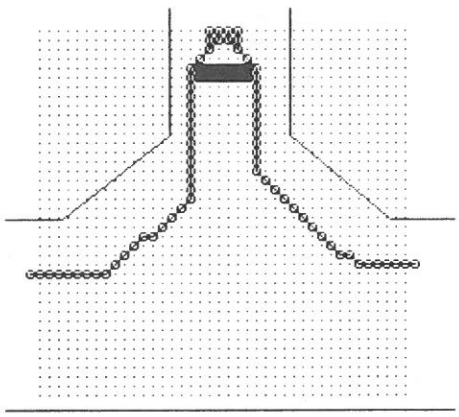

Abb. 9d: Skelett nach Inhibition

Abbildung 9a zeigt das Skelett zum TAF von 2a (ohne Verwaschungen) mit einem

überlagerten Hindernis. In Abbildung 9b wurde die Inhibitionsoperation angewandt. Es ist wieder ein 'glattes' Aktivierungsmuster entstanden. Man kann sagen, daß alle für diese Situation unbrauchbaren Erfahrungen herausgenommen wurden.

9c zeigt ein weiteres Beispiel. Der TAF wurde dreimal verwaschen. Das Hindernis wurde so gewählt, daß keine der ursprünglichen relativen Maxima erhalten bleiben. Nach Anwendung der Inhibition (in 9d) sind neue relative Maxima entstanden, die für eine Vorhersage benutzt werden können.

Zusammenfassung

In diesem Abschnitt wurde ein Verfahren vorgestellt, das es erlaubt, Erfahrungen auf veränderte Situationen anzuwenden. Am Beispiel von Hindernissen in einer Strassenverkehrsszene wurde gezeigt, wie eine Hindernisumfahrung berechnet werden kann. Dabei ist es möglich, mithilfe von Generalisierungen und Anpassungen Erwartungen (Trajektorienvorhersagen) zu generieren, welche so nicht vorher erfahren wurden. TAFs erweisen sich dadurch als ein flexibler Repräsentationsmechanismus.

5. Diskussion und Ausblick

Wir haben mit den 'Trajectory Accumulation Frames' ein szenenhaftes Modell für zeitabhängige Ereignisse vorgelegt. Ein TAF ist ein erfahrungsbasierter Repräsentationsformalismus, der es u.a. erlaubt:

- Vorhersagen über typisches Verhalten zu machen,
- unscharfes Wissen zu repräsentieren,
- konkrete Erfahrungen zu generalisieren,
- von Einzelheiten zu abstrahieren,
- und Erfahrungen auf veränderte Situationen anzuwenden.

Diese funktional wichtigen Eigenschaften wurden erreicht durch einen lokalen Vorhersagemechanismus und die Anwendung lokaler Operationen wie Verwaschung, Konvergenz und Inhibition.

Besonders im Kontext natürlichsprachlich gesteuerter Szenenanalyse erweisen TAFs sich als flexibler Formalismus. Ausgehend von der verbzentrierten Tiefenstruktur einer Äußerung gelingt es, geometrische und zeitliche Randbedingungen für eine niedere Bilddeutung zu berechnen. Die gewonnene Information kann direkt als Filter z.B. in eine Bewegungsanalyse eingehen.

Die Orientierung an kognitiv plausiblen Repräsentationen und Prozessen erweist sich als zweckdienlich. Insbesondere hat die Forderung nach Erlernbarkeit des Modells zu einem erfahrungsbasierten Verfahren geführt, bei dem keine willkürliche Trennung zwischen konkreten Beispielen und prototypischem Wissen vorgenommen wird.

Durch die beschriebenen Arbeiten wurden weitere grundlegende Fragestellungen aufgeworfen, welche z.Zt. über den Rahmen des Projektes hinausgehen.

- Wie kann man von der stationären Umgebung abstrahieren, d.h. wie können TAFs transformiert werden, um bei stärker veränderten Situationen anwendbar zu werden? Wir denken z.B. an die Anwendung eines für eine bestimmte Kreuzung entstandenen TAF auf eine neue Kreuzung. Dabei muß z.B. geklärt werden, ob die vorgeschlagenen Verfahren zur Hindernisbewältigung verwandt werden können.

- Inwieweit muß in bestimmten Situationen ein grösserer örtlicher und zeitlicher Zusammenhang berücksichtigt werden (der Vorhersagealgorithmus berücksichtigt bisher nur den lokalen Kontext)?

- Sollten die 4-dimensionalen Zustandsbeschreibungen um weitere Dimensionen erweitert werden, um eine reichhaltigere Situationsbeschreibung zu ermöglichen?

6. Literatur

BINFORD 82

Survey of Model-Based Image Analysis Systems, Thomas O. Binford, The International Journal of Robotics Research, Vol.1, Nr.1, 18-64, Spring 1982

BLOCK 81

Imagery, N. Block (ed.), MIT Press, Cambridge/Mass., 1981

FINKE 80

Levels of Equivalence in Imagery and Perception, R.A. Finke, Psychological Review 1980, Vol.87, Nr.2, 113-132

FINKE 85

Theories Relating Mental Imagery to Perception, R.A. Finke, Psychological Bulletin 1985, Vol.98, Nr.2, 236-259

KOSSLYN 80

Image and Mind, Stephen M. Kosslyn, Harvard University Press 1980

NEUMANN and NOVAK 83

Event Models for Recognition and Natural Language Description of Events in Real-World Image Sequences, B. Neumann und H.-J. Novak, Proc. IJCAI-83, Karlsruhe 1983, pp. 724-726

NEUMANN 84

Natural Language Description of Time-Varying Scenes, B. Neumann, Technical Report, University of Hamburg, IfI-HH-B-105/84, 1984

NOVAK and NEUMANN 86

Text Generation based on Visual Data: Descriptions of Traffic Scenes, H.-J. Novak und B. Neumann, Proc. of the 2nd International Conference on Artificial Intelligence, Methodology, Systems, Applications, September 1986, Varna, Bulgaria

Danksagung

Dieses Projekt wird durch die Deutsche Forschungsgemeinschaft unterstützt.