

**Eine hybride Repräsentation von  
Objektbewegungen: von analogen  
zu propositionalen Beschreibungen**

Michael Mohnhaupt

FBI-HH-M-177/90

März 1990

## Zusammenfassung

Für ein universelles intelligentes System ist eine adäquate Modellierung von Objektbewegungen ein wichtiger Bestandteil. Die verschiedenen zu lösenden Aufgaben im Zusammenhang mit Objektbewegungen, z.B. Pfadplanung, raumzeitliches Schließen, Ereigniserkennung und Lernen von Objektbewegungen werden in der Literatur meist getrennt behandelt. In diesem Beitrag wird dagegen ein hybrides Repräsentationssystem diskutiert, welches je nach Anforderung unterschiedliche Repräsentationsformate bereitstellt. Das Repräsentationssystem besteht einerseits aus einer 'klassischen' propositionalen Langzeitrepräsentation und andererseits aus einer analogen quantitativen Kurzzeitrepräsentation, welche bei Bedarf instanziiert wird. Es wird insbesondere gezeigt, wie abstrakte propositionale Modelle aus konkreten Beobachtungen gelernt werden können. Beobachtete Beispiele werden zunächst in der analogen Repräsentation gespeichert und verarbeitet. Anschließend wird unter Ausnutzung von perzeptuellen Primitiven eine propositionale Beschreibung berechnet.

Dieser Beitrag erscheint ebenfalls in 'Repräsentation und Verarbeitung räumlichen Wissens', C. Freksa & C. Habel (Hrsg.), Springer-Verlag 1990.

## **Abstract**

An important task for intelligent systems is the adequate modeling of object motion. Different tasks related to object motion, e.g. pathplanning, spatiotemporal reasoning, event recognition, and learning of object motion, are often treated separately. In this paper, we discuss a hybrid representational system, in which different modes of representation are exploited. The representational system includes a propositional abstract long-term representation, and an analog quantitative short-term representation, which is instantiated on demand. We show how propositional models can be learned from concrete observations. First, observed examples are stored and processed using the analogical representation, and second, propositional description are computed by exploiting perceptual primitives.

This paper also appears in 'Repräsentation und Verarbeitung räumlichen Wissens', C. Freksa & C. Habel (Ed.), Springer-Verlag 1990.

# Eine hybride Repräsentation von Objektbewegungen: Von analogen zu propositionalen Beschreibungen

Michael Mohnhaupt

Universität Hamburg, Fachbereich Informatik

Arbeitsbereich 'Kognitive Systeme'

Bodenstedtstr. 16, D-2000 Hamburg 50

mohnhaupt@rz.informatik.uni-hamburg.dbp.de

## 1 Einleitung

Das 'Verstehen' von Objektbewegungen ist ein wichtiger Bestandteil höherer kognitiver Prozesse, sowohl für biologische, als auch für maschinelle Systeme. Ein allgemeines Ziel einer adäquaten Modellierung von Objektbewegungen ist möglichst große Vorhersagekraft. Sie ist z.B. erforderlich, um zeitabhängige Ereignisse zu erkennen, für eine Bewegungsplanung, um Hindernissen auszuweichen und um Schlüsse über zeitveränderliche Umgebungen zu ziehen. Bei geeignetem Vorwissen über eine Szene ist damit auch eine Steuerung und Verbesserung visueller Prozesse möglich.

Entscheidende Information über Objektbewegungen liefern die Prozesse der niederen Bilddedeutung. Im Folgenden wird davon ausgegangen, daß die einzelnen Objekte einer Szene bereits erkannt sind, und daß ein 'Verstehen' von Objektbewegungen, also zeitübergreifenden Zusammenhängen, darauf aufbauen kann. Damit sind einige spezielle Situationen ausgeklammert, in denen eine Objekterkennung erst aufgrund des zeitlichen Verhaltens des Objektes möglich ist.

Eine schwierige Frage betrifft die geeignete Repräsentation für Objektbewegungen und die dazugehörigen Prozesse. In diesem Beitrag wird ein hybrider Ansatz vertreten, d.h. es werden verschiedene Repräsentationen und verschiedene Arten von Prozessen verwendet, um die unterschiedlichen Aufgaben im Zusammenhang mit Objektbewegungen zu bewältigen. Eine einzelne Repräsentation ist nicht flexibel genug, um die sehr unterschiedlichen Aufgaben im Zusammenhang mit Objektbewegungen zu bewältigen. Dies wird gestützt durch theoretische Resultate über die begrenzte Behandelbarkeit von wichtigen Problemen in einem rein propositionalen und logik-basierten Ansatz (siehe z.B. *Levesque 86*). Auch in *Lindsay 88* wird für unterschiedliche Repräsentation mit unterschiedlichen Inferenzmechanismen argumentiert.

Für das Erkennen von zeitübergreifenden Ereignissen und für eine Langzeitspeicherung werden propositionale qualitative Modelle vorgeschlagen, auf denen ein logik-ähnlicher Inferenzmechanismus abläuft. Für wichtige Aspekte des Lernens, für Visualisierungen und für Probleme des raum-zeitlichen Schließens wird eine analoge quantitative Repräsentation vorgeschlagen, welche bei Bedarf instantiiert werden kann. Die darauf ablaufenden Prozesse sind einfach, lokal und parallel.

Das vorgeschlagene Modell wird einerseits durch eigene Untersuchungen motiviert,

andererseits wird dabei versucht, bestehende Untersuchungen zu verschiedenen Einzelaspekten von Objektbewegungen in einem kohärenten Ansatz zu integrieren. Dabei werden sowohl Arbeiten über maschinelle Modellierung von Objektbewegungen berücksichtigt, als auch Arbeiten über mögliche mentale Repräsentationen von Objektbewegungen und den darauf ablaufenden Prozessen. Die hybride Repräsentation von Objektbewegungen wird in Abschnitt 2 diskutiert.

Abschnitt 3 befaßt sich mit wichtigen Aspekten des Zusammenwirkens beider vorgeschlagener Repräsentationen. Insbesondere wird untersucht, inwieweit die abstrakten propositionalen Ereignismodelle aus konkreten Beobachtungen und deren Modellierung in einem temporären raum-zeitlichen Puffer abgeleitet werden können. Es zeigt sich, daß lokale Prozesse im raum-zeitlichen Puffer gut dazu geeignet sind, räumliche und zeitliche Beziehungen zwischen Objekten zu berechnen. Raum-zeitliche Beziehungen zwischen Objekten bilden eine Menge von perzeptuellen Primitiven. Eine Untermenge dieser Primitive wird zur Beschreibung eines jeweiligen Ereignisses herangezogen, nämlich diejenigen, die invariant für einen bestimmten Ereignistyp sind. Eine qualitative Beschreibung der zunächst quantitativ beschriebenen Primitive führt dann zu propositionalen Modellen. Damit wird eine Verbindung von Ergebnissen visueller Prozesse bis hin zu sprachlich orientierten Modellen gezogen. Wichtige Bestandteile des Modells werden dabei durch experimentelle Ergebnisse gestützt.

## **2 Eine hybride Repräsentation von Objektbewegungen**

In diesem Abschnitt wird eine hybride Repräsentation für die Modellierung von Objektbewegungen skizziert. Sie besteht im wesentlichen aus einer quantitativen analogen Kurzzeit-Repräsentation, einem raum-zeitlichen Puffer, und einer abstrakteren, propositionalen und qualitativen Langzeit-Repräsentation.

Bisher wurden verschiedene Aspekte von Objektbewegungen meistens isoliert betrachtet und in unterschiedlichen Teilgebieten der KI behandelt, z.B. in der höheren Bilddeutung, bei der Pfadplanung, oder im Bereich raum-zeitlichen Schließens. Hier wird versucht, die verschiedenen Arbeiten in einem Ansatz zu integrieren. Außerdem werden Erkenntnisse über mentale Repräsentationen berücksichtigt. Der Ansatz liefert einen Rahmen für weitere detailliertere Untersuchungen zur Modellierung von Objektbewegungen.

Die Abbildung 1 zeigt die wichtigsten Komponenten zur Modellierung von Objektbewegungen. Auf der linken Seite der Abbildung sind verschiedene Repräsentationsformate mit einer kurzen Charakterisierung und assoziierten Prozessen aufgeführt. Auf der rechten Seite stehen Einzelaufgaben, welche mithilfe der Repräsentationen gelöst werden können.

### **• Niedere Bilddeutung**

Beginnend am unteren Ende der Abbildung, sind zunächst die Prozesse der niederen Bilddeutung und der Objekterkennung zu finden. Ausgehend von einer Bildfolge werden relevante Informationen über Form, Position und Identität der beteiligten Objekte berechnet. Im weiteren wird davon ausgegangen, daß eine

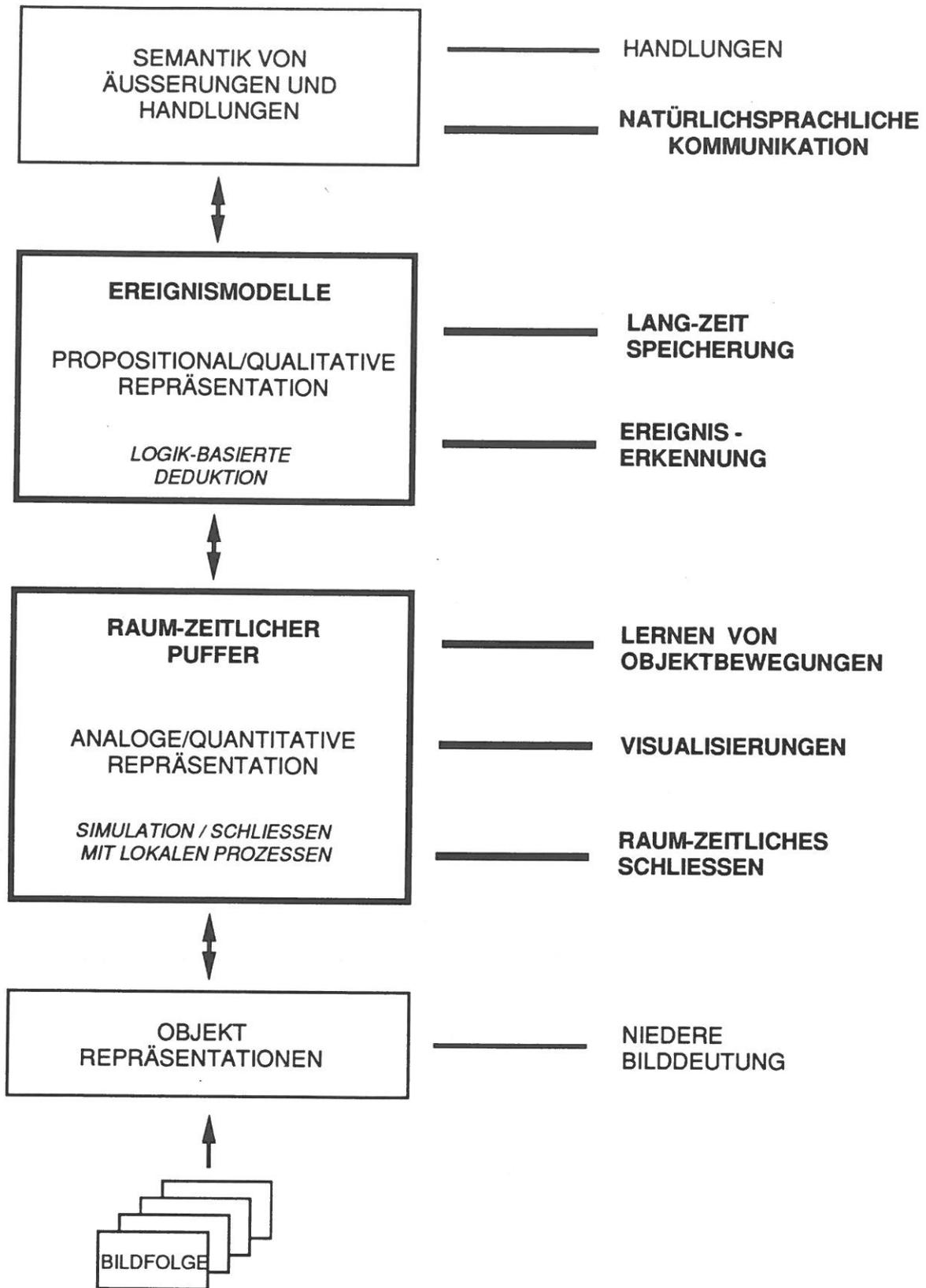


Abbildung 1: Eine hybride Repräsentation von Objektbewegungen

Objektrepräsentation vorliegt und die nachfolgenden Prozesse darauf aufbauen können.

- **Raum-zeitliches Schließen**

Verschiedene Autoren modellieren Teile des raum-zeitlichen Schließens unter Benutzung einer analogen quantitativen Repräsentation. In *Funt 80* kann das Zusammentreffen von fallenden Objekten vorhergesagt werden, basierend auf einem 2-dimensionalen Feld und lokalen Operationen, um Bewegung zu simulieren. Auch *Gardin + Meltzer 89* verwenden eine analoge Repräsentation, um physikalische Effekte qualitativ zu modellieren. Das Verhalten von nicht starren Objekten und Flüssigkeiten wird durch lokale Interaktionsregeln in einem 2-dimensionalen Feld modelliert. Inferenzen können durch Simulation berechnet werden. *Steels 88* schlägt für Pfadplanungen eine analoge Repräsentation und ein Reaktions-Diffusionsmodell für die darauf arbeitenden Prozesse vor. In unseren Arbeiten (z.B. *Neumann + Mohnhaupt 88*, *Mohnhaupt + Neumann 90*) werden eine analoge 4-dimensionale Repräsentation und lokale Prozesse verwendet, um das Verhalten von Objekten in Straßenverkehrsszenen vorherzusagen. Die Modelle erlauben die Bewältigung von einfachen Hindernissen und Vorhersagen über das Zusammentreffen von Objekten.

- **Visualisierungen**

Einige Forschungen beschäftigen sich mit der Modellierung von räumlichen Beziehungen zwischen Objekten, dem Verstehen von abstrakten Beschreibungen (z.B. von geometrischen Figuren) und dem Verstehen von Sprache unter Benutzung einer quantitativen räumlichen Repräsentation. Durch Visualisierung des Inhalts der Beschreibungen in einem 2-dimensionalen (oder mehrdimensionalen) Puffer können Inferenzen vereinfacht werden, kann vorher implizite Information explizit gemacht werden, und kann Konsistenz geprüft werden (siehe z.B. *Gelernter 63*, *Waltz + Boggess 79*, *Adorni and Di Manzo 83*, *Pribbenow 90*, *Khenkhar 90*). *Neumann + Novak 86* kommen zum Ergebnis, daß Visualisierungen ebenfalls benötigt werden für adäquate Hörermodelle in einigen Domänen. In unseren Arbeiten werden Visualisierungen benutzt zur Berechnung von raum-zeitlichen Beziehungen, die vorher nur implizit vorhanden waren (siehe Abschnitt 3), zur Vorhersage von typischen Bewegungen und zur Steuerung der niederen Bilddeutung (siehe *Mohnhaupt + Neumann 90*, *Mohnhaupt + Fleet 88*). Der raum-zeitliche Puffer kann dabei sowohl von visuellen Prozessen, als auch von höheren kognitiven Prozessen beeinflusst werden. Er dient als gemeinsame Repräsentation. Dies ist verträglich mit psychologischen Untersuchungen (siehe z.B. *Finke 85*), nach denen auch bei Menschen eine gemeinsame bildhafte Repräsentation für visuelle und höhere kognitive Prozesse existiert.

- **Lernen von Objektbewegungen**

In *Mohnhaupt + Neumann 89* und *Mohnhaupt + Neumann 90* werden verschiedene Aspekte des Lernens von Objektbewegungen untersucht, unter Benutzung einer quantitativen 4-dimensionalen Repräsentation von Bewegungsverläufen

(zwei Orts- und zwei Geschwindigkeitskoordinaten). Ausgehend von diesen elementaren physikalischen Größen, welche zunächst beobachtete Ereignisinstanzen beschreiben, können typische Objektbewegungen gelernt werden. Dabei werden einfache lokale Operationen verwendet. Generalisierte Objektbewegungen können anschließend auch auf Situationen angewendet werden, für die keine direkten Beobachtungen vorliegen. Auf weitere wichtige Aspekte des Lernens innerhalb einer Pufferrepräsentation und anschließende Überführung in propositionale Beschreibungen wird im nächsten Abschnitt detaillierter eingegangen. Dazu werden perzeptuelle Primitive wie Distanzen und relative Orientierungen durch einfache sich ausbreitende Aktivierungsprozesse innerhalb des analogen quantitativen Puffers berechnet.

- **Langzeit-Speicherung**

Eine weitere wichtige Aufgabe ist die Langzeit-Speicherung von Objektbewegungen. Es werden abstrakte propositionale Beschreibungen dafür verwendet. Der Hauptgrund dafür ist deren Effizienz. *Neumann + Novak 83*, *Neumann + Novak 86* und *Andre + Bosch + Herzog + Rist 86* schlagen propositionale Ereignisbeschreibungen für eine Langzeit-Speicherung von Objektbewegungen in Straßenverkehrs- und Fußballszenen vor. Dies ist auch verträglich mit psychologischen Untersuchungen (siehe z.B. *Marschark 88*). Danach gibt es keine Evidenz dafür, daß bildhafte analoge Repräsentationen im Langzeitspeicher gehalten werden. Die Untersuchungen sprechen dafür, daß abstraktere propositionale Beschreibungen im Langzeitspeicher verwendet werden. *Kosslyn 80* schlägt in seinem Modell ebenfalls propositionale Beschreibungen für eine Langzeitspeicherung von Objekten und deren raum-zeitlichen Beziehungen vor.

- **Erkennung von Objektbewegungen und natürlichsprachliche Kommunikation**

Propositionale Beschreibungen haben sich ebenfalls als nützlich erwiesen bei der Erkennung von Ereignissen, z.B. von Objektbewegungen in Straßenverkehrsszenen (*Neumann + Novak 83*, *Neumann + Novak 86*), oder bei der inkrementellen Erkennung von Spielerbewegungen in Fußballszenen (siehe *Andre + Bosch + Herzog + Rist 86*). Ereignismodelle beschreiben hierbei die charakteristischen Eigenschaften von Objekttrajektorien mithilfe von Prädikaten, welche erfüllt sein müssen, damit z.B. ein 'überholen' oder 'abbiegen' erkannt werden kann. Die propositionale Modellierung gestattet eine effiziente Erkennungsstrategie. In *Mohnhaupt + Neumann 90* wird vorgeschlagen, für eine detailliertere Ereigniserkennung instantiierte Visualisierungen zu Hilfe zu nehmen, um z.B. nach der Groberkennung eines Überholvorganges mit propositionalen Modellen zu entscheiden, wie typisch oder atypisch das erkannte Ereignis ist. Dabei muß mit der propositionale Repräsentation Information assoziiert werden, mit der ein Puffer instantiiert werden kann. Dazu gehören die relevanten Dimensionen und eine Beschreibung typischer Verläufe oder prominenter Beispiele, mit denen der Puffer gefüllt, und eine Simulation gestartet werden kann.

Die propositionalen Modelle können Ausgangspunkt für die natürlichsprachliche Beschreibung einer Szene sein. *Neumann + Novak 86* füllen mit instantiierten

Ereignismodellen Kasusrahmen als Tiefenstruktur für natürlichsprachliche Äußerungen. Anschließend können damit kohärente Szenenbeschreibungen generiert werden (Novak 87).

Es ist den meisten genannten Untersuchungen gemeinsam, daß räumliche Beziehungen innerhalb eines analogen Puffers repräsentiert werden. Dadurch sind sie einfach berechenbar mit lokalen Operationen. Die Dimensionalität der analogen Repräsentation differiert in verschiedenen Arbeiten (siehe eine detaillierte Diskussion in Pinker 88). Physikalische Plausibilität wird einerseits durch intrinsische Eigenschaften des Puffers (z.B. nur ein Objekt pro Ort) gewährleistet, andererseits dadurch, daß die Repräsentation auf Beispielen basiert und damit auf physikalisch möglichen Bewegungen beruht. In den genannten Ansätzen zum raum-zeitlichen Schließen wird eine Simulation innerhalb eines Puffers ausgenutzt, um Vorhersagen zu berechnen und Inferenzen zu ziehen. Außerdem besteht Einigkeit darüber, daß neben einer quantitativen analogen Repräsentation auch qualitative propositionale Modelle benötigt werden, für eine Langzeit-Speicherung und eine effiziente Ereigniserkennung.

### 3 Von analogen zu propositionalen Beschreibungen

Ein wichtiges Kriterium für die Leistungsfähigkeit einer hybriden Repräsentation ist das Zusammenspiel der Einzelkomponenten. Außerdem muß davon ausgegangen werden, daß eine plausible und flexible Repräsentation von Gegenständen und Ereignissen der visuellen Welt aus konkreten Beobachtungen erlernbar und durch sie veränderbar sein muß.

Angewendet auf die im letzten Abschnitt vorgestellte hybride Repräsentation für Objektbewegungen bedeutet dies, daß eine Modellierung im raum-zeitlichen Puffer ableitbar sein muß aus Informationen, welche von der niederen Bilddeutung geliefert werden, und daß die qualitative propositionale Ereignisrepräsentation aus Information ableitbar sein muß, die im raum-zeitlichen Puffer angehäuft worden ist. Dieser Abschnitt beschäftigt sich mit diesen beiden Aspekten der hybriden Repräsentation.

Angenommen, ein kognitives System beobachtet Verkehr in Straßenszenen. Die wahrgenommenen Daten bestehen u.a. aus einer großen Anzahl von Objekttrajektorien, welche verschiedene Ereignisse charakterisieren und aus aktuellen oder früheren Beobachtungen bestehen können. Es wird angenommen, daß die Objekttrajektorien in raum-zeitlichen Koordinaten von der niederen Bilddeutung geliefert werden. Außerdem wird zunächst ein stationärer Beobachter und ein bekannter stationärer Hintergrund angenommen. Ein sinnvolles Resultat der Beobachtungen ist eine kompakte Beschreibung der verschiedenen Ereignisklassen (z.B. überholen, abbiegen, parken), um damit Vorhersagen generieren zu können und Inferenzen ziehen zu können.

Es ist plausibel, auf dieser Ebene eine analoge quantitative Repräsentation zu benutzen, weil kein spezifisches a priori Wissen angenommen wird, welches sofort eine abstraktere Beschreibung ermöglichen könnte. Daher werden zunächst elementare physikalische Größen, wie Position und Geschwindigkeit der Objekte betrachtet. Es wird eine Repräsentation benutzt, welche bezüglich dieser Dimensionen analog ist, um zunächst

soviel Information wie möglich zu konservieren. Erst nach der Aufnahme von mehreren Ereignisinstanzen in einer Szene kann Information für weitere Abstraktionen berechnet werden (siehe Abschnitt 3.1). Die Anhäufung von Beispielen im raum-zeitlichen Puffer führt zu einer natürlichen Repräsentation von akkumulierten Beobachtungen. Ähnlichkeit von Beispielen wird in der Repräsentation durch Nähe erfaßt. Daher sind Abstraktionen durch lokale Operationen berechenbar.

Ein nächster Schritt besteht darin, szenenunabhängige Ereignisbeschreibungen zu berechnen, z.B. ein generisches Modell für 'abbiegen', welches unabhängig von einer bestimmten Kreuzung ist. Hierzu werden invariante Ereigniseigenschaften aus einer Basismenge von perzeptuellen Primitiven bestimmt (siehe Abschnitt 3.2). Eine qualitative Beschreibung der invarianten Ereigniseigenschaften kann dann zu einem propositionalen Ereignismodell führen (siehe Abschnitt 3.3).

### 3.1 Puffer im Trajektorienanhäufungsmodus

In diesem Abschnitt wird zum Lernen von Objektbewegungen die Anhäufung von Einzeltrajektorien im analogen raum-zeitlichen Puffer diskutiert. Es wird hier ein beispielbasierter Ansatz vertreten. Eine detaillierte Beschreibung dieser Untersuchungen ist in *Mohnhaupt + Neumann 90* zu finden. Andere in der Literatur vorgeschlagene beispielbasierte Ansätze untersuchen primär propositionale Repräsentationen (siehe *Bradshaw 87*, *Stanfill + Waltz 86* und *Kibler + Aha 87*). Es ist bemerkenswert, daß beispielbasierte Ansätze für wichtige Teile der menschlichen Konzeptbildung favorisiert werden (siehe z.B. *Smith + Medin 81*).

Im Trajektorienanhäufungsmodus besteht der raum-zeitliche Puffer aus einem 4-dimensionalen Feld ( $x$ ,  $y$ , und zwei Geschwindigkeitskoordinaten  $v$ ,  $b$ ), das einen bestimmten Ausschnitt der  $xy$ -Ebene bedeckt (einen Teil einer Szene). In der Domäne Straßenverkehr wird auf die dritte räumliche Dimension verzichtet, weil es sich um Bewegungen in einer Ebene handelt.

Im Puffer existieren für jedes  $xy$ -Paar Zählerzellen für alle möglichen Geschwindigkeitswerte, jeweils repräsentiert durch Betrag  $v$  und Richtung  $d$ . Der Zustandsvektor  $S = (x, y, v, d)$  beschreibt den Zustand eines Objektes zu einer bestimmten Zeit. Die vier Größen können durch einen Beobachter wahrgenommen werden und erfordern nur elementares Wissen über Positionen und die Änderung von Positionen identifizierter Objekte.

Für jede beobachtete Trajektorie wird eine Spur von Zustandsvektoren registriert durch Inkrementierung der entsprechenden Zähler. Falls mehrere Objekte beobachtet werden, werden viele (möglicherweise dieselben) Zellen inkrementiert, ohne zwischen verschiedenen Objekten zu unterscheiden. Eine Trajektorie wird dabei diskretisiert entsprechend der Auflösung in den verschiedenen Dimensionen des Puffers. Die Repräsentation führt automatisch zu Gebieten mit höherer Evidenz ohne extra Berechnungen. Die Repräsentation kann auf allen beobachteten Beispielen basieren, oder auf einer Menge von aktuellen Beispielen. Im zweiten Fall werden alle Zellen kontinuierlich dekrementiert, wenn neue Beispiele gespeichert werden. Alte Information wird dadurch gewissermaßen mit der Zeit 'vergessen'.

Nach vielen Einzelbeispielen kann der Puffer als Dichtefeld angesehen werden, indem hohe Werte Beobachtungen repräsentieren, die durch viele Beispiele gestützt ist.

Pfade entlang von lokalen Dichtemaxima definieren ein Muster typischer Erfahrungen, welches hier **Skelett** des Puffers genannt wird. Die Verteilung lokaler Maxima ist abhängig von der Auflösung, in welcher der Puffer betrachtet wird. Es werden zwei lokale Operationen verwendet, um einfache Generalisierungen und Abstraktionen zu berechnen. Eine **Verwaschungsoperation**, welche die Zählerwerte in eine lokale Nachbarschaft verteilt. Dadurch wird Erfahrungen auf Trajektorien extrapoliert, welche den beobachteten Beispielen bezüglich der vier Dimensionen des Puffers ähnlich sind. Zusätzlich wird eine lokale **Konvergenzoperation** benutzt, welche lokal hohe Zählerstände stützt und niedrige Zählerstände unterdrückt. Ein Skelett repräsentiert nach Anwendung der Konvergenzoperation eine Abstraktion der Beispiele, da es nur die wesentlichen, am meisten beobachteten Verläufe enthält.

In Abbildung 2 wurden 10 Abbiegetrajektorien im Puffer gespeichert. Links sind zwei Dimensionen ( $x, y$ ) des Puffers zu sehen; rechts ist das Skelett des Puffers abgebildet, nach der Anwendung der lokalen Operationen. Man kann sehen, daß das Skelett nur zwei typische Abbiegeverläufe enthält und von Details abstrahiert hat. Information über Betrag und Richtung der Geschwindigkeiten ist ebenfalls im Puffer unterscheidbar, in der Abbildung aber nicht zu sehen.

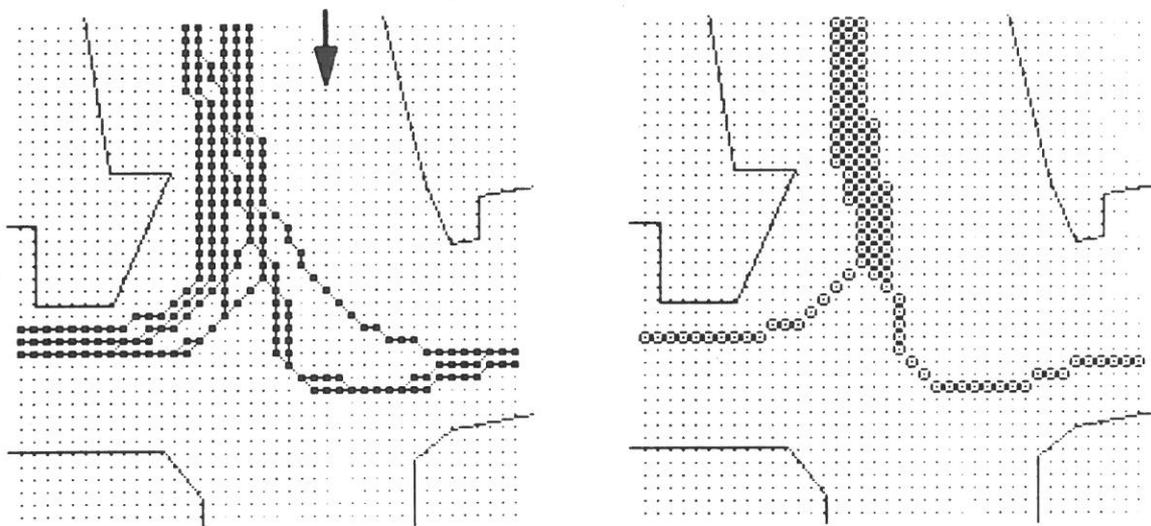


Abbildung 2: Zwei Dimensionen ( $x, y$ ) des Puffers nach der Aufnahme von 10 Abbiegetrajektorien (links); Skelett desselben Puffers nach Anwendung lokaler Operationen (rechts)

Ein Skelett enthält die wesentliche Information des dazugehörigen Buffers, eine Art 'Grobsicht' auf die angesammelten Beispiele. Es kann deshalb für eine effiziente und abstrakte Beschreibung genutzt werden. Als Beschreibung für Skelette wird ein erweiterter Kettencode verwendet (in mehreren Dimensionen). Kettencodes wurden ursprünglich für die Beschreibung von 2-dimensionalen räumlichen Konturen vorgeschlagen (siehe z.B. *Ballard + Brown 82*). Im mehrdimensionalen Buffer beschreibt der Kettencode ein Skelett als Folge von Elementen, welche jeweils eine bestimmte Richtung mit bestimmter Geschwindigkeit repräsentieren. Dabei sind Verzweigungen zugelassen. Die

Sequentialität der Zeit ist implizit in der Beschreibung; denn man kann auf die einzelnen raum-zeitlichen Elemente nur über den jeweiligen Vorgänger zugreifen.

Kettencodebeschreibungen können ebenfalls einzelne Trajektorien repräsentieren. Die Zwischenspeicherung von Einzeltrajektorien kann z.B. erforderlich sein, wenn einzelne Ereignisbeispiele zeitlich weit auseinanderliegen und sie zwischengespeichert werden müssen, bevor ein Ereignismodell berechnet werden kann.

Die Beschreibung eines Skeletts wird mit abstrakten propositionale Ereignisbeschreibungen assoziiert (siehe *Mohnhaupt + Neumann 90*). Z.B. enthält das Modell für 'abbiegen' sowohl Propositionen, welche für eine Ereigniserkennung benutzt werden, als auch eine Skelettbeschreibung. Um ein typisches Ereignis zu visualisieren, muß dann ein Buffer instantiiert und mit dem Skelett des Ereignisses gefüllt werden.

Ein Ziel gegenwärtiger Untersuchungen ist eine über Kettencodeelemente hinausgehende komplexere Beschreibung der raum-zeitlichen Trajektorienelemente. Außerdem ist es für eine detailliertere Beschreibung eventuell nötig, nicht nur die Maxima (das Skelett), sondern auch andere Informationen über die Häufigkeitsverteilung im raum-zeitlichen Buffer zu speichern.

## 3.2 Invariante Ereignisseigenschaften

Die im letzten Abschnitt beschriebene Akkumulierung von Beispielen liefert typische Information über Objektbewegungen, ist aber dahingehend spezifisch, daß sie auf Beobachtungen in einer bestimmten geometrischen Umgebung beruht. Ein nächster Schritt in Richtung generischer Ereignismodelle besteht darin, Ereignisseigenschaften abzuleiten, die unabhängig von einer bestimmten Umgebung sind, z.B. ein Modell für 'abbiegen', welches unabhängig von der Kreuzungsgeometrie gültig ist. Beispiele aus unterschiedlichen Umgebungen müssen integriert werden und erlauben dann auch Vorhersagen für neue Umgebungen, in denen bisher keine Beispiele angehäuft worden sind. Dies ist notwendig, da im allgemeinen Fall nicht angenommen werden kann, daß für jede denkbare Situation direkte Beobachtungen vorliegen.

### 3.2.1 Perzeptuelle Primitive

Das zentrale Problem besteht darin, invariante Ereignisseigenschaften zu berechnen. Sie sollten für die Beschreibung typischer Ereignisse notwendig und hinreichend sein und sie damit eindeutig charakterisieren. Die Idee ist, die invarianten Eigenschaften aus einer Menge von mächtigen, aber kompakten perzeptuellen Primitiven zu extrahieren. Die Primitive sollten die zu lösenden Aufgaben vereinfachen und sollten robust und effizient aus den Eingabedaten berechenbar sein. Außerdem muß die Menge der Primitive möglichst vollständig sein, und die Primitive sollten voneinander weitgehend unabhängig sein, bezüglich der Information, die explizit ist, d.h. ohne oder nur mit geringen Kosten erhältlich ist (siehe *Levesque 86*). Die Unabhängigkeit zweier Primitive bedeutet also hier: nicht voneinander ableitbar in der zur Verfügung stehenden Zeit. Dies hängt natürlich auch von den möglichen Operationen ab. Zum Beispiel können die Positionen zweier Objekte und deren relativer Abstand als voneinander unabhängig angesehen werden, obwohl der Abstand aus den Positionen ableitbar ist; denn zwei Positionen enthalten ihren Abstand nur implizit.

Diejenigen perzeptuellen Primitive sind geeignete Kandidaten für eine beschreibende Untermenge, die Konstantheiten über verschiedene Beispiele aufweisen, und damit invariante Ereigniseigenschaften beschreiben. Kandidaten sind zunächst die elementaren physikalischen Größen Ort und Geschwindigkeit einzelner Objekte, welche schon im Skelett enthalten sind (siehe Abschnitt 3.1). Es ist sinnvoll neben den elementaren Größen zusätzlich deren Ableitungen sowie elementare Größen relativ zu Referenzobjekten zu berechnen und damit explizit zu machen. Die folgende Menge von Primitiven ist geeignet, aus ihnen invariante Ereigniseigenschaften zu gewinnen:

1. Position
2. Orientierung
3. Orientierungsänderung
4. Geschwindigkeit
5. Beschleunigung
6. Position relativ zu einem Referenzobjekt (Distanz)
7. Distanzänderung
8. Orientierung relativ zu einem Referenzobjekt
9. Orientierungsänderung relativ zu einer Referenzorientierung
10. Geschwindigkeit relativ zur Geschwindigkeit eines Referenzobjektes

Die Menge der perzeptuellen Primitive enthält sowohl Größen, die sich ausschließlich auf Trajektorien beziehen (1-5), als auch Größen, die sich auf Beziehungen zwischen Trajektorien und Referenzobjekten beziehen (6-10).

Es besteht eine prinzipielle Ähnlichkeit zwischen den hier aufgezählten Primitiven und denjenigen qualitativen Primitiven, welche für eine Ereigniserkennung in *Neumann + Novak 86* vorgeschlagen wurden. Die dort vorgeschlagenen Primitive können ausgehend von einer quantitativen Szenenbeschreibung berechnet werden. Konstantheiten (wie z.B. konstante Geschwindigkeit), eingeschränkte Werte (wie 'parallel', 'dicht' oder 'neben'), Vergleichswerte und konstante Ableitungen (wie konstante Beschleunigung) werden dort verwendet. Ein Hauptziel der Autoren ist eine natürlichsprachliche Beschreibung einer Straßenverkehrsszene; dabei ist eine propositionale Ereignisbeschreibung abgeleitet aus den Primitiven ein sinnvoller Zwischenschritt. Die Auswahl der Primitive basiert in den genannten Arbeiten auf einer Analyse von Bewegungsverben, während hier die oben genannten Kriterien und die Randbedingungen einer analogen Repräsentation ausschlaggebend sind. In Abschnitt 3.3 wird deutlich, daß die Ähnlichkeit relevanter Primitive in beiden Arbeiten den angestrebten Übergang von einer analogen quantitativen zu einer propositionalen qualitativen Repräsentation vereinfacht und damit den vorgeschlagenen Ansatz stützt.

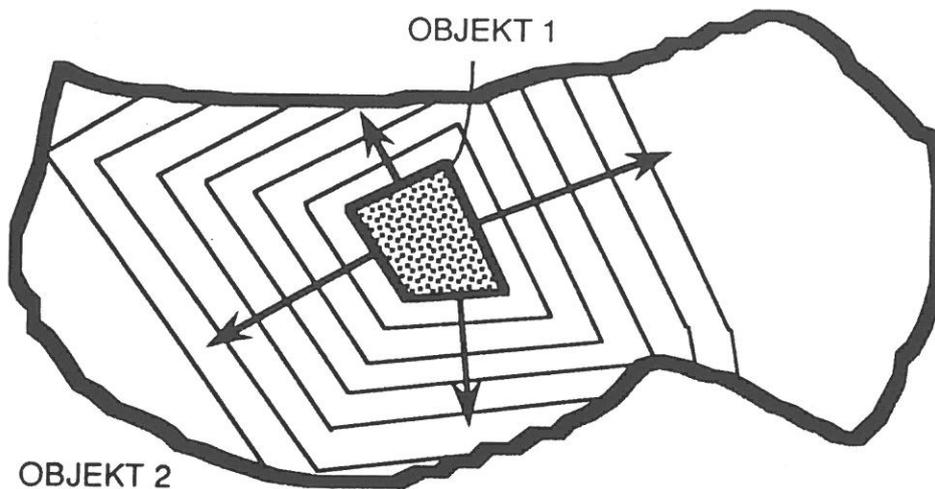


Abbildung 3: Berechnung der Distanzen zwischen zwei Objekten

### 3.2.2 Berechnung der perzeptueller Primitive

Eine wichtige Frage betrifft die Berechnung der perzeptueller Primitive. Im Allgemeinen gibt es in einer Szene viele interessante Objekte und die Anzahl möglicher Beziehungen zwischen Objekteigenschaften wächst exponentiell mit der Anzahl der Objekte. Deshalb wird davon ausgegangen, daß nicht alle Primitive simultan mit der durch visuelle Prozesse eintreffenden Information extrahiert werden. Weil die perzeptuellen Primitive implizit vorhanden sind, wenn stationärer Hintergrund und die Trajektorien bewegter Objekte (durch Skelette oder kodierte Einzelbeispiele) gespeichert sind, kann die Berechnung der Primitive durch nachfolgende Simulationen im raum-zeitlichen Puffer geleistet werden.

Dazu muß der Puffer zunächst instantiiert werden mit einer bestimmten stationären Szenenumgebung und mit einer oder mehreren relevanten Objekttrajektorien. Dann kann eine Simulation gestartet werden, bei der die perzeptuellen Primitive mit Prozessen sich ausbreitender Aktivierung berechnet werden können.

Als Beispiel ist in Abbildung 3 die Berechnung des perzeptuellen Primitivs Distanz für zwei Objekte skizziert. Objekt 1, das eine bestimmte Position in  $xy$  einnimmt, sendet Aktivierungen zu seinen Nachbarzellen. Die Nachbarn senden den Impuls jeweils an ihre Nachbarn mit einer konstanten Verringerung der Aktivierung. Die Zellen, welche den Koordinaten des Objektes 2 entsprechen, werden nach einigen Schritten von den sich ausbreitenden Aktivierungen erreicht. Der erste ankommende Aktivationswert repräsentiert die jeweils kürzeste Entfernung zu Objekt 1, wenn jede Zelle pro Zeiteinheit zu ihren jeweiligen Nachbarzellen propagiert.

### 3.2.3 Generische Modelle

Nachdem beschrieben wurde, daß invariante Ereignisseigenschaften ausgehend von einer Menge von perzeptuellen Primitiven bestimmt werden können, und ein Beispiel skizziert wurde, wie diese Primitive mithilfe von Simulationen im raum-zeitlichen Puffer berechnet werden können, werden hier einige Beispiele für generische Modelle gezeigt.

Zuerst wird auf ein generisches Modell für 'abbiegen' eingegangen. Im Abschnitt über Trajektorienanhäufung im raum-zeitlichen Puffer (3.1) wurden zunächst Information über  $x$ ,  $y$ ,  $v$  und  $d$  einzelner Trajektorien gesammelt. Perzeptuelle Primitive können als zusätzliche Dimensionen der Beschreibung angesehen werden. Sie sind implizit in der  $(x, y, v, d)$  Darstellung mit stationärem Hintergrund enthalten und müssen explizit gemacht werden durch lokale Prozesse innerhalb des Puffers (siehe 3.2.). Dieser zunächst multidimensionale Raum wird danach auf die invarianten Dimensionen reduziert und beschreibt dann ein generisches Modell für das entsprechende Ereignis.

Wenn Objektbewegungen von Abbiegeereignissen aus unterschiedlichen Szenen betrachtet werden, zeigt sich, daß  $x$ ,  $y$  und  $d$  variieren, die folgenden Primitive aber in erster Näherung invariant sind:

- Betrag der Geschwindigkeit ( $v$ ),
- relative Orientierung zwischen Fahrzeug und Straßenbegrenzung ( $ro$ ),
- Abstand zwischen Fahrzeug und Straßenbegrenzung ( $di$ ).

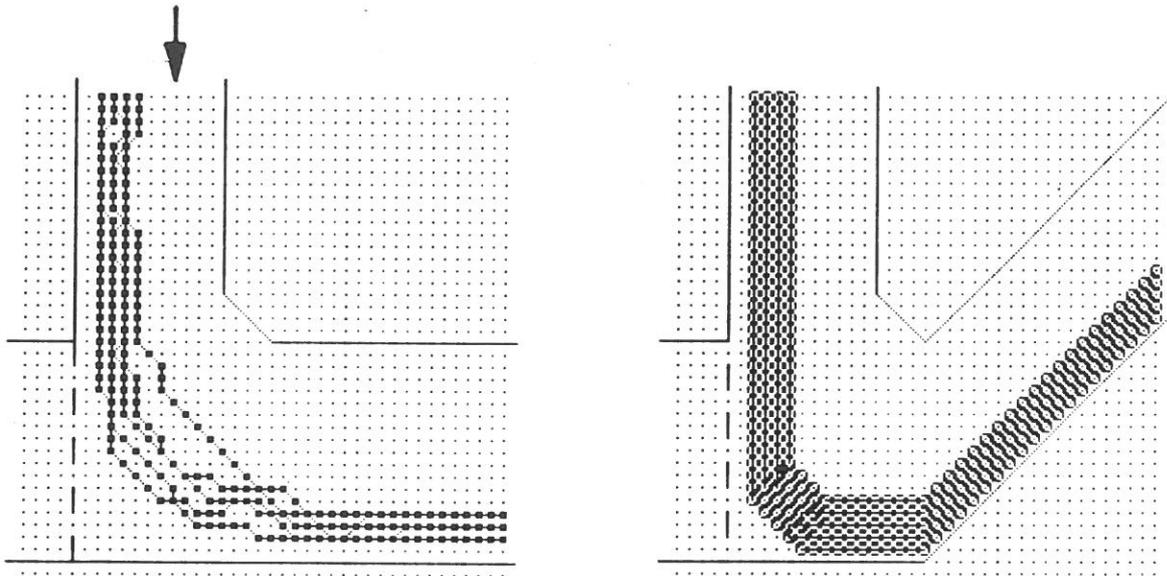


Abbildung 4: Puffer für 'abbiegen' nach 5 Beispielen (links). Auf neue Kreuzung übertragene Information mithilfe perzeptueller Primitive (rechts)

Das folgende Beispiel zeigt die Nützlichkeit der beschriebenen Invarianten. In Abbildung 4 sind links die Spuren von mehreren Abbiegevorgängen zu sehen. Nach Anwendung der lokalen Verwaschungsoperation und der Berechnung eines Skeletts wurden

die beschriebenen invarianten Dimensionen ( $v$ ,  $r_0$ ,  $d_i$ ) berechnet, und damit ein generisches Modell für 'abbiegen'. Dann wurde das Modell auf eine neue Kreuzung mit einer unterschiedlichen Geometrie übertragen (rechts). Dabei müssen wieder die zur neuen Umgebung passenden Dimensionen  $x$ ,  $y$ ,  $v$ ,  $d$  berechnet werden. Dann können dort Visualisierungen und z.B. Vorhersagen über weitere Verläufe von angefangenen Trajektorien berechnet werden.

Information aus einer bestimmten raum-zeitlichen Szenenumgebung wurde damit so transformiert, daß sie auf eine unterschiedliche Umgebung mit unterschiedlicher Geometrie anwendbar ist. Es wurden invariante Ereignisseigenschaften ausgenutzt. Gleichzeitig wurde von varianten Dimensionen abstrahiert. Die Abbildung zeigt berechnete Visualisierungen von Abbiegeereignissen für eine bestimmte Kreuzung in Analogie zu beobachteten Ereignissen auf einer anderen Kreuzung.

Im Folgenden wird ein zweites Beispiel betrachtet. Nach der Aufnahme von unterschiedlichen Überholvorgängen zwischen Fahrzeugen an unterschiedlichen Orten kann ein generisches Modell auf drei perzeptuellen Primitiven basieren:

- Abstand zwischen den Fahrzeugen,
- ihre relative Orientierung,
- ihre relative Geschwindigkeit.

Die beobachteten Trajektorien folgen alle einem ähnlichen Pfad durch den von den drei Größen aufgespannten 3-dimensionalen Raum. Dabei können Abstand und relative Orientierung über der Zeit wie in Abbildung 5 aussehen:

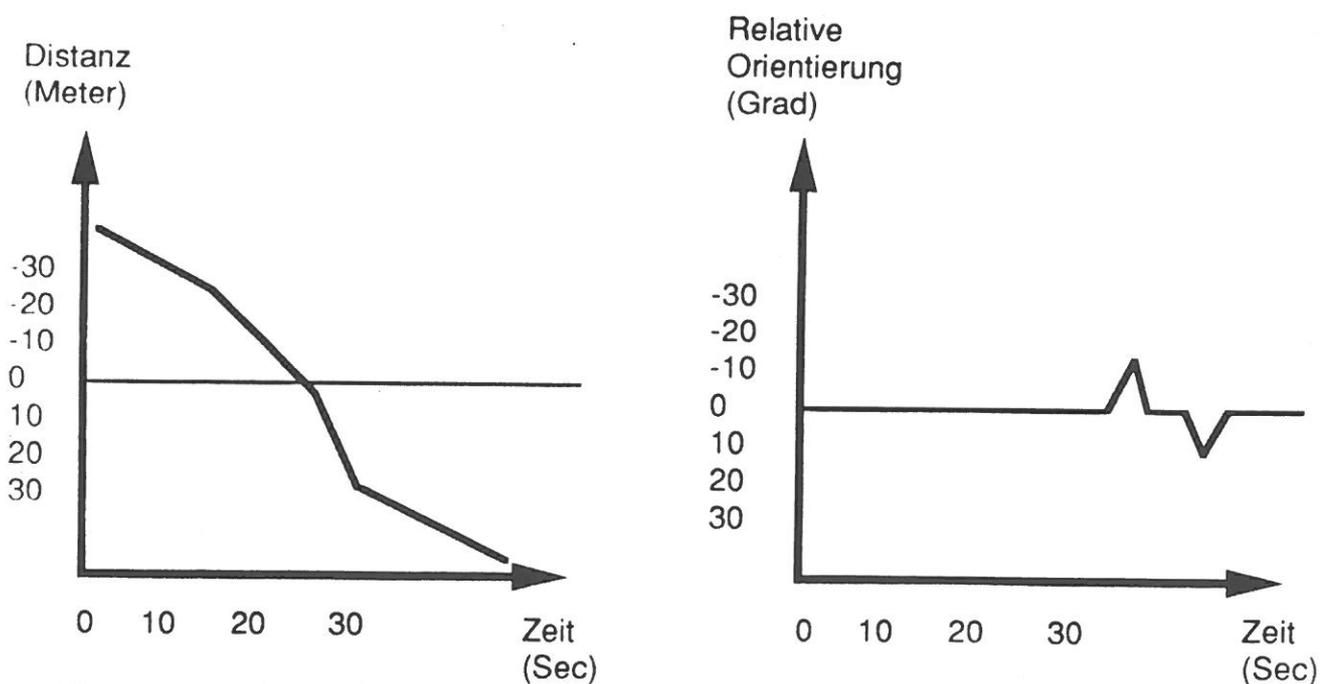


Abbildung 5: Generisches Modell für 'überholen'

Das generische Modell kann auf die gleiche Weise beschrieben werden wie das Skelett eines Puffers mit den Dimensionen  $(x, y, v, d)$ . Um den Pfad durch den im Allgemeinen mehrdimensionalen Raum zu beschreiben, kann ebenfalls der erweiterte Kettencode benutzt werden (zur vollständigen Spezifizierung des Modells gehört auch ein Anfangspunkt). Diese Beschreibung kann ebenfalls mit einem abstrakten propositionalen Ereignismodell (siehe *Neumann + Novak 86* und Abschnitt 2) assoziiert werden, um bei Bedarf einen Puffer für eine konkrete Situation zu instantiieren.

### 3.3 Qualitative Prädikate

Vom generischen Modell, welches noch analog bezüglich der invarianten Ereignisdimensionen ist, kann in einem nächsten Schritt eine qualitative und propositionale Ereignisbeschreibung abgeleitet werden, welche der in Abschnitt 2 eingeführten Repräsentation entspricht. Diese kann dann für eine grobe Ereigniserkennung, für natürlichsprachliche Kommunikation und für eine Langzeit-Speicherung genutzt werden.

Es wird das Beispiel 'überholen' betrachtet, für welches im letzten Abschnitt ein generisches Modell beschrieben worden ist. Die zentrale Idee besteht darin, den quasi-kontinuierlichen Verlauf durch die invarianten Dimensionen in bedeutungsvolle Intervalle einzuteilen, und sie dann mit Prädikaten zu beschreiben, welche an natürliche Sprache angelehnt sind.

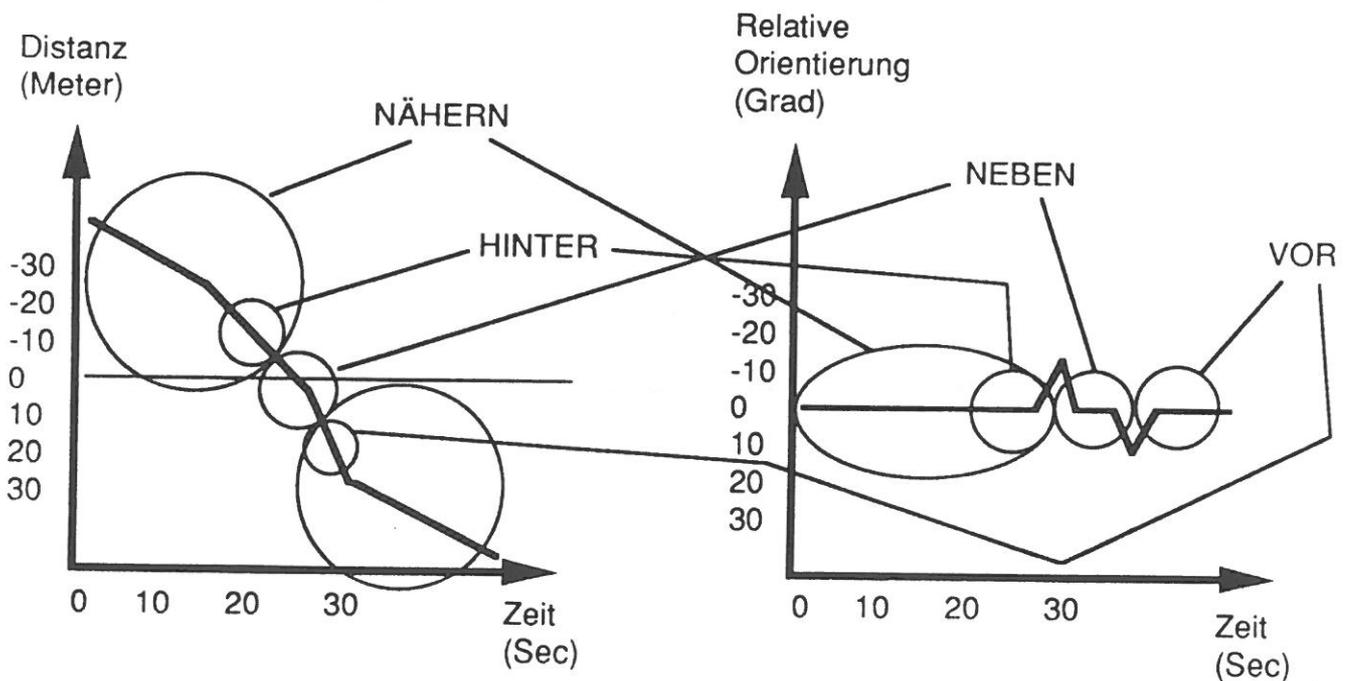


Abbildung 6: Prädikation des generischen Modells für 'überholen'

In Abbildung 6 ist dies am Beispiel 'überholen' skizziert. Die abgebildeten Segmente korrespondieren näherungsweise mit den Prädikaten 'sich nähern', 'hinter', 'neben', 'vor', und 'sich entfernen'. Es ist eine offene und schwierige Frage, einen geeigneten Algorithmus zu entwerfen für die Zuordnung von Segmenten des generischen Modells

und Prädikaten, die an Sprache angelehnt sind. Es hängt nicht nur vom Ereignistyp ab, sondern auch vom Vergleich mit anderen möglichen Ereignissen und deren Ähnlichkeiten. Das Beispiel zeigt, daß ein natürlicher Übergang von generischen Modellen zu propositionalen qualitativen Beschreibungen möglich ist.

## 4 Zusammenfassung

Das 'Verstehen' von Objektbewegungen ist eine wichtige Aufgabe für kognitive Systeme. In diesem Beitrag wurde ein hybrider Ansatz für die Repräsentation von Objektbewegungen diskutiert. Die verschiedenen Probleme im Zusammenhang mit Objektbewegungen, wie z.B. Pfadplanung, raum-zeitliches Schließen, Ereigniserkennung und Lernen von Objektbewegungen werden in der Literatur meist getrennt behandelt. Hier werden die verschiedenen Aufgaben auf einheitliche Repräsentationen zurückgeführt. Unter Berücksichtigung eigener Arbeiten und anderer Forschungen zur Modellierung von Objektbewegungen wurden dabei zwei verschiedene Repräsentationen verwendet, auf denen unterschiedliche Prozesse ablaufen.

Erstens wurde eine quantitative analoge Repräsentation in einem raum-zeitliche Puffer vorgestellt, in der einfache, lokale und parallele Prozesse ausgenutzt wurden um Inferenzen zu ziehen. Und zweitens wurden abstraktere qualitative und propositionale Ereignismodelle beschrieben, für welche logik-ähnliche Inferenzmechanismen verwendet werden. Der Puffer dient als temporäre Repräsentation, die bei Bedarf instantiiert wird, und dabei für Visualisierungen, zum Lernen von Objektbewegungen und für Aufgaben des raum-zeitlichen Schließens genutzt wird. Er kann sowohl durch perzeptuelle, als auch durch höhere kognitive Prozesse beeinflusst werden.

Die Ereignismodelle gestatten eine effiziente Ereigniserkennung, eine kompakte Langzeitspeicherung, und sie sind Startpunkt für natürlichsprachliche Kommunikation. Zusätzlich erlauben die Ereignismodelle eine Initialisierung des raum-zeitlichen Puffers, da mit ihnen Information über typische Ereignisse und deren relevante Beschreibungsdimensionen assoziiert werden kann.

Es gibt drei verschiedene Begründungen für diesen Ansatz. Erstens liegen theoretische Ergebnisse über Komplexitätsgrenzen rein propositionaler und logik-basierter Ansätze vor (siehe z.B. *Levesque 86*). Zweitens gibt es psychologische Evidenz für bildhafte (dem Puffer ähnliche) Repräsentationen, neben den allgemein nicht bezweifelten propositionalen Repräsentationen (siehe 'Imagery'-Debatte, z.B. *Kosslyn 80*, *Block 81*). Und drittens stellt der Ansatz einen Vorschlag dar, die erkannte Komplexität einiger Aufgaben dadurch behandelbar zu machen, daß mit dem raum-zeitlichen Puffer eine adäquat eingeschränkte (d.h. der Natur der Aufgaben entsprechende) Repräsentation verwendet wird. Der Puffer vereinfacht Aufgaben, bei denen räumliche und zeitliche Beziehungen zwischen Objekten wichtig sind, da diese Beziehungen durch einfache lokale und parallele Prozesse berechenbar sind. Außerdem sind durch die Repräsentation einige physikalische Randbedingungen internalisiert (siehe *Palmer 78*), sie müssen daher nicht extra berechnet werden. Der Puffer erlaubt darüberhinaus einen natürlichen Übergang von einzelnen Beispielen zu akkumulierter Erfahrung, ebenfalls unter Verwendung einfacher lokaler Prozesse.

Es wird hier angenommen, daß eine geeignete Repräsentation das Erlernen der Mod-

elle auf plausible Weise ermöglichen muß. D.h. der diskutierte Ansatz ist nur dann tragbar, wenn die Modelle im raum-zeitlichen Puffer aus den Ausgaben der niederen Bilddeutung gewonnen werden können, und wenn die propositionalen Ereignismodelle abgeleitet werden können aus Information, die im Puffer angehäuft wurde. Die Untersuchungen haben gezeigt, daß es möglich, ist aus Beispielen von Objektbewegungen typische (aber noch szenenabhängige) Ereignisinformation zu gewinnen und anschließend durch die Ausnutzung von invarianten perzeptuellen Primitiven generische (szenenunabhängige) Ereignisbeschreibungen abzuleiten. In einem nächsten Verarbeitungsschritt können propositionale Ereignismodelle berechnet werden, durch geeignete qualitative Beschreibungen der quantitativen generischen Modelle. Damit entsteht ein natürlicher Übergang von analogen zu propositionalen Ereignismodellen, welche zunächst voneinander unabhängig entworfen wurden. Dies ist eine weitere Stützung des hybriden Ansatzes.

Einige der beschriebenen Schritte wurden durch experimentelle Ergebnisse gestützt. Es ist aber deutlich, daß die vorgeschlagenen Repräsentationen und deren Prozesse an vielen Stellen weiter und detaillierter untersucht werden müssen. Einige offene Fragen wurden bereits in Abschnitt 3 angesprochen, z.B. die Auswahl geeigneter raum-zeitlicher Formprimitive, oder die algorithmische Zuordnung von Segmenten des generischer Modelle zu Primitiven für natürlichsprachliche Beschreibungen. Außerdem fehlt weiterhin eine fundierte Theorie mit der eine (möglichst) optimale Repräsentation ausgewählt werden kann, gegeben eine bestimmte Klasse von Problemen.

#### **Danksagung:**

Ich danke Bernd Neumann für seine Unterstützung und zahlreiche interessante Diskussionen und Anregungen. Außerdem danke ich Simone Pribbenow und David Fleet für konstruktive Kritik.

## **Literatur**

- [Adorni and Di Manzo 83] *Top-down Approaches to Scene Interpretation*. G. Adorni, M. Di Manzo. Proc. CIL 83, Barcelona, Spain, June 1983.
- [Andre + Bosch + Herzog + Rist 86] *Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions*. E. Andre, G. Bosch, G. Herzog, T. Rist. in Proc. 7th ECAI, 1986.
- [Ballard + Brown 82] *Computer Vision*. Dana H. Ballard, Christopher M. Brown. Prentice Hall, 1982.
- [Block 81] *Imagery*. Ned Block(Ed.). MIT Press, Cambridge, Mass. 1981.
- [Bradshaw 87] *Learning about speech sounds: the NEXUS Project*. Gary Bradshaw. 4th Int. Workshop on Machine Learning, Irvine 1987, pp. 1-11.
- [Denis + Engelkamp + Richardson 88] *Cognitive and Neuropsychological Approaches to Mental Imagery*. M. Denis, J. Engelkamp, J. T. E. Richardson (Ed.). Martinus Nijhoff Publisher 1988.

- [Finke 85] *Theories Relating Mental Imagery to Perception*. Ronald A. Finke. *Psychological Bulletin* **98** (1985) 236-259.
- [Funt 80] *Problem solving with diagrammatic representations*. B. V. Funt. *Artificial Intelligence* **13** (1980) 201-230.
- [Gardin + Meltzer 89] *Analogical Representation of Naive Physics*. Francesco Gardin, Bernhard Meltzer. *Artificial Intelligence* **38** (1989) 139-159.
- [Gelernter 63] *Realization of a geometry theorem-proving machine*. H. Gelernter. in 'Computer and thought', E. Feigenbaum and J. Feldman (Ed.), McGraw Hill 1963.
- [Khenkhar 90] *Eine objektorientierte Darstellung von Depiktionen auf der Grundlage von Zellmatrizen*. M. N. Khenkhar. in 'Repräsentation und Verarbeitung räumlichen Wissens', C. Freksa und C. Habel (Hrsg.), Springer-Verlag 1990.
- [Kibler + Aha 87] *Learning Representative Exemplars of Concepts: An Initial Case Study*. Dennis Kibler, David W. Aha. 4th Int. Workshop on Machine Learning, Irvine 1987, pp. 24-30.
- [Kosslyn 80] *Image and Mind*. Stephen M. Kosslyn. Harvard University Press, 1980.
- [Levesque 86] *Making Believers out of Computers*. Hector J. Levesque. *Artificial Intelligence* **30** (1986) 81-108.
- [Lindsay 88] *Images and inference*. Robert K. Lindsay. *Cognition* **29** (1988) 229-250.
- [Marschark 88] *The Functional Role of Imagery in Cognition*. Marc Marschark. in *Denis + Engelkamp + Richardson 88*, pp. 405-417.
- [Mohnhaupt + Fleet 88] *Raum-zeitliche Filter für eine top-down Steuerung der Bewegungsanalyse*. Michael Mohnhaupt, David Fleet. Proc. German Workshop on Artificial Intelligence GWAI-12, 1988, 296-305.
- [Mohnhaupt + Neumann 89] *Some aspects of learning and reorganisation in an analogical representation*. Michael Mohnhaupt, Bernd Neumann. in 'Knowledge representation and organisation in machine learning', K. Morik (Ed.), lecture notes in artificial intelligence, Springer Verlag 1989, pp. 50-64.
- [Mohnhaupt + Neumann 90] *Understanding Object Motion: Recognition, Learning and Spatio-Temporal Reasoning*. Michael Mohnhaupt, Bernd Neumann. to appear in 'Journal of Robotics and Autonomous Systems', North Holland 1990.
- [Neumann + Mohnhaupt 88] *Propositionale und analoge Repräsentation von Bewegungsverläufen*. Bernd Neumann, Michael Mohnhaupt. *Künstliche Intelligenz* **1** (1988) 4-10.
- [Neumann + Novak 83] *Event models for recognition and natural-language description of events in real-world image sequences*. Bernd Neumann, Hans-Joachim Novak. Proc. Int. Joint Conf. on Art. Intell. IJCAI-8, 1983, 724-726.

- [Neumann + Novak 86] *NAOS: Ein System zur natürlichsprachlichen Beschreibung zeitveränderlicher Szenen*. Bernd Neumann, Hans-Joachim Novak. *Informatik Forsch. Entw.* 1 (1986) 83-92.
- [Novak 87] *Textgenerierung auf der Grundlage visueller Daten: Beschreibungen von Straßenszenen*. Hans-Joachim Novak. Springer Verlag 1987.
- [Palmer 78] *Fundamental Aspects of Cognitive Representation*. S. P. Palmer. in: E. Rosch, B.B. Lloyd (Ed.): *Cognition and Categorisation*, Hillsdale, N.Y.: Erlbaum Press 1978.
- [Pinker 88] *A computational theory of the mental imagery medium*. Steven Pinker. in Denis + Engelkamp + Richardson 88, pp. 17-32.
- [Pribbenow 90] *Interaktion von propositionalen und bildhaften Repräsentationen*. Simone Pribbenow. in 'Repräsentation und Verarbeitung räumlichen Wissens', C. Freksa und C. Habel (Hrsg.), Springer-Verlag 1990.
- [Smith + Medin 81] *Categories and Concepts*. Edward E. Smith, Douglas L. Medin. Harvard University Press 1981.
- [Stanfill + Waltz 86] *Toward Memory-Based Reasoning*. Graig Stanfill, David Waltz. *Communication of the ACM* 29 (1986) 1213-1228.
- [Steels 88] *Steps Towards Common Sense*. Luc Steels. Proc. 8th ECAI, 1988, pp. 49-54.
- [Waltz + Boggess 79] *Visual Analog Representations for Natural Language Understanding*. David L. Waltz, Lois Boggess. Proc. Int. Joint Conf. on Art. Intell. IJCAI-6, 1979, 926-934.