

Das System SOCCER: Simultane Interpretation und natürlichsprachliche Beschreibung zeitveränderlicher Szenen

Gerd Herzog, Gudula Retz-Schmidt

SFB 314, Project VITRA, Universität des Saarlandes
D-66041 Saarbrücken

Zusammenfassung

Im Hinblick auf die Integration von maschinellem Sehen und der Verarbeitung natürlicher Sprache setzt sich der vorliegende Beitrag genauer mit dem Problem der simultanen Analyse und natürlichsprachlichen Beschreibung einer Realweltbildfolge auseinander.

Als Diskursbereich für die Entwicklung des Systems *Soccer*, einem natürlichsprachlichen Zugangssystem zu einem Bildfolgenanalyse-System, wurden kurze Szenen aus Fußballspielen gewählt. In dieser Domäne können die Beschreibungen der Textsorte Fußballdirektreportage nachempfunden werden, für die bereits zahlreiche linguistische Untersuchungen vorliegen. Die Eingabe für das System *Soccer* bilden die aus Realweltbildfolgen extrahierten Trajektorien der bewegten Objekte. Diese rein geometrische Repräsentation der betrachteten Szene wird automatisch weiter analysiert, d.h. es werden interessante Ereignisse erkannt, mit dem Ziel, die Geschehnisse in der Szene sprachlich zu beschreiben. Im Gegensatz zu bisherigen Ansätzen soll dabei jedoch nicht von einer bereits vollständig analysierten Bildfolge ausgegangen werden; vielmehr sind Ereignisse so zu repräsentieren, daß sie simultan zu ihrem Auftreten in der Szene detektiert und natürlichsprachlich beschrieben werden können.

Neben der Erkennung und Beschreibung rein visueller Phänomene sollen die Beobachtungen auch interpretiert werden. Im Teilsystem *Replai* wird dazu mithilfe von Planerkennungsmethoden vom beobachteten Verhalten auf die zugrundeliegenden Intentionen zurückgeschlossen. Hierfür ist es von Nutzen, daß in der Domäne "Fußballspiel" relativ klar definierte Ziele und Abhängigkeiten zwischen Zielen bestehen, die für die Intentionserkennung ausgenutzt werden können.

Dieser Beitrag ist erschienen in: In: J. Perl (Hrsg.), Sport & Informatik, pp. 95–119. Schorndorf: Hofmann, 1989.

1 Einführung

Kernziel des Projektes *Vitra*¹ ist es, die komplexen Informationsverarbeitungsprozesse des Menschen, die der Interaktion von Sprachproduktion und visueller Wahrnehmung zugrundeliegen, mit informatischen Mitteln exakt zu beschreiben und zu erklären. Derzeit ist man von einem universell einsetzbaren KI-System, das beliebige Bildfolgen sprachlich beschreibt, noch sehr weit entfernt und muß sich bei der Systementwicklung jeweils auf eingeschränkte Diskursbereiche konzentrieren. Im Projekt *Vitra* werden unterschiedliche Diskursbereiche und Kommunikationssituationen betrachtet, um die Übertragbarkeit der entwickelten Konzepte und Methoden auf andere Domänen möglichst frühzeitig prüfen zu können (vgl. Wahlster (1987)). Als Diskursbereich für das System *Soccer* (vgl. André et al. (1988)), in dem der Forschungsschwerpunkt auf der simultanen Berichterstattung über beobachtbare Ereignisse während des Ablaufs einer Bildsequenz liegt, wurden kurze Szenen aus Fußballübertragungen gewählt. In dieser Domäne können die Beschreibungen der Textsorte Fußballdirektreportage nachempfunden werden, für die bereits zahlreiche linguistische Untersuchungen vorliegen (vgl. Rosenbaum (1969), Schneider (1974), Brandt (1983)).

Neben der Erkennung und Beschreibung rein visueller Phänomene (Ereignisse) sollen — ähnlich wie bei menschlichen Beobachtern — die Beobachtungen auch interpretiert werden. D.h. es soll vom beobachteten Verhalten auf die zugrundeliegenden Intentionen zurückgeschlossen werden. Zu diesem Zweck werden Planerkennungsmethoden verwendet. Hierfür ist es von Nutzen, daß in der Domäne "Fußballspiel" relativ klar definierte Ziele und Abhängigkeiten zwischen Zielen bestehen die für die Intentionserkennung ausgenutzt werden können.

Die einzelnen Schritte von der digitalisierten Bildfolge hin zur simultanen natürlichsprachlichen Beschreibung, die den Aufbau einer geometrischen Zwischenrepräsentation, die hieran anschließende inkrementelle Erkennung von Ereignissen und Intentionen sowie deren natürlichsprachliche Beschreibung umfassen, sollen in diesem Beitrag am Beispiel des Systems *Soccer* näher diskutiert werden.

2 Bildfolgenanalyse

Hauptaufgabe der Bildanalyse ist es, anhand von Bildern eine symbolische computerinterne Beschreibung einer Szene zu erzeugen. Bei der Bildfolgenanalyse steht hierbei insbesondere die Analyse und Interpretation bewegungsbedingter Änderungen im Vordergrund.² Am Beispiel des für die Kopplung mit *Vitra* eingesetzten Systems *Actions* wird im nachfolgenden Abschnitt gezeigt, auf welche Weise Information über bewegte Objekte aus einer Folge digitisierter Bilder abgeleitet werden kann. Hieran anschlie-

¹Visual TRAnslator

²Eine Einführung in das Bildverstehen bieten beispielsweise Ballard und Brown (1982) und Neumann (1982). Die Arbeit Nagel (1985) setzt sich besonders mit dem Bereich der Bildfolgenanalyse auseinander.

ßend wird die Schnittstelle zwischen Bildfolgenanalyse und *Vitra*, d.h. weitergehender Szeneninterpretation, motiviert und vorgestellt.

2.1 Das Bildfolgenanalyse-System ACTIONS

Das am Institut für Informations- und Datenverarbeitung (IITB) der Fraunhofergesellschaft, Karlsruhe, entwickelte System *Actions*³ leistet die automatische Detektion und Verfolgung von Objekten in Bildfolgen. Kernziel der Arbeit an *Actions* ist es, robuste, allgemein anwendbare Methoden zur Analyse von Realweltszenen zu entwickeln. Eine zusammenfassende Darstellung der bislang erzielten Ergebnisse findet sich in Sung und Zimmermann (1986) und Sung (1988). Abb. 1 bietet einen Überblick zu den einzelnen Verarbeitungsschritten. Bei dem betrachteten Bildmaterial handelt es sich um Aufnahmen von einer Straßenkreuzung, die von einem ca. 35m hohen Gebäude aus gemacht wurden, sowie um eigene Aufnahmen von einer Begegnung aus der Fußballbundesliga. Das Material wurde jeweils mit einer stationären, monokularen Kamera aufgenommen. Für die Analyse werden zur Zeit bis zu 132 Sekunden dauernde Ausschnitte von 3300 Vollbildern mit $512 * 512$ Bildpunkten zu je 8 Bit Grauwertaufösung verwandt.

Bewegte Objekte werden durch die Berechnung und Analyse von Verschiebungsvektorfeldern erkannt. Die Berechnung der Verschiebungsvektoren basiert dabei auf der Zuordnung von charakteristischen lokalen Grauwertverteilungen, sogenannten *Merkmalen*. Zur Bestimmung der Merkmale wird jedes Pixel mit einer festen Anzahl (im konkreten Fall 8) umliegender Pixel verglichen. Je nach Anzahl der Vergleichspunkte, deren Grauwert kleiner bzw. größer ist, wird das betrachtete Pixel einer entsprechenden Klasse zugeteilt. Zusammenhängende Punkte aus derselben Klasse können zu Flecken zusammengefaßt werden. Im weiteren werden dann nur noch Flecken aus einer der beiden extremen Klassen, d.h. Kuppen und Senken des Grauwertgebirges, berücksichtigt.

Um bei der Ermittlung von Verschiebungsvektoren Fehler durch zufällige Schwankungen der Merkmalspositionen zu verringern, werden die Schwerpunkte der Flecken über mehrere Bilder hinweg verfolgt. Hieraus ergeben sich lokale Verschiebungsvektoren vom n -ten zum $(n + 4)$ -ten Bild. Die so gewonnenen Verschiebungsvektorfelder werden hinsichtlich Betrag, Richtung und Bildposition auf Ballungen ähnlicher Vektoren untersucht. Die Ballungen werden jeweils durch ein entsprechend dem mittleren Bewegungsvektor ausgerichtetes umschreibendes Rechteck markiert. Diese *Rahmen* können als Kandidaten für das Abbild bewegter starrer Objekte gelten. Der geometrische Mittelpunkt eines Rahmens dient als Repräsentant des bewegten Objektes in der Bildebene.

Die Korrespondenzlinien der Rahmenpositionen entsprechen den Trajektorien der Vektorballungen und somit der Objektkandidaten in der Bildebene. Zwei Rahmen aus aufeinanderfolgenden Aufnahmen werden einander zugeordnet, falls ihre Bewegungs-

³Automatic Cueing and Trajectory estimation in Imagery of Objects in Natural Scenes

ACTIONS : "Fußball"

Automatic Cueing and Trajectory estimation in Imagery of Objects in natural Scenes

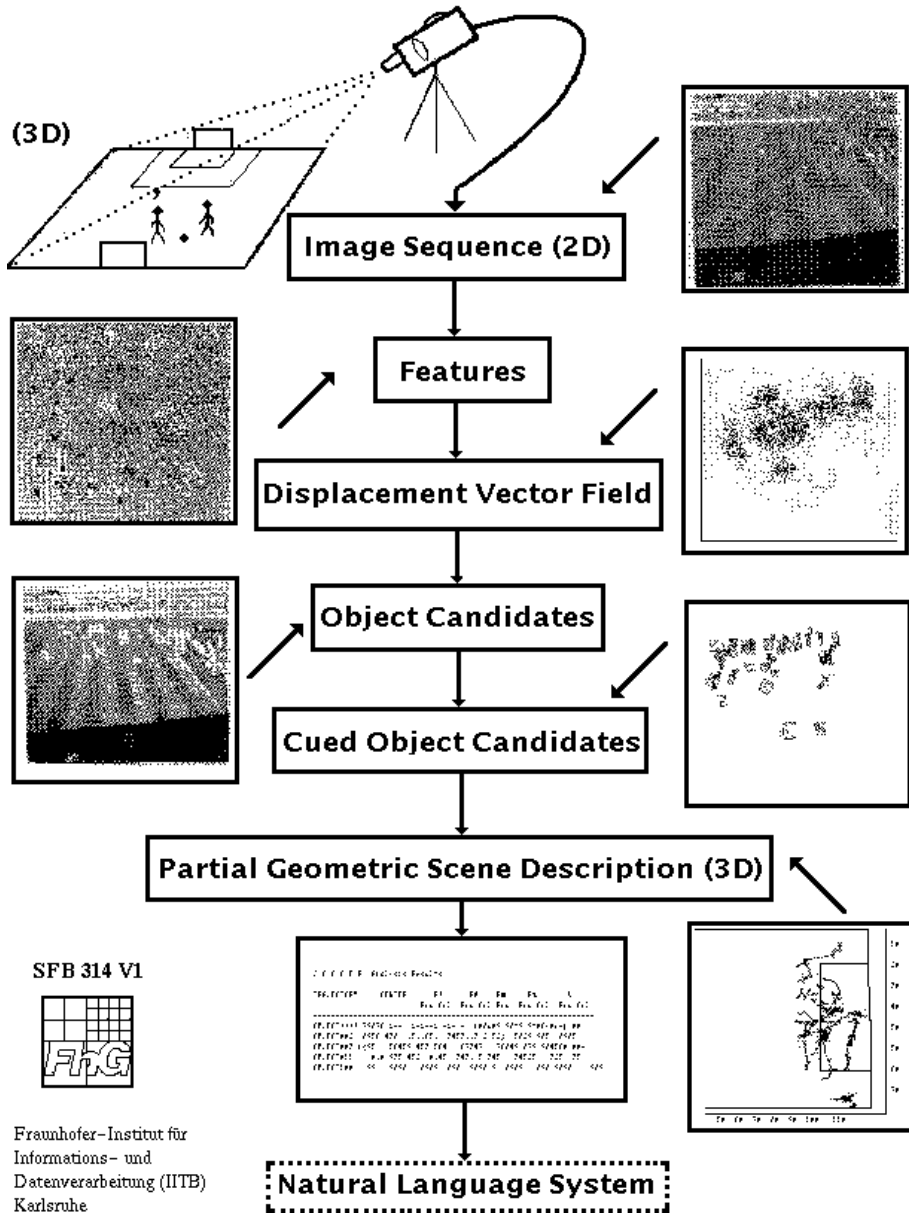


Abbildung 1: Verarbeitungsschritte in *Actions*

richtung annähernd übereinstimmt und ihr räumlicher Abstand kleiner ist als der doppelte Betrag des mittleren Verschiebungsvektors. Bei noch verbleibender Mehrdeutigkeit wird der Zuordnungskandidat mit dem kleinsten räumlichen Abstand gewählt. Die Koordinatenangaben der so gewonnenen Korrespondenzlinien werden unter Berücksichtigung von Kameraposition, Aufnahmewinkel und der Geometrie des statischen Szenenhintergrundes in Szenenkoordinaten rücktransformiert und in der (eingeschränkten) geometrischen Szenenbeschreibung zusammengefaßt. Die Klassifikation von Objektkandidaten und deren Zuordnung zu a priori bekannten Objekten kann derzeit von *Actions* noch nicht geleistet werden und erfolgt daher interaktiv.

2.2 Schnittstelle zwischen Bildfolgenanalyse und Interpretation

Die *geometrische Szenenbeschreibung* (GSB) wurde in Neumann und Novak (1986) als Repräsentation für die Ausgabe des Bildanalyseprozesses eingeführt. Ziel dieser Repräsentation ist es, die ursprüngliche Bildfolge prinzipiell vollständig und ohne Informationsverlust darzustellen. Das bedeutet, daß die mit der Kamera aufgenommene Bildfolge (im Prinzip) aus der GSB rekonstruiert werden könnte. Die geometrische Szenenbeschreibung enthält:

- Für jedes Einzelbild der Bildfolge: Zeitpunkt, alle in der Szene sichtbaren Objekte, Betrachterstandpunkt sowie Beleuchtungsdaten
- für jedes Objekt:
 - Identität (über die Bildfolge)
 - 3D-Position und Orientierung in Weltkoordinaten für jedes Einzelbild
 - 3D-Form und physikalische Oberflächeneigenschaften (Farbe)
 - Klassenzugehörigkeit und gegebenenfalls Identität in bezug auf mögliches Vorwissen (und somit weitere verbalisierbare Eigenschaften wie z.B. Namen)

Hierbei ist zu beachten, daß die Klassifizierung und die Identifizierung bereits vorher bekannter Objekte nur mit Hilfe zusätzlicher Wissensquellen möglich ist und nicht allein aus der reinen Bildinformation gewonnen werden kann.

Das Konzept der geometrischen Szenenbeschreibung stellt eine idealisierte Schnittstelle zwischen der Bildanalyse und einem darauf aufsetzenden natürlichsprachlichen System dar. In Anwendungen, wie z.B. dem von Neumann und Novak entwickelten System *Naos* zur natürlichsprachlichen Beschreibung von Straßenverkehrsszenen, wird die GSB jedoch entsprechend den konkreten Anforderungen eingeschränkt. In den *Vitra*-Systemen werden z.B. Betrachterstandpunkt, Beleuchtungsdaten sowie die vollständige 3D-Form der Objekte nicht berücksichtigt. Die Information über den statischen Szenenhintergrund wird nicht von der Bildanalyse geliefert, sondern liegt

vielmehr als instantiiertes Modell des betrachteten Weltausschnitts vor. Derartige Einschränkungen finden sich derzeit in allen Ansätzen, da man von einem universell einsetzbaren KI-System, das beliebige Bildfolgen vollständig analysieren kann, noch sehr weit entfernt ist.

3 Inkrementelle Ereigniserkennung

Die Aufgabe eines weitergehenden Verstehensprozesses besteht zunächst darin, raum- und zeitabhängige Konzepte aus der von einem Bildfolgenanalyse-System erzeugten geometrischen Beschreibung der sichtbaren Objekte und ihrer Trajektorien zu abstrahieren. Solche konzeptuellen Einheiten, die hier als Ereignisse bezeichnet werden, dienen zur Erfassung des Geschehens in einer zeitveränderlichen Szene. Ereignisse spielen dabei die Rolle von einzelsprachunabhängigen Bedeutungseinheiten; sie sind hier also diejenigen wahrnehmbaren Veränderungen der Welt, über die man üblicherweise spricht (vgl. Miller und Johnson-Laird (1976)).

3.1 Simultane versus retrospektive Interpretation

Neben der Frage, welche Konzepte aus einer geometrischen Szenenbeschreibung zu extrahieren sind, ist es entscheidend, wie der Erkennungsprozeß realisiert wird. Eine Betrachtung der relevanten Forschungsarbeiten zeigt, daß bisher nur Ansätze verfolgt wurden, die eine retrospektive Beschreibung der zu analysierenden zeitveränderlichen Szene zum Ziel haben. Die Systeme *Ham-Ans* (vgl. Wahlster et al. (1983)), *Naos* (vgl. Neumann und Novak (1986)), *Epex* (vgl. Walter (1987)) und das in *Vitra* entwickelte System *Citytour* (vgl. André et al. (1986), Schirra et al. (1987)) bedienen sich zur Interpretation von Straßenverkehrsszenen einer *a-posteriori-Strategie*, bei der vor Analysebeginn eine vollständige geometrische Szenenbeschreibung vorhanden sein muß. Im Gegensatz dazu erfolgt die Szenenanalyse in dem zur Herzkammerdiagnose eingesetzten System *Alven* (vgl. Tsotsos (1981)) datengetrieben und sukzessive. Dennoch sollen auch in diesem Fall die erkannten Ereignisse erst nach der vollständigen Analyse einer Szene zur Verfügung gestellt werden.

Eine völlig neue Problemstellung ergibt sich, wenn es darum geht, Ereignisse simultan zu ihrem Auftreten in der Szene zu erkennen. Dabei stellt sich insbesondere die Frage, wie teilweise erkannte Ereignisse zu repräsentieren sind, um sie zur weiteren Verarbeitung zur Verfügung zu stellen. Im Hinblick auf die Erzeugung simultaner Szenenbeschreibungen in natürlicher Sprache wird dieses Problem unmittelbar deutlich: Zur Zentrierung der Beschreibung auf das aktuelle Szenengeschehen ist es häufig sinnvoll, Ereignisse bereits dann zu verbalisieren, während sie ablaufen und nicht erst im Nachhinein, wenn sie vorbei sind. Beispiele hierfür wären die Beschreibung eines gerade auftretenden Überholvorgangs in einer Straßenverkehrsszene oder die Schilderung eines Angriffs in einem Fußballspiel bei einer Live-Reportage eines Radioreporters.

Diese Problemstellung stellt besondere Anforderungen an die Modellierung von Ereignissen, denen auch zeitlogischen Formalisierungen, wie die in Allen (1984) und McDermott (1982) vorgestellten Ansätze, nicht gerecht werden, da sie formal nur zwischen aufgetretenen und nicht aufgetretenen Ereignissen unterscheiden. Um zu einer feineren Beschreibung des Auftretens eines Ereignisses zu gelangen, bietet es sich an, die verschiedenen Phasen eines Ereignisses, d.h. den Beginn, den Ablauf und das Ende des Auftretens, zu berücksichtigen. Um den Zustand eines auftretenden Ereignisses zu einem vorgegebenen Zeitpunkt charakterisieren zu können, führen wir zusätzlich folgende Prädikate ein:

TRIGGER ($t_i \text{ event}$) für den Beginn,

PROCEED ($t_i \text{ event}$) für das Fortschreiten,

STOP ($t_i \text{ event}$) für das Ende und

SUCCEED ($t_i \text{ event}$) für das Andauern eines auftretenden Ereignisses.

Das Prädikat dient zur Modellierung von Ereignissen, die als solche bereits vollständig erkannt sind, deren Auftreten jedoch andauert. Hierzu zählen beispielsweise die zu durativen Bewegungsverben korrespondierenden Ereigniskonzepte wie fahren, laufen oder gehen.

Zur Veranschaulichung betrachte man einen Überholvorgang in einer Straßenverkehrsszene. Abb. 2 zeigt hierzu vier markante, aus einer Realbildfolge stammende Einzelbilder. Jedes Einzelbild korrespondiert zu je einem der diskreten Zeitpunkte T_1 bis T_4 . Zum Zeitpunkt T_1 nähert sich der Pkw dem Kleinbus. Zum Zeitpunkt T_2 schert der Pkw aus; er beginnt, den Kleinbus zu überholen. Die Tatsache, daß das Ereignis beginnt, wird formal durch **TRIGGER**(T_2 (Überholen PKW1 BUS1)) repräsentiert. Zum Zeitpunkt T_4 schert der Pkw vor dem Kleinbus ein und beendet somit den Überholvorgang. Das Ereignis (Überholen PKW1 BUS1) ist jetzt vollständig aufgetreten und es gilt: **STOP**(T_4 (Überholen PKW1 BUS1)). Zwischen T_2 und T_4 ist der Überholvorgang in der Szene zu beobachten. Das Ereignis tritt also gerade auf, ist jedoch während dieses Zeitraums noch nicht vollständig erkannt. Würde der Pkw beispielsweise zum Zeitpunkt T_3 in eine Seitenstraße abbiegen, so könnte nicht mehr von einem Überholvorgang gesprochen werden. Für noch nicht vollständig erkannte Ereignisse steht das Prädikat **PROCEED** bereit. In der im Beispiel dargestellten Situation gilt somit für alle t_i mit $T_2 < t_i < T_4$ die Prädikation **PROCEED**(t_i (Überholen PKW1 BUS1)).

Ein Beispiel für das Andauern eines Ereignisses bezüglich obiger Bildfolge wäre etwa die Tatsache, daß PKW1 als fahrend erkannt wird. Es gilt für alle t_i mit $T_1 \leq t_i \leq T_4$ die Prädikation **SUCCEED**(t_i (Fahren PKW1)).

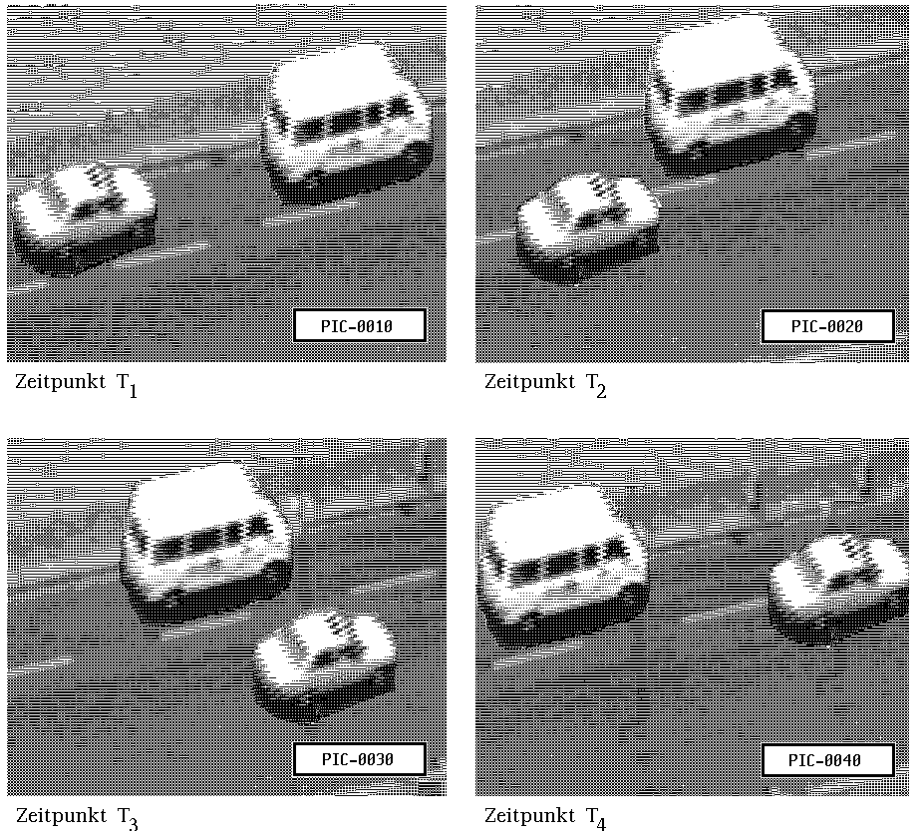


Abbildung 2: Ein Überholvorgang in einer Straßenverkehrsszene

3.2 Ereignismodelle zur inkrementellen Ereigniserkennung

In Analogie zu Objektmodellen werden Ereignisse konzeptuell durch *Ereignismodelle* beschrieben (vgl. Neumann und Novak (1983)). Ereignismodelle repräsentieren a-priori-Wissen zur Erfassung des Szenengeschehens, insbesondere Wissen über interessante Bewegungsabläufe der Szenenobjekte. Sie dienen als Bindeglied zwischen den in 3.1 vorgestellten Ereignisprädikaten und denen aus einer geometrischen Szenenbeschreibung zu abstrahierenden Ereigniskonzepten. Ein wichtiges Prinzip besteht darin, die Ereignismodelle in einer aus unterschiedlichen Abstraktionsstufen bestehenden Hierarchie anzuordnen. Auf der untersten Stufe stehen dabei die direkt über den geometrischen Daten definierten Konzepte, wie z.B. *exist* oder *move*, die zur Definition komplexerer Ereignismodelle, wie beispielsweise *überholen*, herangezogen werden. Erkannte Ereignisse sind Instantiierungen entsprechender Ereignismodelle; sie werden im folgenden auch als Ereignisinstanzen bezeichnet.

Die hier betrachtete Aufgabenstellung, die simultane Erkennung von Ereignissen, kann nur mit Hilfe einer inkrementell arbeitenden Erkennungsstrategie geleistet werden, bei der die zeitlichen Beziehungen zwischen den Subereignissen den Detektions-

prozeß steuern und Ereignisse entsprechend ihrem Ablauf schrittweise erkannt und explizit in der Wissensbasis des Systems repräsentiert werden. Mit der im Rahmen des Systems *Soccer* entwickelten Methodik zur Modellierung von Ereigniskonzepten (vgl. Rist et al. (1987), Herzog und Rist (1988)) soll diesen Anforderungen Rechnung getragen werden. Ein Ereignismodell in *Soccer* umfaßt:

- Rollen

Rollen stehen als existenzquantifizierte Platzhalter für die an einem Ereignis beteiligten Objekte. In den Ereignisinstanzen sind diese Rollen mit entsprechenden Bezeichnern für konkrete Szenenobjekte gefüllt.

- Rollenrestriktionen

Rollenrestriktionen schränken die Menge der möglichen Rollenfüller bei der Instantiierung eines Ereignismodells ein. Obligatorisch sind hierbei Typrestriktionen, die angeben, welcher Objektklasse die Rollenfüller angehören müssen. Des weiteren können durch Rollenrestriktionen auch Bedingungen formuliert werden, die sich auf Abhängigkeiten zwischen den einzelnen Rollenfüllern beziehen. Solche Restriktionen sind typischerweise von der Form: 'Falls der Füller der Rolle *A* die Eigenschaft *p* besitzt, dann muß der Füller der Rolle *B* die Eigenschaft *q* besitzen'.

- Ablaufschema

Das Kernstück eines Ereignismodells ist sein Ablaufschema. Es dient dazu, den prototypischen Ablauf eines Ereignisses zu spezifizieren.

Das Ablaufschema eines Ereignismodells, formal als endlicher gerichteter markierter Graph definiert, spezifiziert die Sub-Ereignisse bzw. den situativen Kontext, der in einer Szene beobachtbar sein muß, um von einem Auftreten des entsprechenden Ereignisses sprechen zu können. Der zugrundeliegende Gedanke ist der, daß die temporalen Aspekte so repräsentiert werden, daß das Erkennen eines Ereignisses zu einer Traversierung des dazugehörigen Ablaufschemas korrespondiert. Eine solche Traversierung erfolgt dabei schrittweise innerhalb eines vorgegebenen Zeittaktes — also inkrementell. Zur Demonstration sei als Beispiel das Konzept `Pass_in_den_Lauf` gewählt. Es beschreibt die Situation, in der ein Spieler seinem laufenden Mannschaftskameraden den Ball zuspielt. Im Formalismus von Allen (vgl. Allen (1984)) könnte dieses Konzept wie folgt definiert werden:

```
OCCUR(timeintervall1 (Pass_in_den_Lauf Sp1 Ball Sp2))  
<==>  
  EXIST  timeinterval2  
    DURING(timeinterval2 timeintervall1)  
    OCCUR(timeinterval2 (Laufen Sp2))  
    OCCUR(timeintervall1 (Zuspiel Sp1 Ball Sp2))
```

Abb. 3 zeigt das entsprechende Ablaufschema, das sich durch Projektion der intervallweise angegebenen Gültigkeitsbedingungen auf diskrete Zeitpunkte ergibt (vgl. dazu Herzog und Rist (1988)). Die Typmarkierungen an den Kanten des Ablaufschemas werden zur Definition der elementaren Ereignisprädikate herangezogen.

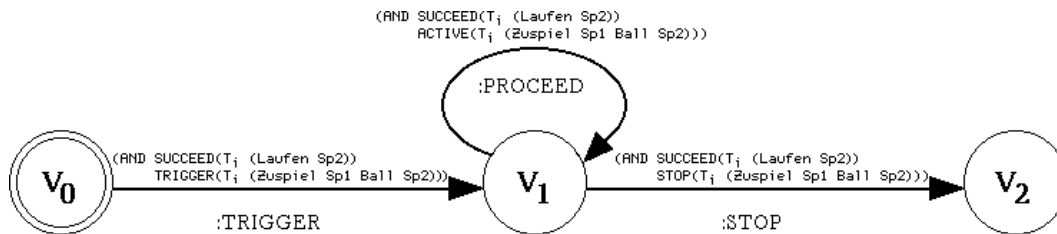


Abbildung 3: Ablaufschema des Konzeptes `Pass_in_den_Lauf`

4 Erkennung von Handlungsintentionen

4.1 Grundlagen

Wenn menschliche Beobachter Geschehnisse verfolgen und beschreiben, an denen menschliche Akteure beteiligt sind, dann lassen sie es selten bei einer Beschreibung rein visueller Phänomene bewenden. Darüber hinaus interpretieren sie, was sie sehen, d.h. sie unterstellen den Akteuren Intentionen, die deren Handlungen zugrundeliegen. Das wird z.B. auch in den folgenden Zitaten ausgedrückt: "... the 'stream of behaviour' attains orderliness in the eyes of other humans to the extent that goals and motives are imputed to the behaviour." (Dickman, 1963, p. 40). "... understanding involves inferring the intentions (i.e. the plans and goals) of the characters, speakers, or actors." (Foss und Bower, 1986, p. 94).

Solche Interpretationen spiegeln sich auch in sprachlichen Beschreibungen von visuell wahrgenommenem Geschehen wider. So können z.B. Intentionen explizit formuliert werden, wie in "Sie will A tun". Weiterhin können bestimmte Verben verwendet werden, die eine Intention implizieren, wie z.B. "verfolgen" (vgl. (Miller und Johnson-Laird, 1976, p. 101, 475), (Pleines, 1975, p. 67), (Kenny, 1963, p. 184)). Es können auch Beziehungen zwischen Intentionen oder zwischen Handlungen und Intentionen ausgedrückt werden, z.B. durch Finalsätze, wie in "Er tat A, um B zu erreichen". Dies sind nur einige wenige Beispiele aus der Fülle von Möglichkeiten, die die natürliche Sprache dafür bietet.

Ein Ziel des Projektes *Vitra* ist es, geeignete Verfahren für die automatische Bildung von Hypothesen über Intentionen aus beobachtetem Verhalten (wir sprechen im folgenden von *Intentionserkennung*) zu entwickeln. Dies ist die Grundlage dafür, daß bei der natürlichsprachlichen Beschreibung auch solche Formulierungen verwendet werden können, die Intentionen ausdrücken.

Unter *Intentionen* sollen im folgenden solche Ziele des Akteurs verstanden werden, die von ihm innerhalb der nächsten Zeit angestrebt werden, und die ihm auch in diesem Zeitraum erreichbar erscheinen. Nicht unter den Begriff der Intention fallen also längerfristige Ziele und Ziele, über deren Umsetzung sich der Akteur noch keine Gedanken gemacht hat (vgl. Baier (1970), Cohen (1986), Cohen und Levesque (1987)). Infolgedessen stehen Intentionen in enger Verbindung zu Plänen. „An intention is a want that is distinguished by having been incorporated into an agent's plan.”(Appelt, 1985, p. 48). The term '*intention*' is used to refer to the uncompleted parts of a plan whose execution has already begun”((Miller et al., 1960, p. 61), im Text hervorgehoben). Für die Erkennung von Intentionen kann also auf Planerkennungsverfahren zurückgegriffen werden. Bekannte Ansätze hierfür stellen z.B. die Arbeiten von Schmidt, Sridharan und Goodson Schmidt et al. (1978) sowie von Kautz Kautz und Allen (1986) dar.

Als Information stehen dem Beobachter einerseits die beobachteten Handlungssequenzen⁴ und andererseits allgemeines Wissen über die Zusammenhänge zwischen Zielen sowie zwischen Zielen und Plänen⁵ in gewissen Bereichen zur Verfügung. Dabei sind Ziele hierarchisch organisiert. Sie lassen sich in Konjunktionen oder Disjunktionen von Subzielen dekomponieren. Psychologische Experimente zeigen, daß menschliche Beobachter für die Erkennung von Intentionen auf solcherart strukturiertes Wissen zurückgreifen (Foss und Bower (1986), Brewer und Dupree (1983), Lichtenstein und Brewer (1980)). Es gibt auch KI-Systeme, die solche hierarchischen Zielstrukturen aus der Beobachtung lernen (Soloway und Riseman (1977)).

4.2 Das System REPLAI

Im folgenden wird das in *Soccer* integrierte Teilsystem *Replai*⁶ dargestellt, das – aufbauend auf die von *Soccer* erkannten Ereignisse – Intentionen in kurzen Ausschnitten aus Fußballspielen erkennt.

In Abb. 4 ist der Aufbau von *Replai* dargestellt. *Replai* bekommt als Eingabe zu jedem Zeitpunkt die von *Soccer* inkrementell erkannten Ereignisse sowie - auf Anfrage - weitere Informationen (wie z.B. Mannschaftszugehörigkeit von Spielern oder räumliche Relationen zwischen Spielern). Diese Informationen werden von der Komponente *scene handler* in das Format überführt, mit dem *Replai* arbeitet.

Das System verfügt über zwei Wissensquellen, die *plan hierarchy*, die allgemeines Wissen über Ziele und Pläne im Fußball enthält, und das *agent model*, das Wissen über einzelne Spieler und deren Präferenzen enthält.

In der bestehenden Version von *Replai* ist die *plan hierarchy* ein gerichteter Baum,

⁴Sequences of acts are important in identifying an agent's purpose because often one engages in an extended program of action all aimed at a single goal.”(Goldman, 1970, p. 117)

⁵“... people are forced to use their general knowledge of human intentionality to fill in the missing information; they do this by generating expectations and drawing inferences in order to come up with a plan that explains an actor's behavior.”(Foss und Bower, 1986, p. 94)

⁶REcognition of PLans And Intentions; siehe auch Retz-Schmidt (1988)

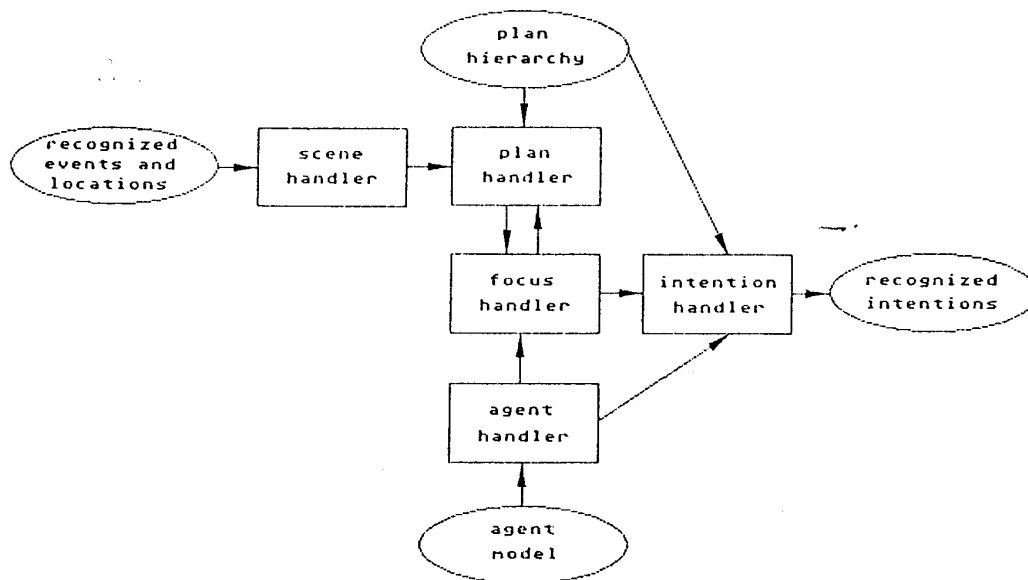


Abbildung 4: Die Architektur von *Replai*

dessen Knoten Ziele repräsentieren, die Handlungskonzepten im Fußball entsprechen (vgl. z.B. (Talaga, 1979, p. 53)). Die Ziele sind durch AKO⁷-Kanten miteinander verbunden, werden also disjunktiv in ihre Subziele zerlegt. Ein Ausschnitt aus dieser Planhierarchie ist in Abb. 5 zu sehen⁸. An den Blättern dieses Baumes hängen sogenannte *plan schemata*. Dabei handelt es sich um Folgen von direkt beobachtbaren (d.h. von *Soccer* erkennbaren) Ereignissen. Deren zeitliche Ordnung ist durch Kanten, die zeitliche Aufeinanderfolge repräsentieren, dargestellt. Es können auch Alternativen und Wiederholungen vorkommen. Beispiele für solche Planschemata sind in Abb. 6 und 7 zu sehen. Die Planhierarchie ähnelt in ihrem Aufbau dem in dem Planerkenntnisansatz von Azarewicz et al. verwendeten Wissen. Dort gibt es eine *hierarchische Komponente* mit einer festen Anzahl von Ebenen sowie eine Menge *deterministischer endlicher Automaten*, in denen die zeitliche Abfolge von Ereignissen repräsentiert wird (Azarewicz et al. (1986).

Das agent model enthält für einzelne Spieler numerische Bewertungen, die deren Präferenzen für einzelne Aktionen (z.B. Dribbeln), für ganze Pläne (z.B. Sololauf) oder für Interaktionen mit anderen Spielern (z.B. einem bestimmten Mannschaftskameraden oder dem direkten Gegenspieler) ausdrücken. Dieses Wissen kann ebenfalls von Bedeutung für die Intentionserkennung sein (vgl. (Goldman, 1970, p. 119): „Information about an agent's likes and dislikes [...] can be used to help determine his current wants.”)

⁷A Kind Of

⁸Die Variablen sind auf folgende Weise abgekürzt: at - agent's team, ot - opposing team, a - agent, o - opponent, r - recipient. Weitere Abkürzungen: p - preconditions, is - intended state.

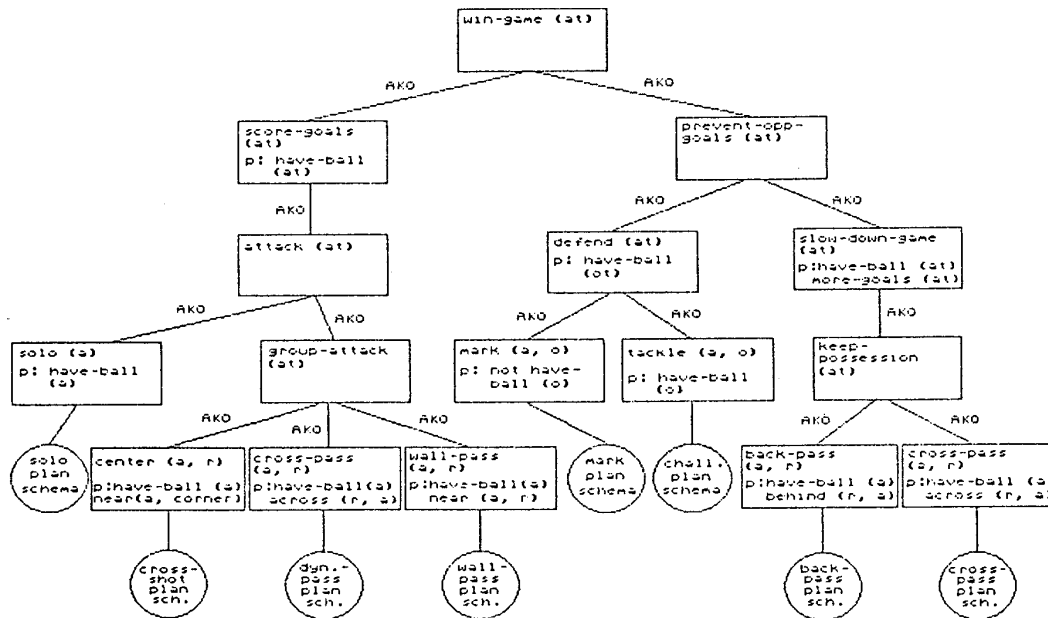


Abbildung 5: Ein Teil der Planhierarchie in *Replai*

Der eigentliche Planerkennungsprozeß wird von der Komponente *plan handler* ausgeführt und besteht aus zwei Teilprozessen. Der *top-down-Prozeß* startet an der Wurzel der Planhierarchie, geht zu den Nachfolger-Knoten und überprüft, ob die dort angegebenen *preconditions* in der aktuellen Situation für einen bestimmten Akteur erfüllt sind. Ist das der Fall, werden die jeweiligen Nachfolger-Knoten überprüft und so weiter, bis die Blätter der Hierarchie erreicht sind. Auf diese Weise werden ein oder mehrere Pfade durch den Baum, vom allgemeinsten bis zu einem sehr spezifischen Ziel gefunden. Es wird also ein Subgraph ausgesondert (cf. (Pollack, 1986, p. 17)), der nur noch den Teil der Hierarchie enthält, der unter den aktuellen Bedingungen plausibel erscheint. Ist ein solcher Subgraph gefunden, überprüft der *Plan-Schema-Prozeß* für jeden gefundenen Pfad durch die Planhierarchie, ob die Aktion und die *preconditions* des Startknotens des entsprechenden Planschemas für den jeweiligen Akteur erfüllt sind. Ist das der Fall, wird dieses Planschema für den entsprechenden Akteur aktiviert, d.h. es wird eine Instanz dieses Planschemas mit den entsprechenden Variablenbindungen angelegt. Dabei kann es auch vorkommen, daß es für eine Variable (wie z.B. *recipient*) mehrere mögliche Bindungen gibt. Dann werden mehrere Instanzen angelegt. Im weiteren Verlauf werden die aktivierten Planschemata vom Plan-Schema-Prozeß weiterverfolgt und aktualisiert. D.h. es wird überprüft, ob zu den (zeitlichen) Nachfolger-Knoten übergegangen werden kann, ob das Planschema schon fertig traversiert worden ist, oder aus anderen Gründen nicht mehr weiterverfolgt werden soll.

Der *focus handler* dient dazu, die Anzahl der betrachteten Akteure einzuschränken, da es weder sinnvoll noch möglich erscheint, die Handlungen aller 22 Spieler über

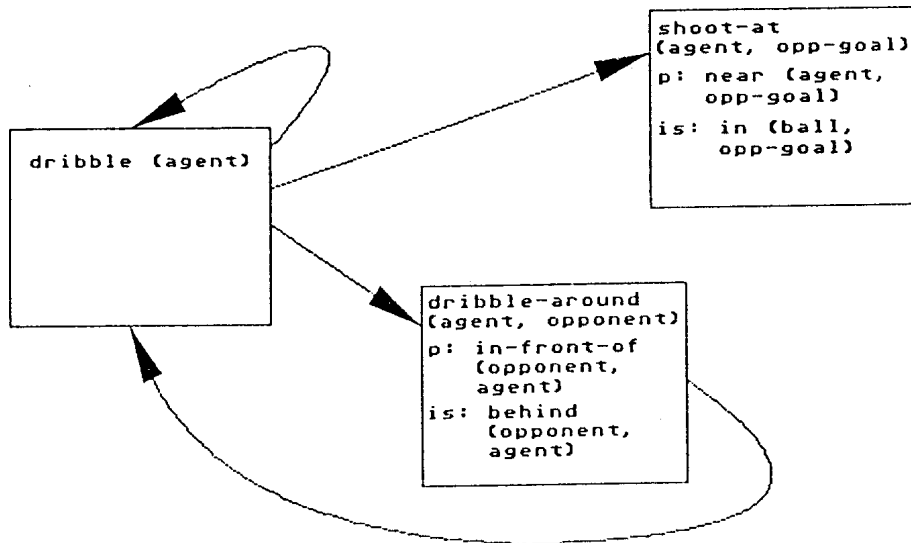


Abbildung 6: Planschema für Sololauf

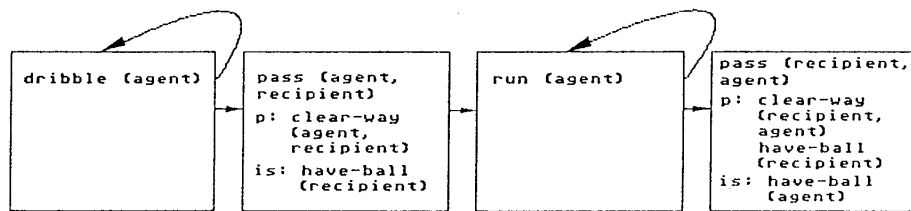


Abbildung 7: Planschema für Doppelpaß

die gesamte Zeit hinweg zu verfolgen. Die Auswahl der Akteure, die in den Fokus hinein bzw. wieder heraus kommen, geschieht mit Hilfe von Heuristiken, die in Abb. 8 dargestellt sind. Interessante Aktionen sind z.B. dribbeln oder zuspielen. – Nur fokussierte Akteure werden im Planerkennungprozess berücksichtigt. D.h. der top-down- und der Plan-Schema-Prozess werden nur für Akteure, die in den Fokus gelangen, angestoßen.

Der *agent handler* benutzt das Wissen aus dem agent model über die Präferenzen einzelner Spieler, um die Hypothesen über aktuell von einem Akteur verfolgte Pläne und Intentionen auf ihre Plausibilität hin zu bewerten. Diese Bewertungen werden dann von zwei Komponenten verwendet: vom focus handler, um Akteure mit zu unwahrscheinlichen Plänen aus dem Fokus zu entfernen, d.h. um die Menge der betrachteten Akteure einzuschränken, und vom intention handler (s.u.), um unter konkurrierenden Hypothesen für die aktuellen Intentionen eines Akteurs auszuwählen.

Der *intention handler* hat die Aufgabe, aus den erkannten Plänen die Intentionen der Akteure abzuleiten. Dabei werden drei Arten von Intentionen unterschieden: Ein

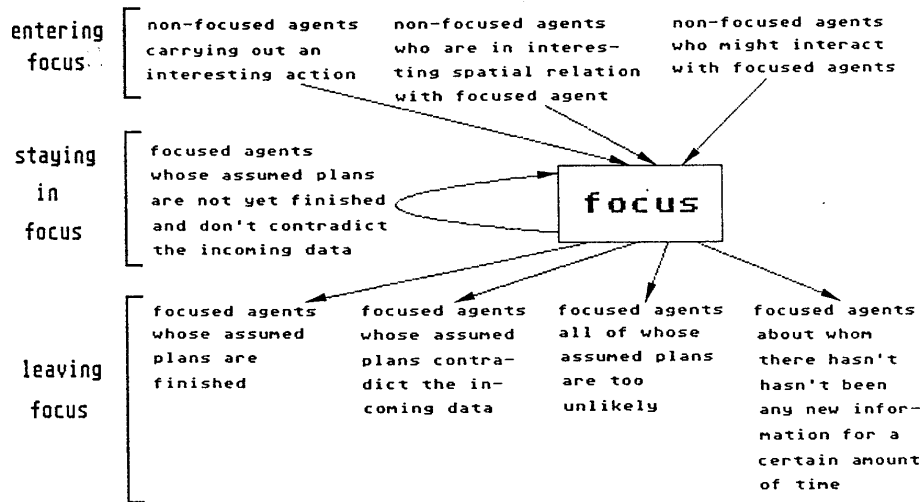


Abbildung 8: Heuristiken zur Auswahl der Fokus-Elemente in *Replai*

intendierter Zustand ist ein Zustand, den der Akteur durch seine aktuelle Handlung anstrebt. Eine *intendierte Aktion* ist eine Handlung, die der Akteur in der nächsten Zeit auszuführen plant. Ein *Oberziel* ist ein Ziel, das die aktuell ausgeführte Handlungssequenz veranlaßt hat. Ähnliche Unterscheidungen finden sich auch vielfach in philosophischen Handlungstheorien (z.B. Anscombe (1957), Audi (1973), Goldman (1970)) und in der KI-Literatur (z.B. Appelt (1985)). Der intention handler erkennt die verschiedenen Arten von Intentionen auf die folgende Weise: Der intendierte Zustand ist in einem Slot des aktuellen Plan-Schema-Knotens angegeben. Hat die aktuelle Aktion einen (zeitlichen) Nachfolger-Knoten, dessen preconditions schon erfüllt sind, so ist die entsprechende Aktion die intendierte Aktion. Das Oberziel ergibt sich aus dem dem aktuellen Plan-Schema übergeordneten Knoten der Planhierarchie. — Der intention handler wird die Schnittstelle zu einer noch zu entwickelnden Komponente für die natürlichsprachliche Beschreibung von Intentionen sowie von Zusammenhängen zwischen Intentionen sein.

5 Sprachproduktion

Um zu gewährleisten, daß sich das Wahrgenommene unmittelbar auf die Sprachproduktionsprozesse auswirken kann, arbeiten die Prozesse zur Ereigniserkennung und Sprachproduktion in *Soccer* virtuell parallel. Dies ermöglicht die Simulation von Phänomenen wie inkrementelle Sprachproduktion, unerwarteter Themenwechsel oder Auslassung ganzer Teilszenen (vgl. hierzu André et al. (1987), André (1988)). Sprachproduktion wird in *Soccer* unterteilt in *Selektion*, *Linearisierung* und *Enkodierung*.

5.1 Selektion, Linearisierung und Enkodierung

Die Auswahl der mitzuteilenden Information erfolgt durch eine *Selektionskomponente*, der insbesondere folgende Aufgaben zufallen: die Vorauswahl der erkannten Ereigniskonzepte, die Belegung der Tiefenkasus sowie die Bestimmung der Attribute bei der Referenzierung von Objekten. Diese Komponente wird durch eine Regelbasis unterstützt, die Pakete von diskursspezifischen Kommunikationsregeln enthält. Die Austauschbarkeit der Regelbasis gewährleistet die Übertragbarkeit auf andere Diskursbereiche.

Die *Linearisierungskomponente* entscheidet über die Reihenfolge, in der die vorselektierten Ereignisse erwähnt werden. Es wird ein kurzfristiger Textplan erstellt, der jedoch u.U. abgeändert werden muß, um der Situation des Simultanberichts gerecht zu werden. Beispielsweise kann das Eintreten herausragender Ereignisse dazu führen, daß vorselektierte Ereignisse nicht mehr genannt werden, da das aktuelle Geschehen von größerer Relevanz ist.

Unter die *sprachliche Enkodierung* fallen u.a. Prozesse zur Wortwahl, zur Nominalphrasengenerierung, zur Bereitstellung morphosyntaktischer Information (z.B. Numerus, Modus, Tempus) sowie zur Oberflächentransformation. Die Enkodierung eines selektierten Ereignisses beginnt mit der Auswahl eines Verbs, die sich im wesentlichen danach richtet, ob die zu dem entsprechenden Ereigniskonzept gehörigen Rollenfüller als Oberflächenkasus des Verbs auftreten können. Die Belegung des Fokus⁹ wird bei dieser Auswahl insofern berücksichtigt, als nach Möglichkeit ein Verb selektiert wird, bei dem das am stärksten fokussierte Objekt als Subjekt stehen kann. Nach der Instantiierung des Verbrachmens erfolgt die Überführung systeminterner Bezeichner in Nominalphrasen.

5.2 Objektreferenzierung

Die Referenzierung von Objekten erfolgt nach dem Prinzip der eindeutigen Benennung unter Zugriff auf ein Partnermodell. Die Koreferenz zwischen dem Wissen des Systems und dem beim Partner vermuteten Wissen wird durch ein Netzwerk, das sogenannte Koreferenzennetz repräsentiert (vgl. Jameson und Wahlster (1982)). Pronomina werden in SOCCER nur dann generiert, wenn sichergestellt ist, daß der Hörer die Referenz aufgrund von syntaktischen oder semantischen Kriterien auflösen kann. Psychologischen Modellen folgend (vgl. Guindon (1985)), fordern wir als notwendige Bedingung bei der Referenzierung durch Pronomina, daß das Objekt im Fokus ist.

Deskriptive Nominalphrasen haben nicht nur die Funktion, ein Objekt von einer Menge alternativer Objekte abzugrenzen, sondern dienen auch zur Mitteilung zusätzlicher Information. Besondere Aufmerksamkeit wurde folgenden Möglichkeiten zur Spezifikation von Objekten gewidmet:

⁹Als Fokus wird hier die Menge derjenigen Objekte bezeichnet, auf die ein Sprecher zum Zeitpunkt seiner Äußerung seine Aufmerksamkeit richtet (vgl. Sidner (1983)). Dabei handelt es sich nicht um denselben Fokus wie in *Replai*.

- Verwendung nicht-räumlicher Objektbeziehungen
Objekte werden durch Beziehungen zu anderen Objekten charakterisiert (z.B. ein Spieler als Mitglied einer Mannschaft)
- Beschreibung von Lokationen
Objekte werden durch räumliche Relationen beschrieben, wobei zwischen komparativen (z.B. *'die linke Ecke'*) und absoluten Relationen (z.B. *'der Spieler im Strafraum'*) unterschieden wird.
- Beziehung auf Ereignisse
Objekte können durch aktuelle oder vorerwähnte Ereignisse näher spezifiziert werden (z.B. *'Walter, der von Becker angegriffen wird'*).

Eine Dimensionspräferenzliste gewährleistet, daß die für den Hörer leichter verifizierbaren Attribute bevorzugt zur Abgrenzung eines Objekts von Alternativobjekten herangezogen werden.

6 Zusammenfassung und Ausblick

Bei der Konzeption, die bei der Kopplung von *Vitra* und *Actions* verfolgt wird, besteht die Aufgabe der Bildanalyse in der Erzeugung einer quantitativen Beschreibung grundlegender wahrnehmbarer Größen, wie z.B. Position und Geschwindigkeit von Objekten. Ausgehend von dieser geometrischen Szenenbeschreibung stellen weitergehende Interpretationsprozesse qualitative Beschreibungen räumlicher Anordnungen und zeitübergreifender Vorgänge in Form von räumlichen Relationen und Ereignissen zur Verfügung. Hierauf aufbauend können in einem weiteren Interpretationsschritt Intentionen der agierenden Objekte abgeleitet werden. Diese verschiedenen konzeptuellen Strukturen sind das Bindeglied zwischen visuellen Daten und sprachlichen Einheiten wie räumliche Präpositionen bzw. Bewegungs- und Handlungsverben.

Eine Besonderheit des in *Soccer* verfolgten Ansatzes ist die Zielsetzung, eine Szene simultan zu ihrem Ablauf zu interpretieren und in natürlicher Sprache zu beschreiben. Langfristig betrachtet wird damit die Realzeitverarbeitung bei der Analyse, Interpretation und natürlichsprachlichen Beschreibung zeitveränderlicher Szenen angestrebt (vgl. auch Nagel (1988)).

Die Leistungsfähigkeit von *Soccer* läßt sich demonstrieren, indem kurze Szenen in einem Graphikfenster visualisiert und dazu simultan natürlichsprachlich beschrieben werden. Dabei kann wahlweise geschriebener oder gesprochener deutscher Text erzeugt werden. Abb. 9 zeigt einen Ausschnitt aus einer typischen Beschreibung. Zum Zeitpunkt [00:11:20] wird als mitteilenswertig selektiert, daß `spieler11` in Ballbesitz ist und zum Tor läuft. Während der Enkodierung dieser Information wird festgestellt, daß der Spieler angegriffen wird. Die zeitliche Überschneidung der beiden zuletzt genannten Ereignisse wird durch das Temporaladverb "dabei" ausgedrückt.

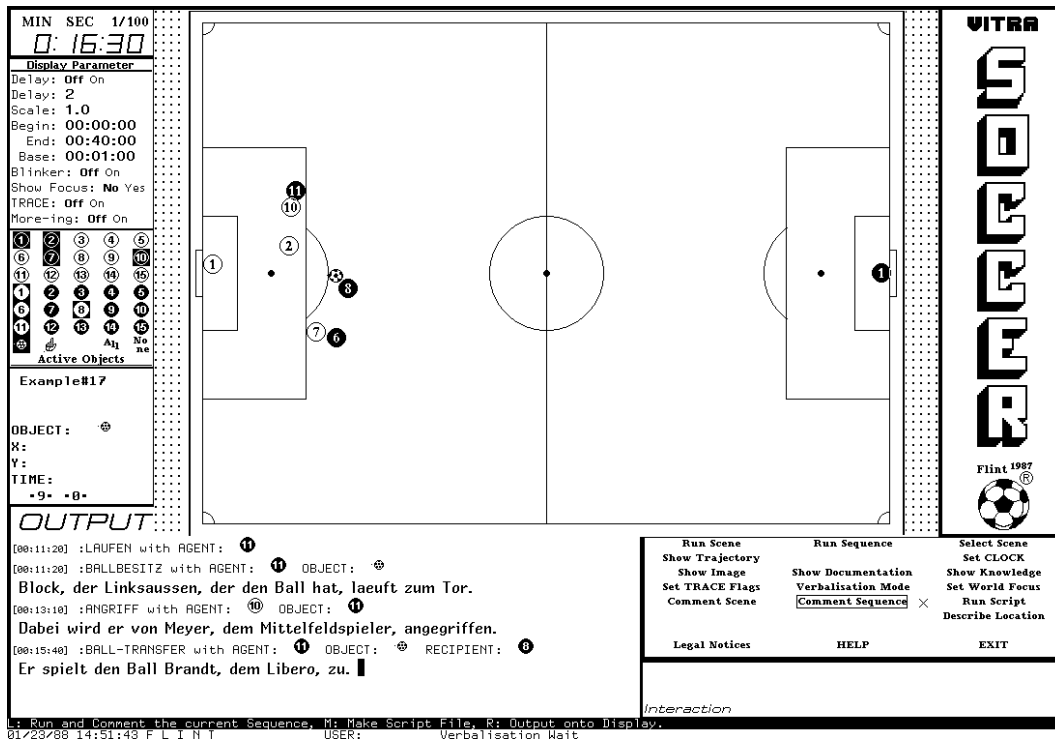


Abbildung 9: Die Soccer-Bildschirmmaske

Da *spieler11* st rker fokussiert ist als *spieler10* wird eine Passivkonstruktion gew hlt. Schlielich wird das Zuspiel zu *spieler8* mitgeteilt. Bemerkte sei, da es sich bei den Ereignissen unseres Beispiels ausschlielich um Ereignisse handelt, die zum Zeitpunkt der Enkodierung noch nicht abgeschlossen sind, was die Wahl des Pr sens als *Tempus* erkl rt.

Um die Interpretationsleistung des Systems *Soccer* und auch die Qualit t der erzeugten nat rlichsprachlichen Simultanbeschreibungen noch zu verbessern, sollen in einer weiteren Ausbaustufe die folgenden Erweiterungen durchgef hrt werden, die z.T. auch unter dem Gesichtspunkt des Sports von Interesse sein k nnten:

- Verbalisierung erkannter Intentionen

Die erkannten Intentionen sollen in nat rlicher Sprache beschrieben werden. Dabei soll von den verschiedenen M glichkeiten der nat rlichen Sprache Gebrauch gemacht werden. D.h. Intentionen sollen explizit (durch Verwendung von "wollen") und implizit (durch Verwendung von Verben, die Intentionen implizieren) ausgedr ckt werden. Weiterhin sollen Beziehungen zwischen mehreren Intentionen eines Akteurs sowie zwischen Intentionen und Handlungen (z.B. durch Finals tze) formuliert werden. Schlielich sollen die erkannten Pl ne genauer daraufhin analysiert werden, inwieweit sie zu einem erfolgreichen Abschlu gebracht werden konnten, bzw. was im Falle von Mierfolgen die Ursache

chen waren. Die Ergebnisse dieser Analyse sollen ebenfalls natürlichsprachlich mitgeteilt werden.

- **Behandlung von Objektgruppen**

Läßt man als Rollenfüller bei der Definition von Ereignissen auch Objektgruppen zu, dann könnten auch Konzepte wie beispielsweise *Angriff_des_Sturms* innerhalb des Formalismus definiert werden. Auch für das Teilsystem *Replai* ist eine solche Behandlung von Objektgruppen interessant. So sollen in Zukunft Intentionen von ganzen Gruppen von Spielern, die zu einem koordinierten Verhalten führen (wie z.B. bei einer Abseitsfalle), erkannt werden können. Eine besondere Schwierigkeit hierbei liegt sicherlich darin, daß sich Objektgruppen, im Gegensatz zu einzelnen Objekten, neu formieren, vergrößern, verkleinern oder auflösen können; d.h. ihre Gestalt als auch ihre zeitliche Existenz ist variabel. Hinzu kommt noch, daß die Menge aller möglichen Objektgruppen exponentiell zur Menge der dynamischen Szenenobjekte wächst.

- **Entwicklung eines visuellen Hörermodells**

Im Hinblick auf die Generierung adäquater Beschreibungen scheint es sinnvoll, die beim Hörer vermutlich erzeugten visuellen Vorstellungen explizit zu modellieren und bei der weiteren Sprachproduktion zu berücksichtigen (vgl. Wahlster (1987), Schirra (1989)). Durch Abgleich der Imagination des Hörers mit den tatsächlichen Szenendaten läßt sich dann verifizieren, ob eine Äußerung den intendierten Effekt hat.

7 Technische Anmerkungen

Die aktuelle Version des *Vitra*-Systems wurde in Commonlisp und Flavors auf Symbolics Lisp-Maschinen vom Typ 3600 und 3640 unter Release 7.1 implementiert, wobei eine der Maschinen mit einer Farbgraphikerweiterung OP36-C108 und einem zusätzlichem Farbbildschirm ausgestattet ist.

Die Sprachausgabe erfolgt mithilfe eines *AEG-SVS* Sprachsynthesemoduls, das über eine serielle Schnittstelle mit der Symbolics verbunden ist. Als Eingabe erwartet das Sprachsynthesemodul ASCII-Zeichen und generiert daraus gesprochenes Deutsch.

Danksagung

Die hier beschriebene Arbeit wurde teilweise unterstützt vom Sonderforschungsbereich 314 der Deutschen Forschungsgemeinschaft, "Künstliche Intelligenz und wissensbasierte Systeme", Projekt N2: VITRA (VIsual TRAnslator).

Für Anmerkungen und Kommentare zu einer Vorversion dieses Beitrags danken wir Jörg Schirra und unseren Kollegen vom Institut für Informations- und Datenverar-

beitung (IITB) der Fraunhofer-Gesellschaft in Karlsruhe, die uns auch die Abbildung zur Funktionsweise des Systems *Actions* zur Verfügung gestellt haben.

Literatur

- J. F. Allen.** Towards a General Theory of Action and Time. *Artificial Intelligence*, **23** (2), 123–154, 1984.
- E. André.** Generierung natürlichsprachlicher Äußerungen zur simultanen Beschreibung von zeitveränderlichen Szenen: Das System SOCCER. Memo 26, Universität des Saarlandes, SFB 314 (VITRA), Saarbrücken, 1988.
- E. André, G. Bosch, G. Herzog, T. Rist.** Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions. In: *Proc. of the 7th ECAI*, Band 2, S. 1–8, Brighton, UK, 1986.
- E. André, G. Herzog, T. Rist.** On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: The System SOCCER. In: *Proc. of the 8th ECAI*, S. 449–454, Munich, 1988.
- E. André, T. Rist, G. Herzog.** Generierung natürlichsprachlicher Äußerungen zur simultanen Beschreibung zeitveränderlicher Szenen. In: K. Morik (Red.), *GWAI-87. 11th German Workshop on Artificial Intelligence*, S. 330–337, Springer, Berlin, Heidelberg, 1987.
- G. Anscombe.** *Intention*. Basil Blackwell, Oxford, 1957.
- D. E. Appelt.** *Planning English Sentences*. Cambridge University Press, Cambridge, London, 1985.
- R. Audi.** Intending. *Journal of Philosophy*, **70** (13), 387–403, 1973.
- J. Azarewicz, G. Fala, R. Fink, C. Heithecker.** Plan Recognition for Airborne Tactical Decision Making. In: *Proc. of AAAI-86*, S. 805–811, Philadelphia, PA, 1986.
- A. Baier.** Act and Intent. *Journal of Philosophy*, **67** (19), 648–658, 1970.
- D. H. Ballard, C. M. Brown.** *Computer Vision*. Prentice-Hall, Englewood Cliffs, NJ, 1982.
- W. Brandt.** *Zeitstruktur und Tempusgebrauch in Fußballreportagen des Hörfunks*. Elwert, Marburg, 1983.
- W. Brewer, D. Dupree.** Use of Plan Schemata in the Recall and Recognition of Goal-Directed Actions. *Journal of experimental Psychology: Learning, Memory, and Cognition*, **9**, 117–129, 1983.

- P. R. Cohen.** Communication as Rational Interaction. *CSLI Monthly*, **2** (2), 1986.
- P. R. Cohen, H. J. Levesque.** Intention = Choice + Commitment. In: *Proc. of AAAI-87*, S. 410–415, Seattle, WA, 1987.
- H. Dickman.** The Perception of Behavioral Units. In: R. Barker (Red.), *The Stream of Behavior*, S. 23–41, Appleton-Century-Crofts, New York, 1963.
- C. Foss, G. Bower.** Understanding Actions in Relation to Goals. In: N. Sharkey (Red.), *Advances in Cognitive Sciences 1*, S. 94–124, Ellis Horwood, Chichester, 1986.
- A. Goldman.** *A Theory of Human Action*. Prentice-Hall, Englewood Cliffs, NJ, 1970.
- R. Guindon.** Anaphora Resolution: Short-Term Memory and Focusing. In: *Proc. of the 23th ACL*, S. 218–227, Chicago, IL, 1985.
- G. Herzog, T. Rist.** Simultane Interpretation und natürlichsprachliche Beschreibung zeitveränderlicher Szenen: Das System SOCCER. Memo 25, Universität des Saarlandes, SFB 314 (VITRA), Saarbrücken, 1988.
- A. Jameson, W. Wahlster.** User Modelling in Anaphora Generation. In: *Proc. of the 5th ECAI*, S. 222–227, Orsay, France, 1982.
- H. Kautz, J. Allen.** Generalized Plan Recognition. In: *Proc. of AAAI-86*, S. 32–37, Philadelphia, PA, 1986.
- A. Kenny.** *Action, Emotion and Will*. Humanities, New York, 1963.
- E. Lichtenstein, W. Brewer.** Memory for Goal-Directed Actions. *Cognitive Psychology*, **12**, 412–445, 1980.
- D. McDermott.** A Temporal Logic for Reasoning about Processes and Plans. *Cognitive Science*, **6**, 101–155, 1982.
- G. Miller, E. Galanter, K. Pribram.** *Plans and the Structure of Behaviour*. Holt, Rinehart and Winston, New York, 1960.
- G. A. Miller, P. N. Johnson-Laird.** *Language and Perception*. Cambridge University Press, Cambridge, London, 1976.
- H.-H. Nagel.** Analyse und Interpretation von Bildfolgen. Teil I und II. *Informatik Spektrum*, **8**, 178–200, 312–327, 1985.
- H.-H. Nagel.** From Image Sequences Towards Conceptual Descriptions. *Image and Vision Computing*, **6** (2), 59–74, 1988.

- B. Neumann.** Bildverstehen. In: W. Bibel, J. Siekmann (Red.), *Künstliche Intelligenz*, S. 285–355, Springer, Berlin, Heidelberg, 1982.
- B. Neumann, H.-J. Novak.** Event Models for Recognition and Natural Language Description of Events in Real-World Image Sequences. In: *Proc. of the 8th IJCAI*, S. 724–726, Karlsruhe, FRG, 1983.
- B. Neumann, H.-J. Novak.** NAOS: Ein System zur natürlichsprachlichen Beschreibung zeitveränderlicher Szenen. *Informatik Forschung und Entwicklung*, **1**, 83–92, 1986.
- J. Pleines.** Kausale Relationen und Intentionalität. In: V. Ehrich, P. Finke (Red.), *Beiträge zur Grammatik und Pragmatik*, S. 55–70, Scriptor, Kronberg/Taunus, 1975.
- M. Pollack.** *Inferring Domain Plans in Question Answering*. Doktorarbeit, Dept. of Computer Science, Moore School, Univ. of Pennsylvania, Philadelphia, PA, 1986.
- G. Retz-Schmidt.** Various Views on Spatial Prepositions. *AI Magazine*, **9** (2), 95–105, 1988.
- T. Rist, G. Herzog, E. André.** Ereignismodellierung zur inkrementellen High-level Bildfolgenanalyse. In: E. Buchberger, J. Retti (Red.), *3. Österreichische Artificial-Intelligence-Tagung*, S. 1–11, Springer, Berlin, Heidelberg, 1987.
- D. Rosenbaum.** *Die Sprache der Fußballreportage im Hörfunk*. Doktorarbeit, Fachbereich Germanistik, Univ. des Saarlandes, 1969.
- J. R. J. Schirra.** Ein erster Blick auf ANTLIMA: Visualisierung statischer räumlicher Relationen. In: D. Metzger (Red.), *GWAI-89: 13th German Workshop on Artificial Intelligence*, S. 301–311, Springer, Berlin, Heidelberg, 1989.
- J. R. J. Schirra, G. Bosch, C.-K. Sung, G. Zimmermann.** From Image Sequences to Natural Language: A First Step Towards Automatic Perception and Description of Motions. *Applied Artificial Intelligence*, **1**, 287–305, 1987.
- C. Schmidt, N. Sridharan, J. Goodson.** The Plan Recognition Problem: An Intersection of Psychology and Artificial Intelligence. *Artificial Intelligence*, **11**, 45–83, 1978.
- P. Schneider.** *Die Sprache des Sports: Terminologie und Präsentation in Massenmedien*. Schwann, Düsseldorf, 1974.
- C. L. Sidner.** Focusing in the Comprehension of Definite Anaphora. In: M. Brady, R. C. Berwick (Red.), *Computational Models of Discourse*, S. 267–330, MIT Press, Cambridge, MA, 1983.

- E. Soloway, E. Riseman.** Knowledge-Directed Learning. In: *ACM SIGART Newsletter 63. Proc. of the Workshop on Pattern-Directed Inference Systems*, S. 49–55, New York, 1977.
- C.-K. Sung.** Extraktion von typischen und komplexen Vorgängen aus einer langen Bildfolge einer Verkehrsszene. In: H. Bunke, O. Kübler, P. Stucki (Red.), *Mustererkennung 1988; 10. DAGM Symposium*, S. 90–96, Springer, Berlin, Heidelberg, 1988.
- C.-K. Sung, G. Zimmermann.** Detektion und Verfolgung mehrerer Objekte in Bildfolgen. In: G. Hartmann (Red.), *Mustererkennung 1986; 8. DAGM-Symposium*, S. 181–184, Springer, Berlin, Heidelberg, 1986.
- J. Talaga.** *Fußballtaktik*. Sportverlag Berlin, Berlin, 1979.
- J. K. Tsotsos.** Temporal Event Recognition: An Application to Left Ventricular Performance. In: *Proc. of the 7th IJCAI*, S. 900–907, Vancouver, Canada, 1981.
- W. Wahlster.** Ein Wort sagt mehr als 1000 Bilder. Zur automatischen Verbalisierung der Ergebnisse von Bildfolgeanalyse-Systemen. *Annales, Forschungsmagazin der Univ. des Saarlandes*, 1 (1), 82–93, 1987.
- W. Wahlster, H. Marburger, A. Jameson, S. Busemann.** Over-answering Yes-No Questions: Extended Responses in a NL Interface to a Vision System. In: *Proc. of the 8th IJCAI*, S. 643–646, Karlsruhe, FRG, 1983.
- I. Walter.** EPEX: Bildfolgendeutung auf Episodenebene. In: K. Morik (Red.), *GWAI-87. 11th German Workshop on Artificial Intelligence*, S. 21–30, Springer, Berlin, Heidelberg, 1987.