Perzeptuelle Verankerung ausgedehnter Landmarken während der instruierten Navigation

Felix Lindner

Überarbeitete Version vom 4. November 2009
Originalabgabedatum: 10. Juli 2009

Diplomarbeit im Studiengang Informatik am
Arbeitsbereich Wissens- und Sprachverarbeitung,
Zentrum für Intelligente Systeme und Robotik,
Department Informatik,
Fakultät für Mathematik, Informatik und Naturwissenschaften,
Universität Hamburg

Erstgutachterin: Dr. Carola Eschenbach AB WSV

Zweitgutachter: Prof. Dr. Jianwei Zhang AB TAMS

Zusammenfassung

Symbolische Künstliche Intelligenz trägt entscheidend dazu bei, die kognitive Leistungsfähigkeit von Robotern zu steigern und ihre Einsatzmöglichkeiten zu erweitern. Eine Voraussetzung dafür ist die Verknüpfung symbolischer Repräsentationen mit korrespondierenden Repräsentationen auf Ebene der Sensorik und Aktorik innerhalb einer Roboterarchitektur. Die Verknüpfung von Symbolen und Perzepten, die dieselben physikalischen Objekte in der Welt denotieren, wird in der Literatur unter der Bezeichnung „perzeptuelle Verankerung“ aktuell diskutiert. Während der instruierten Navigation löst ein Agent das Problem der perzeptuellen Verankerung, wenn er die in der Routeninstruktion beschriebenen räumlichen Konstellationen in der realen Welt identifiziert. Der Beitrag der vorliegenden Arbeit besteht in der Modellierung der grundlegenden Strukturen für die perzeptuelle Verankerung ausgedehnter Landmarken in der Architektur eines instruierten Agenten. Basierend auf der Analyse des Rahmenwerks von Coradeschi & Saffiotti (2000) wird ein Verankerungsverfahren für den Fall ausgedehnter Landmarken vorgeschlagen. Es werden Verfahren zur Wiedererkennung von Landmarken, zur Rekonstruktion von Landmarkengeometrien und ihrer Aggregation zu einer kognitiven Karte untersucht. Der Agent erwirbt während der Navigation ein internes Modell seiner Umgebung, das er für räumliches Schließen und zur lokalen Aktionsplanung nutzt. Die Ergebnisse werden schließlich experimentell anhand von Simulationen und in Roboter-Experimenten evaluiert.

Abstract

Symbolic artificial intelligence in robotics significantly improves cognitive skills of robotic systems in general. To benfit from the state of the art in artificial intelligence and cognitive science, it must be possible to connect symbolic representations to corresponding sensory information in a complex robotic system. A crucial instance of this problem, i.e. the connection between percepts and symbolic descriptions of objects within an artificial system (referred to as perceptual anchoring), is just heavily discussed in current literature on cognitive robotics. Any agent that follows a route-instruction solves this problem when it identifies the spatial constellations mentioned in the instruction. In this thesis, a framework for perceptual anchoring of extended landmarks based on an analysis of the framework by Coradeschi & Saffiotti (2000) is proposed. The model consists of components to reidentify landmarks, to incrementally reconstruct landmarks' geometry and to aggregate those geometries to a mental map. The main objective is to construct a experience-based knowledge base that can be used by an agent to reason about and acting in its environment while navigating through. The model will be evaluated in both simulation experiments and within a real-world robot setting.

Danksagung

Ganz herzlich bedanke ich mich bei all jenen Mitarbeitern aus den Arbeitsbereichen WSV und TAMS und bei den Teilnehmern des Graduiertenkollegs CINACS, die in zahlreichen konstruktiven Gesprächen während der letzten 6 Monate meine Sicht auf das Thema dieser Diplomarbeit kommentiert und bereichert haben.

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einführung

Es wird für die Zukunft erwartet, dass Service-Roboter ihren Einzug in den Alltag der Menschen finden werden. Fragt man Menschen, welche Fähigkeiten sie von persönlichen Service-Robotern erwarten, nennen sie Aufgaben wie „Lebensmittel zubereiten“, „Aufgaben in Haushalt und Garten übernehmen“, „Einkäufen tätigen“ und „sich um die Kinder/Alten/Haustiere kümmern“ (Copleston & Bugmann, 2008). Vielversprechende Zielvorstellungen werden zudem im Zusammenhang von Assistenzrobotern formuliert, die die Mobilität und Eigenständigkeit von Menschen mit Einschränkungen im Alltag unterstützen (Oestreicher & Severinson-Eklundh, 2006; Gharpure & Kulyukin, 2008; Shi & Tenbrink, 2009).

Alle genannten Aufgaben haben einen inhärenten Raumbezug. Das ist zunächst nicht verwunderlich, liegt die Besonderheit mobiler Roboter im Gegensatz zu anderen Computersystemen doch gerade in ihrer Möglichkeit, die Welt wahrzunehmen und in ihr zu agieren. So werden Roboter zunächst als Diener des Menschen verstanden, die alltägliche Aufgaben übernehmen oder Menschen dazu verhelfen, Aufgaben eigenständig zu bewältigen, die sie ohne Hilfestellung nicht bewältigen könnten. Das erfordert von Robotersystem die Fähigkeit zur sicheren Navigation durch die natürliche Umgebung und zur Identifizierung und Manipulation realer Objekte. Damit Service-Roboter akzeptiert werden, müssen sie darüber hinaus von ihren Anwendern intuitiv instruiert werden können. Die natürliche Interaktion zwischen Mensch und Roboter sollte so intuitiv sein wie die Interaktion zwischen Menschen.

Menschen sind in der Lage, mit der Komplexität der Welt umzugehen und sie zu beschreiben ohne die grundlegenden physikalischen Prozesse zu kennen und einzubeziehen. Der Mensch verwendet qualitative räumliche Konzepte wie *links von/ rechts von/ vor/ hinter/ bei*, um über den Raum zu räsonieren und zu kommunizieren. Er bezeichnet Objekte bei ihrem Namen oder referenziert auf sie, indem er sie anhand ihrer Eigenschaften beschreibt. Im Gegensatz dazu ist die Wahrnehmung eines Roboters zunächst durch die Messungen seines sensorischen Apparates bestimmt. Er nimmt die Welt in Form quantitativer physikalischer Größen wie Entfernung, Winkel, Helligkeit, Temperatur oder Druck wahr, die zunächst numerisch repräsentiert sind. Die Situation

zwischen Mensch und Maschine gleicht der Situation in der von Wittgenstein (1953) getroffenen Feststellung, dass auch wenn ein Löwe sprechen könnte, der Mensch ihn nicht verstehen könnte. Selbst wenn der Roboter wohlgeformte sprachliche Aussagen über die Welt, wie er sie wahrnimmt, treffen würde, könnten Menschen sie ohne Wissen über die internen Vorgänge im robotischen System nicht mit ihrer eigenen Erfahrungswelt in Verbindung bringen. Ein Ziel der Forschung zur Service-Robotik besteht darin, dem Roboter zu ermöglichen, eine für die Mensch-Roboter-Interaktion geeignete Konzeptualisierung der Welt zu erwerben. Dazu muss die Wahrnehmung des Roboters mit der symbolischen Ebene verknüpft werden, auf der Sprache und räumliches Denken beim Menschen funktioniert.

Die Verknüpfung von Symbolen und Perzepten, die dieselben physikalischen Objekte in der Welt denotieren, wird in der Literatur unter der Bezeichnung „perzeptuelle Verankerung“ aktuell diskutiert (vgl. Coradeschi & Loutfi, 2008). Das Ziel der perzeptuellen Verankerung ist, die Koreferenz[1] zwischen perzeptuellen Eindrücken und symbolischen Beschreibungen zu identifizieren und über die Zeit aufrecht zu erhalten. Dafür werden zwei grundlegende Prozesse unterschieden: die Erkennung und Wiedererkennung von Objekten einerseits, und die Koreferenzauflösung zwischen symbolischen Beschreibungen und internen Modellen der Wahrnehmung. Die Realisierung dieser Fähigkeiten in einem Robotiksystem stellt eine Voraussetzung dafür dar, dass Mensch und Maschine über Objekte in der Welt kommunizieren können.

Eine Domäne, in der die perzeptuelle Verankerung untersucht werden kann, stellen Routeninstruktionen dar. Während der instruierten Navigation löst ein Agent das Problem der perzeptuellen Verankerung, wenn er die in der Routeninstruktion beschriebenen räumlichen Konstellationen in der realen Welt identifiziert. Ein Agent A befindet sich in einer ihm unbekannten Umgebung. Er kennt sein Ziel, nicht aber den Weg dorthin und so bittet er einen Agenten B, der den Weg vom Start zum Ziel kennt, darum, ihm den Weg zu beschreiben. Daraufhin macht sich Agent A auf den Weg entlang der beschriebenen Route. Denis (1997) identifiziert die Beschreibung von Landmarken und die Beschreibung von Aktionen als zwei Hauptbestandteile von Routeninstruktionen. Landmarken sind herausstechende, wahrnehmbare Merkmale der Umgebung (Lynch, 1960). Menschen benutzen Landmarken, um sich im Raum zu orientieren und Plätze wiederzuerkennen. In Routeninstruktionen wird auf Landmarken referenziert, um Aktionen an Landmarkenkonstellationen zu knüpfen – wie beispiels-

[1] Zwei Repräsentationen sind koreferent, wenn ihre Extensionen gleich sind. Im Falle der perzeptuellen Verankerung ist dies gleichbedeutend damit, dass sie dasselbe Objekt in der Welt bezeichnen.

weise in „Nach dem Rathaus biegst Du links ab.“ Während der Navigation identifiziert der instruierte Agent die in der Routeninstruktion spezifizierten Landmarken und führt die damit assoziierten Aktionen aus.

Im Szenario des Geometrischen Agenten (Tschander et al., 2003) gibt ein Mensch dem Agenten eine Routeninstruktion in natürlicher Sprache oder in Form einer Routenskizze. Während der Navigation identifiziert der Agent die Landmarken, Pfade und Entscheidungspunkte, die in der Instruktion erwähnt sind. Reale Landmarken wie Häuser und Wege haben eine räumliche Ausdehnung, die aus der Perspektive des Navigierenden nicht vollständig „auf einen Blick“ erfasst werden kann. Deswegen werden verschiedene Teile von Landmarken während der Navigation im Allgemeinen von unterschiedlichen Perspektiven aus wahrgenommen. Dabei stellt sich das Problem der Wiedererkennung. Wurde ein anderer Teil der Landmarke bereits zuvor gesehen, kann die neue Information dem internen Modell der Landmarke hinzugefügt werden. Während Menschen navigieren, erwerben sie durch die direkte perzeptuelle und motorische Interaktion mit der Welt Wissen über die eigene Bewegung durch den Raum und über die räumlichen Beziehungen zwischen Orten (Montello et al., 2004). In der vorliegenden Arbeit wird gezeigt, wie mittels des Rahmenwerks zur perzeptuellen Verankerung der instruierte Agent während der Navigation ein internes Modell der Umgebung aufbaut, das er für zahlreiche Schlussmechanismen nutzen kann, um der Routeninstruktion erfolgreich zu folgen.

1.1 Verwandte Arbeiten

Einen Überblick über die Arbeiten zur perzeptuellen Verankerung nach Coradeschi & Saffiotti gibt (Coradeschi & Loutfi, 2008). Eine formale Behandlung des Rahmenwerks ist in (Coradeschi & Saffiotti, 2000) gegeben. Die Rolle der perzeptuellen Verankerung für den Erwerb von Umgebungswissen in einem Innenraum-Szenario wird bei Galindo et al. (2008) behandelt. Dabei generiert der Roboter autonom eine mentale, metrische Karte seiner Umgebung und konzeptualisiert diese Umgebung anhand einer Taxonomie, die die typische Gestalt einer Wohnung beschreibt. Die beiden Repräsentationsebenen existieren parallel und können komplementär genutzt werden. Während die metrische Karte zur *Low-Level*-Navigation genutzt wird, ermöglicht die konzeptuelle Ebene effektive Aktionsplanung und Mensch-Roboter-Kommunikation. Einen ähnlichen Ansatz zur Kartierung und Konzeptualisierung von Innenräumen auf verschiedenen Abstraktionsebenen, findet sich basierend auf der *Spatial Semantic Hier-*

archie (Kuipers et al., 2000) bei Zender et al. (2008). Mit dem Fokus auf die natürlichsprachliche Instruktion von Robotern am Beispiel von Routeninstruktionen beschäftigt sich die Arbeiten zur korpusbasierten Robotik von Bugmann et al. (2004). Im Rahmen dieser Arbeiten werden natürlichsprachliche Routeninstruktionen analysiert und ein Verfahren zur Abbildung sprachlicher Konstrukte auf von einem Roboter ausführbare Prozeduren untersucht. Das System wird in einer Miniaturstadt evaluiert, durch die ein Roboter – natürlichsprachlich instruiert – navigieren soll. Einen vergleichbaren Versuchsaufbau findet sich bei Elmogy et al. (2008): mittels einer formalen Sprache zur Roboterinstruktion wird ein Humanoid instruiert, einer Route durch eine Modellstadt zu folgen. Das Verankerungsproblem ist aktuell in den Fokus dieses Projektes geraten und wurde als grundlegend für die Mensch-Roboter-Interaktion im Allgemeinen identifiziert (Elmogy et al., 2009). Ein ambitioniertes Projekt, das sich mit der instruierten Navigation autonomer Roboter in realen Außenumgebungen beschäftigt, ist in (Lidoris et al., 2009) beschrieben. Ein Roboter findet sich in einer ihm unbekannten Umgebung wieder und fragt Passanten nach dem Weg. Derzeit sind die Instruktionen noch sehr einfache Zeige-Gesten und das Verankerungsproblem stellt sich deswegen in diesem Stadium noch nicht in dem Maße, wie es in dieser Arbeit diskutiert wird.

1.2 Zielsetzung

Der Geometrische Agent steht vor dem Problem, einer vom Menschen beschriebenen Route entlang definit referenzierter Entscheidungspunkte zu folgen. Ziel dieser Arbeit ist es, den Bezug zum *Symbol Grounding Problem* herzustellen und auf Basis aktueller Literatur zum Thema „perzeptuelle Verankerung“ einen Lösungsansatz zur perzeptuellen Verankerung von Landmarken zu entwickeln, der sich in die Architektur des Geometrischen Agenten einfügt. Als Folge der Integration perzeptueller Eindrücke wird der Agent in der Lage sein, während der instruierten Navigation – über den unmittelbaren perzeptuellen Fokus hinaus – gedächtnisbasierte Schlüsse über seine Umwelt zu ziehen.

1.3 Gliederung

Die Diplomarbeit gliedert sich wie folgt: Kapitel 2 beschreibt den Geometrischen Agenten und die Simulationsumgebung, in die er eingebettet ist. Aus der Analyse der Fähigkeiten des Agenten und der Herausforderungen, denen der Agent während

der instruierten Navigation gegenübersteht, wird die Notwendigkeit für ein Modul zur perzeptuellen Verankerung abgeleitet. Das Verankerungsproblem wird in Kapitel 3 erläutert. Dabei wird die wissenschaftliche Debatte um das Verankerungsproblem seit den Anfängen der Künstlichen Intelligenz skizziert und die Bedeutung für die Mensch-Roboter-Interaktion hevorgehoben. Relevante Aspekte des Rahmenwerks nach Coradeschi & Saffiotti (2000) werden näher erläutert. Eine Lösung des Problems der perzeptuellen Verankerung ausgedehnter Landmarken im Geometrischen Agenten wird in Kapitel 4 präsentiert. Nach einer formalen Behandlung des Lösungsansatzes wird gezeigt, welche Rolle die perzeptuelle Verankerung für die Erkennung, Lokalisierung und Wiedererkennung ausgedehnter Landmarken spielt, und wie sie als Grundlage zur lokalen Aktionsplanung und informierten Koreferenzauflösung dienen kann. Schließlich wird der Ansatz in Kapitel 5 anhand von Simulationsexperimenten und in einem Realweltszenario experimentell validiert. Kapitel 6 zieht ein abschließendes Fazit und zeigt Perspektiven für zukünftige Arbeiten auf.

2 Der Geometrische Agent

Der Geometrische Agent ist ein Rahmenwerk zur Untersuchung kognitiver Grundlagen der instruierten Navigation (Tschander et al., 2003). In der Simulationsumgebung sollen unterschiedliche Instruktionsformen in einer kontrollierten Welt und unter kontrollierten perzeptuellen Fähigkeiten eines Agenten evaluiert werden können.[1]

Der Ablauf im Szenario des Geometrischen Agenten gliedert sich in die zwei Phasen „Instruktion“ und „Navigation“ (vgl. Abbildung 2.1). Während der Instruktionsphase gibt ein Sprecher dem Agenten eine Routeninstruktion. Dabei sind sowohl der Hörer (der Geometrische Agent) als auch der Sprecher von der beschriebenen Umgebung räumlich getrennt – das heißt, die in der Instruktion beschriebenen Konstellationen werden während der Instruktionsphase von keinem der Teilnehmer wahrgenommen. Das ist deshalb entscheidend, weil anderenfalls durch andere Modalitäten als Sprache oder Skizze (beispielsweise Gestik) bereits während der Instruktion auf die Referenten verwiesen und Einigung über die Koreferenz getroffen werden könnte. Der Sprecher beschreibt dem Hörer die Route mit Bezug auf Entscheidungspunkte und Aktionen, die an den Entscheidungspunkten vom Hörer ausgeführt werden sollen (beispielsweise „Wenn Du bei Haus F stehst, dann gehst Du links.“). Der Hörer baut aus den Beschreibungen des Sprechers eine interne Repräsentation der Route auf. Im Geometrischen Agenten wird die *Conceptual Route Instruction Language* verwendet (im weiteren Verlauf als CRIL abgekürzt). Die Prozesse der syntaktischen und semantischen Sprachverarbeitung und der Interpretation der Routeninstruktion generieren eine CRIL-Repräsentation der Route (wie sie vom Hörer verstanden wurde). Diese Repräsentation wird im Folgenden wegen ihrer graphenartigen Struktur als Instruktionsgraph (kurz: I-Graph) bezeichnet. Diese Repräsentation umfasst die in der Instruktion beschriebenen Informationen über die räumlichen Konstellationen entlang der Route und einen Aktionsplan, in dem die Abfolge von Aktionen spezifiziert ist.

Die Aktionsprimitiva sind:

[1] Das heißt, dass von sensori-motorischen Aspekten der Navigation weitestgehend abstrahiert wird. Abschnitt 5.2 wird zeigen, dass die Modelle, die im Rahmen der Simulationsumgebung implementiert sind, grundsätzlich auch in Realweltanwendungen funktionieren.

!VIEW Der Agent hält nach einer Landmarke Ausschau.
Beispiel: „Dann siehst Du Haus F.“

!GO Der Agent geht entlang eines Pfades. Pfade sind Trajektorien durch den Raum, die – im Falle des Geometrischen Agenten – auf Wegen verlaufen.
Beispiel: „Du gehst zwischen Haus B und Haus C durch“.

!CH_ORIENT Der Agent ändert seine Orientierung in eine Richtung.
Beispiel: „Dann drehst Du Dich nach links.“

!BE_AT Der Agent versichert sich, in einer spezifizierten Region zu sein.
Beispiel: „Dann stehst Du vor der Mensa.“

Im Allgemeinen ist das interne Modell aus der Kommunikation unterspezifiziert und vage. Der Sprecher geht von einem Hörer aus, der menschliche Intelligenz besitzt und so Lücken, Ungenauigkeit oder Fehler in der Instruktion gegebenenfalls während der Navigationsphase selbstständig auflösen kann.

Während der Navigationsphase muss der Geometrische Agent das interne Modell aus der Instruktion mit einem internen Modell seiner tatsächlichen Wahrnehmung in Zusammenhang bringen. Er stellt Koreferenzen zwischen Elementen des aus der Instruktion stammenden internen Modells mit den Elementen des aus der Perzeption stammenden internen Modells her und identifiziert so die in der Instruktion beschriebenen Entscheidungspunkte, um die mit ihnen assoziierten Aktionen auszuführen.

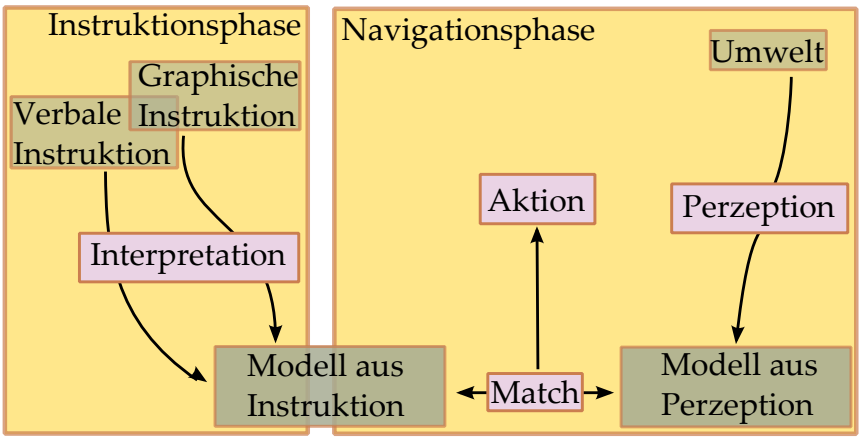


Abbildung 2.1: Instruktionsphase und Navigationsphase im Geometrischen Agenten

2.1 Umgebung und Perzeption

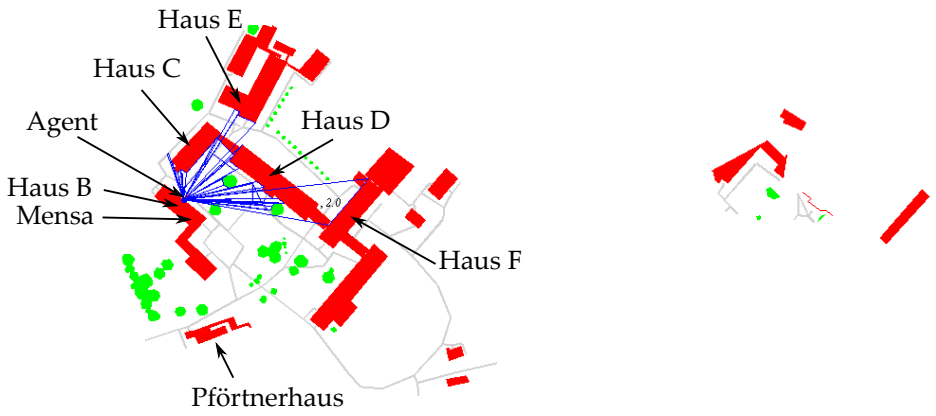
Die virtuelle Umgebung des Geometrischen Agenten ist mittels zweidimensionaler euklidischer Geometrie beschrieben. Die Landmarken in der virtuellen Welt haben eine Form und eine Ausrichtung, die durch Strecken (Wege) oder Polygone (Häuser und Bäume) repräsentiert sind. Daneben werden Regionen als Polygone (Bei-, Zwischen-Regionen) oder Halbebenen (Vor-, Hinter-, Links-, Rechts-Regionen) repräsentiert. Die Pose $< x, y, \theta >$ des Geometrischen Agenten selbst ist als Punkt (für die x- und y-Koordinate seiner Position in der Welt) und einer Richtung (für den Gierwinkel θ, der die Orientierung des Agenten angibt) modelliert. Damit sind die Freiheitsgrade des Agenten definiert. Der Agent nimmt die Perspektive eines Fußgängers ein. Insbesondere kann er nicht fliegen und besitzt somit keine Möglichkeit zur Luftbildaufnahme, was eine wichtige Einschränkung darstellt: Der Agent hat zu keinem Zeitpunkt Landmarken wie Häuser und Bäume in ihrer vollen Ausdehnung im Blick.

Landmarken sind als Polygone in der Welt definiert. Über die geometrischen Eigenschaften hinaus besitzen die Landmarken zudem nicht-geometrische Eigenschaft wie die Zugehörigkeit zu einer Kategorie (HAUS), eine Höhenangabe (HOCH, MEDIUM, NIEDRIG) und ein Label (NAME('F'))[2]. Durch die Höhenangabe der Häuser und Bäume ist die Welt eine 2,5D-Umgebung. Die Angabe spielt insbesondere bei der Berechnung der Sichtbarkeit von Landmarken eine Rolle: Landmarken, die von höheren Landmarken verdeckt werden, können vom Agenten nicht wahrgenommen werden und Landmarken, die von niedrigeren Landmarken (partiell) verdeckt werden, können gesehen werden.

Die Umgebung ist aus der Vogelperspektive in Abbildung 2.2(a) dargestellt. Die namentlich markierten Landmarken sind die Protagonisten der Beispiele, anhand derer im Verlauf der Arbeit verschiedenste Vorgänge exemplarisch demonstriert werden. Die Vogelperspektive ist allerdings – wie bereits betont – nicht die Perspektive des Agenten. Der Agent ist als blauer Punkt in der Übersichtskarte markiert. Sein Sichtfeld (durch blaue Linien angedeutet) bestimmt, welche Teile der Umgebung er wahrnimmt (siehe das sogenannten „Netzhautbild“ in Abbildung 2.2(b)).

Die Perzeption des Geometrischen Agenten ist eine Abbildung der virtuellen Welt in

[2] Unter einer *Unique-Name-Assumption* kann das Label als Identifikationsnummer einer Landmarke fungieren, was die Wiedererkennung einer Landmarke (zu) stark vereinfacht. Das in dieser Arbeit konstruierte Modell zur perzeptuellen Verankerung funktioniert auch dann, wenn der Agent die Label-Information nicht wahrnimmt.



(a) Die Welt (b) „Netzhautbild“

Abbildung 2.2: Die Umgebung des Geometrischen Agenten: der Informatik-Campus

ein internes Modell, das jeweils die aktuell wahrgenommene Szenerie repräsentiert. Die Repräsentation der Perzeption wird analog zur Repräsentation der Instruktion als P-Graph bezeichnet.

Das Sichtfeld des Agenten hat einen festen Öffnungswinkel von 120 Grad. Alle Kanten der geometrischen Elemente in der virtuellen Welt, die im Sichtfeld liegen (und von keinen anderen Elementen verdeckt werden), werden in einem internen Modell der Perzeption rekonstruiert. Die internen Repräsentationen der Wahrnehmungen sind die Perzeptionsknoten. Sie enthalten neben propositionaler Information über die gesehene Landmarke auch analoge, geometrische Repräsentationen der gesehen Kante als Linienzug (der multimodale Charakter des Perzeptionsknoten wird bei der perzeptuellen Verankerung in Kapitel 4 eine zentrale Rolle spielen).

2.2 Koreferenzauflösung

Während der Navigation gleicht der Geometrische Agent die Repräsentationen aus der Routeninstruktion (I-Graph) mit der Wahrnehmung (P-Graph) ab. Die Koreferenzauflösung analysiert die Landmarken und Regionen in der aktuell perzipierten Szene, um die beschriebenen räumlichen Konstellationen aus der Routeninstruktion in der Welt zu identifizieren. Wird eine hinreichend gute Übereinstimmung zwischen dem I-Graphen und dem aktuellen P-Graphen festgestellt, dann wird das interne Modell der Umgebung durch die Information aus dem P-Graphen ergänzt. Auf diese Art schließt die Koreferenzauflösung die Lücke zwischen der sprachlichen Beschreibung der Welt

und der Wahrnehmung der Welt.

Ein Beispiel ist in Abbildung 2.3 dargestellt. Der lila dargestellte Teil des Graphen ist das Modell aus der Routeninstruktion. Der Graph beschreibt eine Region $r5$, die die Vor-Region einer Landmarke $b1$ ist, die durch das Merkmalsbündel[3] {haus, hoch, label(„F“)} beschrieben ist. So könnte die Anweisung aus der Routeninstruktion gelautet haben:„Und dann stehst Du vor Haus F“. Die räumliche Konstellation aus der Äußerung wird als Instruktionsgraph repräsentiert und die Aktion „in der Vor-Region von Haus F sein“ – !BE_AT(r5) – dem Aktionsplan hinzugefügt. In der dargestellten Situation sieht der Agent eine Landmarke, die er als hohes Haus namens „F“ erkennt und bildet die Vor-Region auf Grundlage der gesehenen Kontur. Der Agent glaubt, dass es sich bei dem gesehenen Haus (und der gebildeten Region) um das in der Routeninstruktion erwähnte Haus (und um die beschriebene Region) handelt. Er stellt also die Koreferenz zwischen den beiden Repräsentationen fest und erweitert das Modell der Instruktion um die Information aus der Perzeption (z. B. weiß er nun, dass das Haus F hoch ist).

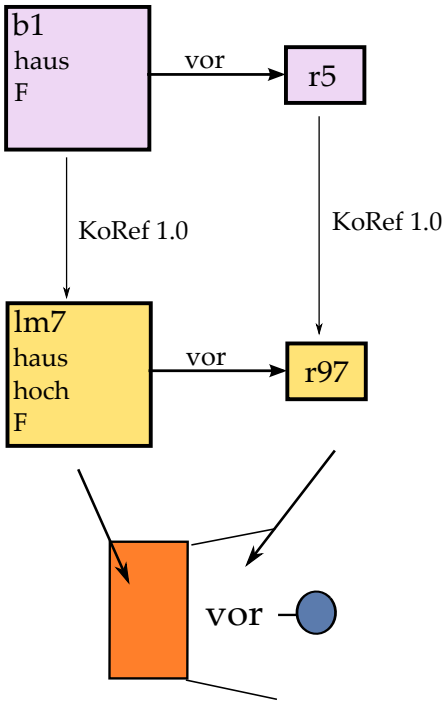


Abbildung 2.3: Koreferenzauflösung zwischen I-Graph (lila) und P-Graph (gelb)

[3]Ein Merkmalsbündel ist eine Menge von Symbolen, die Eigenschaften repräsentieren (vgl. Lakoff, 1990, S. 115).

Eine Menge von Koreferenzrelationen zwischen Instruktionsknoten und Perzeptionsknoten wird im Folgenden als „Match“ bezeichnet.

Definition 1 (Match) *Ein Match ist eine Menge bewerteter Koreferenzzuweisungen zwischen Perzpetionsknoten und Instruktionsknoten.*

Die Entscheidung für einen Match wird immer unter Unsicherheit getroffen. Im Allgemeinen gibt es mehrere Kandidaten, wie das Beispiel in Abbildung 2.4 verdeutlicht. Gesucht ist hier ein mittelhohes Haus. Der Sprecher der Routeninstruktion geht davon aus, dass seine Beschreibung die gemeinte Landmarke von den umliegenden Landmarken diskriminiert. Während der Navigationsphase kann der Agent allerdings vor dem Problem stehen, dass es mehrere Landmarken gibt, auf die die Beschreibung aus der Routeninstruktion zutrifft. Der Agent sieht in dem Beispiel drei Häuser: ein mittelhohes Haus B, ein mittelhohes Haus A und ein hohes Haus F. Das Verfahren zur Koreferenzauflösung bewertet die drei Möglichkeiten. Die Merkmale im Perzeptionsknoten $lm8$ passen weniger gut zur Beschreibung im Instruktionsknoten $b1$ als die Merkmale in den Perzeptionsknoten $lm7$ und $lm9$. Für die Wahl zwischen $lm7$ und $lm9$ gibt es allerdings zunächst kein Kriterium. Es kann keine totale Ordnung über der Menge der Match-Kandidaten aufgestellt werden.

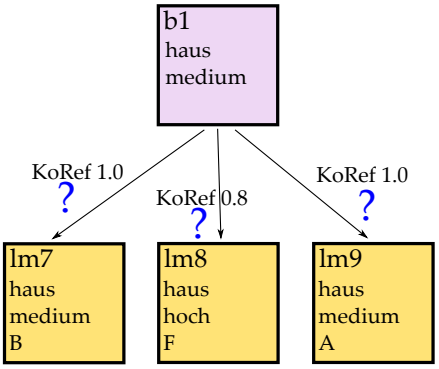


Abbildung 2.4: Entscheidung für einen Match-Kandidaten

Das Verfahren zur Berechnung von Match-Kandidaten zwischen I-Graphen und P-Graphen wird in (Helwich, 2003) entwickelt. Es berücksichtigt die Relationen zwischen Knoten, die Merkmalsbündel der Knoten und taxonomisches Weltwissen über die Konzepte, die durch die Symbole in den Merkmalsbündeln denotiert sind. In der Taxonomie sind Subsumptions- und Exklusivitätsbeziehungen zwischen Konzepten spezifiziert (Abbildung 2.5). Die Ähnlichkeit zwischen CRIL-Graphen wird in dem Verfahren auf die Ähnlichkeit zwischen CRIL-Knoten zurückgeführt und die Ähn-

lichkeit zwischen CRIL-Knoten auf die Ähnlichkeit von Konzepten. Dabei ist i) die Ähnlichkeit von Konzepten umso größer, je größer die Tiefe des spezifischsten Vaters ist, ii) die Ähnlichkeit von Konzepten umso kleiner, je größer die Differenz der Tiefe zwischen ihnen und dem spezifischsten Vater ist und iii) die Ähnlichkeit von Vater und Sohn größer als die Ähnlichkeit von Brüdern.

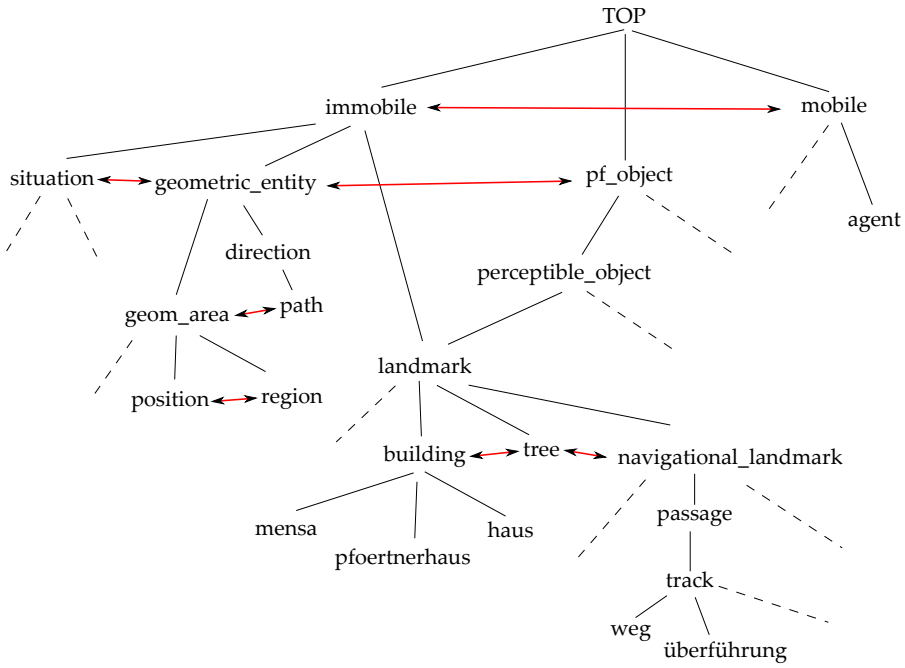


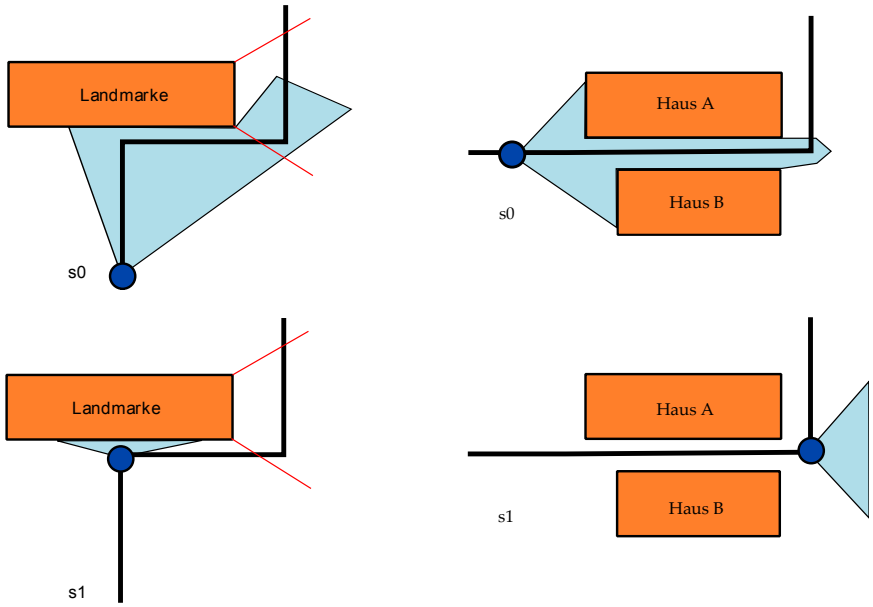
Abbildung 2.5: GA Taxonomie

2.3 Problemanalyse

Der Geometrische Agent zieht für die Koreferenzanalyse in jeder Situation immer nur die sich derzeitig bietende Szenerie in Betracht. Zwar werden diejenigen Perzeptionsknoten persistent in ein Routenmodell integriert, die mit einem Instruktionsknoten als koreferent identifiziert werden – diese Information bleibt aber für nachfolgende Entscheidungen des Agenten gegenwärtig unbeachtet. Der Teil der wahrgenommenen Szene, der mit keinem Instruktionsknoten koreferenziert werden kann, wird zudem komplett „vergessen“.

Aus der Gedächtnislosigkeit ergeben sich vielerlei Probleme: Landmarken, die aus dem Sichtfeld des Agenten geraten, können im weiteren Verlauf nicht zur Entscheidungsfindung berücksichtigt werden. Abbildung 2.6 verdeutlicht zwei Situationen, in

denen sich das fehlende Gedächtnis negativ auswirkt. In der Situation $s0$ im Bildabschnitt 2.6(a) nimmt der Agent einen Großteil der Vorderseite der Landmarke wahr. Er ist in dieser Situation in der Lage, die Rechts-Region der Landmarke zu bilden und einen Pfad in diese Region zu gehen. In Situation $s1$ steht der Agent direkt vor der Landmarke, was sein Sichtfeld deutlich einschränkt. Er sieht nur noch einen – im Vergleich zu der Wahrnehmung in Situation $s0$ – kleinen Teil der Landmarke und findet keinen Weg in die Rechts-Region. Eine vergleichbare Auswirkung der Gedächtnislosigkeit ist in Abbildung 2.6(b) dargestellt. In diesem Beispiel identifiziert der Agent zur Situation $s0$ zwei Landmarken Haus A und Haus B, um zwischen ihnen durchzugehen. Nach der Aktion – in der folgenden Situation $s1$ – findet sich der Agent zunächst in einer landmarkenlosen Welt wieder. Er kann insbesondere nicht schließen, dass er sich jetzt in der Bei-Region von Haus A befindet. Dem Geometrischen Agenten fehlt die Fähigkeit, auf perzeptuell nicht-präsente Dinge in der Welt Bezug nehmen zu können. Höhere navigatorische Fähigkeiten bedingen es, auch Landmarken, die aus dem Sichtfeld geraten, in die Entscheidungsfindung und Planung einzubeziehen. Dafür muss sich der Agent an Landmarken, Landmarkenkonstellationen und Orte erinnern.



(a) Rechtsregion gerät aus dem Fokus (b) Landmarke gerät aus dem Fokus

Abbildung 2.6: Probleme, die sich aus der Gedächtnislosigkeit ergeben

Auch besteht die Möglichkeit, dass der Agent in einer Folgesituation Haus A für Haus

B hält, wie in 2.7 geschehen. Während der Navigationsphase wurde zunächst der Perzeptionsknoten *lm*12 mit dem Instruktionsknoten *b*2 koreferenziert. Zu einem späteren Zeitpunkt wurde zudem *lm*21 mit *b*2 koreferenziert. Während es sich bei *lm*12 tatsächlich um ein Perzeptionsereignis handelt, das mit der Pförtnerei assoziiert ist, hat *lm*21 seinen Ursprung in der Betrachtung eines ganz anderen Hauses (namens „Haus F“). Später wird der Perzeptionsknoten *lm*169, der Haus F entspringt, referentiell korrekt mit *b*2 koreferenziert. In dieser Konfiguration hält der Geometrische Agent die Perzeptionsknoten *lm*12 und *lm*21 für koreferent und den Perzeptionsknoten *lm*169 für nicht-koreferent mit *lm*12 oder *lm*21. Die Wiedererkennung von Landmarken schlägt fehl. Wie bereits anhand von Abbildung 2.4 behandelt, muss der Agent eine Entscheidung bei der Auswahl möglicher Match-Kandidaten fällen. Das bedeutet auch, dass er eine Festlegung trifft, welche Landmarke er für die Landmarke hält, die durch den Instruktionsknoten denotiert ist. Weil er Landmarken nicht sicher wiedererkennen kann, werden mitunter zu verschiedenen Zeitpunkten Match-Kandidaten ausgewählt, die nicht-koreferente Perzeptionsknoten an einen gemeinsamen Instruktionsknoten binden.

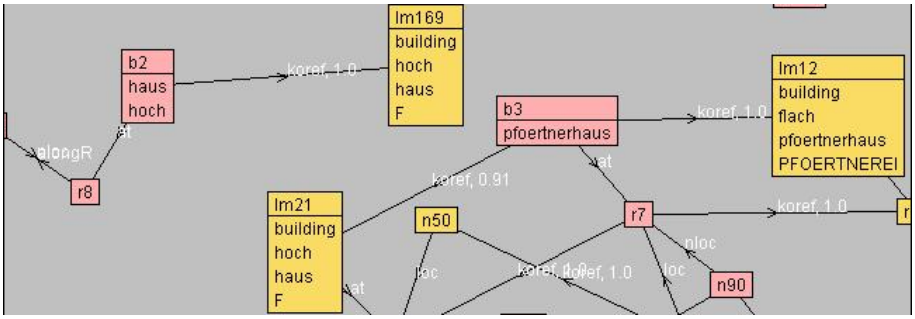


Abbildung 2.7: Fehlgeschlagene Wiedererkennung

Häufig entstehen verschiedene Perzeptionsereignisse zu einer Landmarke auch durch Verdeckungen von Landmarken (Abbildung 2.8). In solchen Fällen ist es wünschenswert, dass der Agent die Zusammengehörigkeit der Perzeptionsknoten ermittelt und einen kohärenten perzeptuellen Eindruck der betrachteten Landmarke erhält.

Die Analyse führt zu der Schlussfolgerung, dass der Geometrische Agent Wissen über die zurückgelegte Route und Übersichtswissen über die Umgebung erwerben und nutzen sollte. Er sollte insbesondere in der Lage sein, Landmarken wiederzuerkennen, um Match-Kandidaten auszuschließen, die konfligierende Koreferenzierungen einführen. Das wird durch die Ausdehnung der Landmarken, der Möglichkeit von Verdeckungen und durch die Perspektive des Agenten erschwert. Es ist dem Agenten

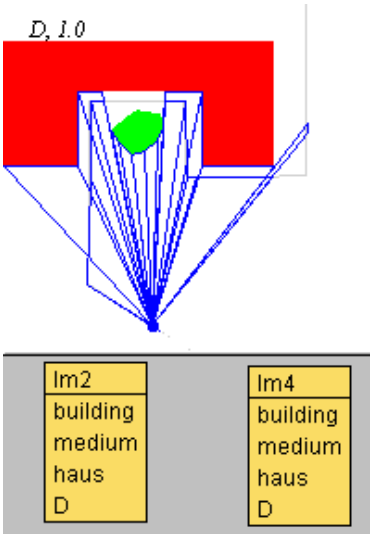


Abbildung 2.8: Fehlende Integration

nicht möglich, von irgendeiner Pose aus eine Landmarke vollständig wahrzunehmen. Darum sollte der Agent in der Lage sein, verschiedene Eindrücke einer Landmarke zu einer kohärenten Repräsentation zusammenzuführen. Dafür wird ein Entscheidungsverfahren entwickelt, das die Zugehörigkeit eines wahrgenommen Landmarkenteils zu einem bestehenden internen Modell einer Landmarke ermittelt. Die Koreferenz zwischen Modellen aus Perzeption und Instruktion werden dann zwischen diesen aggregierten Landmarkenmodellen und dem Instruktionsknoten hergestellt.

Im Folgenden wird untersucht, wie diese Probleme mit dem *Symbol Grounding Problem* im Allgemeinen und dem Problem der perzeptuellen Verankerung im Speziellen zusammenhängen. Im weiteren Verlauf der Arbeit wird eine Lösung der Probleme entwickelt und evaluiert.

3 Das Verankerungsproblem

Seit den Anfängen der Künstlichen Intelligenz (KI) Mitte des 20. Jahrhunderts werden Debatten darüber geführt, unter welchen Bedingungen einer Maschine Intelligenz zugeschrieben werden kann und welche Paradigmen am besten geeignet sind, um intelligentes Verhalten zu modellieren. Die grundlegende Annahme der Kognitionswissenschaft ist, dass Kognition auf Berechnung zurückgeführt werden kann und somit algorithmisch beschreibbar ist (*Cognition as Computation*).

Im Zentrum der klassischen KI steht die von Newell & Simon (1975) formulierte *Physical Symbol System Hypothesis* (PSSH). Ein physisches Symbolsystem besteht aus einer Menge von Symbolen, die zu symbolischen Ausdrücken zusammengesetzt werden können und aus einer Menge von Prozessen, die diese Ausdrücke manipulieren. Die Hypothese lautet, dass ein solches Symbolsystem die notwendigen und hinreichenden Bedingungen für Intelligenz erfüllt. Sie impliziert, dass jede Maschine, die ein Symbolsystem implementiert (z. B. der Computer) prinzipiell zur Intelligenz fähig ist und dass jede intelligente Maschine (z. B. der Mensch) ein physisches Symbolsystem implementiert (Newell & Simon, 1975). Die zentrale Bedeutung dieser Hypothese ist, dass Kognition nicht mehr bloß eine Idee ist, sondern realisiert bzw. implementiert werden kann (deswegen *physical*). Die Art der Implementierung eines PSS – ob als neuronales Netzwerk oder formallogische Inferenzmaschine – wird bewusst offen gelassen. In der Kognitionswissenschaft wird allgemein die Meinung vertreten, dass menschliches Verhalten – inklusive Wahrnehmen, Denken, Entscheiden, Agieren – nur durch die Art der menschlichen Konzeptualisierung der Welt (also ihrer Repräsentation anhand von Konzepten, Eigenschaften und Beziehungen) erklärt werden kann (Fodor, 1980). Es stellt sich jedoch heraus, dass bei der Bildung symbolischer Repräsentationen als Prozess der Wahrnehmung und des Verstehens die Verbindung zwischen der Beschreibung und dem Beschriebenen verloren geht. Es gibt aber einen entscheidenen Unterschied zwischen der Repräsentation der Gegebenheit, dass ein Ding existiert, das bestimmte Eigenschaften besitzt und der Repräsentation der Gegebenheit, dass ein bestimmtes Ding in der Welt bestimmte Eigenschaften besitzt. Für Letzteres wird eine Repräsentation benötigt, die die Dinge in der Welt mit ihrer Konzeptualisierung

innerhalb eines situierten Systems verbindet (vgl. Pylyshyn, 2009).

In der Formulierung der PSSH von Newell & Simon (1975, S. 116) heißt es:

> An expression designates an object if, given the expression, the system can either affect the object itself or behave in ways dependent on the object.

Im Fall klassischer wissensbasierter Systeme wird die Verbindung zwischen symbolischen Ausdrücken und ihren Referenten durch die Interpretation der Menschen hergestellt. Es bleibt aber unklar, wie die Verbindung zwischen den symbolischen Ausdrücken und ihren Referenten in der physikalischen Welt hergestellt wird. Ein Expertensystem kann zweifelsohne beispielsweise über Häuser {Haus-1, Haus-2, ...} und Wege {Weg-1, Weg-2, ...} räsonieren, ohne jemals ein Haus oder einen Weg wahrgenommen zu haben. Ein Roboter, der auf Wegen gehen und sich anhand von Häusern als Landmarken orientieren soll, steht allerdings genau vor diesem Problem, Symbole mit ihren Referenten in der physikalischen Welt in Verbindung zu bringen. Dafür muss der Designer von Robotersystemen verstehen, wie ein Roboter die Welt im Sinne einer Spezifikation einer Konzeptualisierung wahrnehmen kann. Die Lösung dieses Problems ist eine notwendige Voraussetzung für die vollständige Integration von Perzeption, Aktion und wissensbasiertem Schließen (Hertzberg & Chatila, 2008).

Die generelle Herausforderung besteht also darin zu untersuchen, wie symbolische Ausdrücke eines PSS in der Wahrnehmung eines Agenten verankert[1] werden können. Das Problem, die Bedeutung der Symbole in einem Symbolsystem anders zu definieren als nur wiederum durch weitere Symbole (anhand ihrer Beziehung zur realen Welt), ist von Harnad (1990) unter der Bezeichnung *Symbol Grounding Problem*[2] (SGP) beschrieben worden. Harnad identifiziert drei Schritte zur Lösung des SGP: Im ersten Schritt, den er „Ikonisierung" nennt, werden zunächst die Sensordaten repräsentiert. Diese Repräsentation ist analog zur Repräsentation eines visuellen Stimulus auf der Retina. Im zweiten Schritt werden einzelne Objekte voneinander „diskriminiert". Diese Phase korrespondiert mit dem Bindungsproblem, das in den Neurowissenschaften viel Aufmerksamkeit gewonnen hat (u. a. Engel, 2006). Dabei werden distinkte Objekte und ihre invarianten Eigenschaften aus dem sensorischen Stimulus extrahiert. Der dritte Schritt heißt „Identifizierung". Identifizierung erzeugt schließlich eine sym-

[1] Das englische *„to ground"* wird mit „verankern" ins Deutsche übersetzt. Im späteren Verlauf wird auch im Rahmen der perzeptuellen Verankerung (engl. *perceptual anchoring*) der Begriff *„to anchor"* mit „verankern" übersetzt. Konzeptionell ist *„Anchoring"* eine Technik für das *„Grounding"* von Symbolen, die Objekte denotieren. Aus dem Kontext wird ersichtlich, ob das allgemeine Problem oder das speziellere Teilproblem gemeint ist.

[2] Der Begriff *Symbol Grounding* wird in seiner englischen Form als Terminus technicus verwendet.

bolische Repräsentation der erkannten Objekte und ordnet sie in ein Konzeptsystem ein. Nach Harnad (2003) ist ein Symbol verankert, wenn der Agent in der Lage ist, den Referenten des Symbols (bezeichne es ein Objekt, eine Aktion oder ein Ereignis) in der physikalischen Welt zu identifizieren. Dies entspricht einer extensionalen Semantik der Symbole, wonach die Bedeutung eines Symbols durch die Menge definiter Objekte gegeben ist, die es in der externen Welt denotiert.[3]

Die PSSH sah sich seit ihrer Formulierung vor allem in der Robotik vielfacher Kritik ausgesetzt (am prominentesten durch Brooks, 1990, 1991). Nach den ersten Versuchen des symbolischen Planens in der Robotik (z. B. mit dem berühmten Roboter Shakey, Nilsson, 1984), formierte sich Ende der 80er Jahre mit der „Neuen KI“ (*Nouvelle AI*) (Brooks, 1990) ein konkurrierendes Paradigma, das die Situiertheit eines Agenten, seine physische Einbettung in die kontinuierliche Welt betont und symbolische Weltmodelle in Robotersystemen ablehnt. Von den Vertretern der „Neuen KI“ wird die Meinung vertreten, dass ein physikalischer Körper eine notwendige Voraussetzung für Intelligenz ist (auch unter *Embodiment* bekannt). Die Vernachlässigung von Situiertheit eines Agenten und damit der Interaktion mit der physikalischen Umwelt in der klassischen KI bedeutete insbesondere, dass den klassischen Modellen der Bezug zur sensori-motorischen Erfahrungswelt fehlten.

Embodiment im Sinne der von Brooks (1990) aufgestellten *Physical Grounding Hypothesis* (PGH) postuliert, dass Intelligenz im entscheidenen Maße von der Morphologie eines Agenten, seiner Sensoren und Aktoren bestimmt wird und die sensori-motorische Koppelung der Schlüssel zu einem effektiven, autonomen, situierten System darstellt. Symbolische Repräsentationen von Weltausschnitten werden weitestgehend vermieden und die Perzeption der Welt als bestes Modell ihrer selbst aufgefasst (Brooks, 1991). Unter diesem Paradigma entstanden zahlreiche Robotersysteme auf Basis der Subsumptionsarchitektur (Brooks, 1985), in denen das Systemverhalten ausschließlich auf unmittelbar perzipierbarer Information aus der Umwelt basiert. Durch die direkte Koppelung zwischen Perzeption und Aktion eines Roboters sind die Aktionen des Roboters physisch verankert (Brooks, 1991). Tatsächlich vermeiden die unter der PGH konstruierten Systeme zunächst das *Symbol Grounding Problem*: ein Agent ohne symbolische Ebene besitzt keine Symbole, die er verankern müsste. Aber sobald der Agent Sprache verarbeiten können und höhere kognitive Fähigkeiten entwickeln soll, wird er Symbole verarbeiten müssen und das Verankerungsproblem wird sich doch stellen

[3] Searle (1980) argumentiert, dass die Tatsache, dass ein Roboter den Referenten eines Symbols identifizieren kann, nicht hinreichend für die Behauptung sei, dass das Symbol für den Roboter eine (intensionale) Bedeutung hat. Siehe (Chalmers, 1992) für eine Analyse dieses Einwandes.

(Taddeo & Floridi, 2005). Somit besiegelt die Vermeidung des Verankerungproblems gleichzeitig auch die Grenzen des Systems. Allerdings schreibt auch Brooks (1990, S. 12) über das Verhältnis zwischen der symbolischen KI und seiner Idee der „Neuen KI“:

> Thus the two approaches appear somewhat complementary. It is worth addressing the question of whether more power may be gotten by combining the two approaches.

Die folgenden zwei Abschnitte argumentieren für eine symbolische Ebene in Robotik-Architekturen und verdeutlichen damit die Relevanz des Verankerungsproblems für die Robotik. Insbesondere wird auf den für die Service-Robotik zentralen Aspekt der Mensch-Roboter-Interaktion eingegangen.

3.1 Semantisches Wissen in der Robotik

Ein schwerwiegender Nachteil rein reaktiver Ansätze besteht darin, dass das Wissen des Agenten jederzeit auf die Wahrnehmung der Umwelt beschränkt ist. Einerseits bedeutet dies, dass die Performanz des Systems vollständig von der sensorischen Ausstattung des Systems und seiner Verlässlichkeit abhängt. Ein noch viel weitreichenderer Nachteil besteht darin, dass Aspekte der Welt, die nicht im perzeptuellen Fokus stehen oder sich gar prinzipiell der direkten Wahrnehmung entziehen, von solchen Systeme nicht berücksichtigt werden können. Die Anforderungen an intelligente Agenten umfassen allerdings auch höhere kognitive Aufgaben wie die Planung zukünftiger Aktionen (Levesque & Lakemeyer, 2008) und das Schließen über nicht-beobachtete Ereignisse und Zustände. Sloman (2009) führt an, dass Eigenschaften wie Festigkeit, Undurchlässigkeit, Beweglichkeit, Elastizität, Zerbrechlichkeit, Löslichkeit, Genießbarkeit von Objekten im Allgemeinen nicht direkt perzipierbar sind, sondern erst mittels Weltwissen inferiert werden müssen. Dazu muss der Agent Modelle erlernen oder bereits besitzen, die die Eigenschaften der perzipierten Objekte beschreiben – z. B. die Effekte, die Aktionen wie Heben, Schieben, Stoßen, usw. auf die Objekte und ihre Umgebung haben. Frank et al. (2009) demonstrieren beispielsweise die Relevanz von Modellen nicht-perzipierbarer Eigenschaften von Objekten am Beispiel der Bewegungsplanung eines Roboters unter Berücksichtigung deformierbarer Objekte wie Pflanzen oder Vorhänge.

Sprache und Denken erfordern, für bestimmte Zwecke unterschiedliche Dinge un-

ter gemeinsame Konzepte zu subsumieren (Johnson-Laird, 1989). Wissensbasen und Ontologien ermöglichen die Generalisierung lokaler sensorischer Wahrnehmungsereignisse, ihre Verknüpfung mit Weltwissen und unterstützen die kognitiven Fähigkeiten des Agenten in der vorausschauenden Aktionsplanung, der Objekterkennung oder der Lokalisation (vgl. Hertzberg & Saffiotti, 2008). Eine Ontologie ist eine formale Spezifikation einer Konzeptualisierung (Gruber, 1993). Sie definiert ein gemeinsames Vokabular und beschränkt die Interpretationen der Terme derart, dass sie konsistent mit der modellierten Domäne sind (Loutfi et al., 2008). Zudem bietet eine Ontologie Inferenzmechanismen und über sogenannte *Reasoning-Services* die Möglichkeit, Anfragen an die Ontologie zu stellen. Bei den Inferenzmechanismen handelt es sich vor allem um die Klassifikation von Konzepten und Individuen. Konzeptklassifikationen berechnen Subsumptionsbeziehungen zwischen Konzepten einer gegebenen Terminologie (spezifiziert in einer T-Box). Die Klassifizierung von Individuen (oder Objekten) berechnet, ob ein gegebenes Individuum Instanz eines Konzeptes ist (Baader & Nutt, 2003).

Da Ontologien durch ihren Bezug zur realen Domäne motiviert sind, können sie einen Agenten dabei unterstützen, die Welt im Sinne der in der Ontologie spezifizierten Konzepte wahrzunehmen. Für den Einsatz von Ontologien in der Robotik sprechen vielerlei Gründe (vgl. Bateman & Farrar, 2005): Ontologien unterstützen die semantisch kohärente Interpretation auf verschiedenen Wissensrepräsentationensebenen in einer Architektur. Insbesondere kann in Ontologien kodiertes Weltwissen genutzt werden, um unvollständige Sensorinformation mit Weltwissen anzureichern. Außerdem können Ontologien zwischen intelligenten Agenten geteilt werden und so die Kommunikation durch die Benutzung einer gemeinsamen Terminologie erleichtern. Dies gilt sowohl für die Roboter-Roboter-Interaktion als auch für die Mensch-Roboter-Interaktion.

3.2 Die besondere Bedeutung für die Mensch-Roboter-Interaktion

Menschliche Kommunikationspartner schreiben sich gegenseitig intentionale Zustände wie Absichten, Meinungen, Wünsche und Ziele zu, die nicht direkt perzipiert werden können. Die einzige Basis, auf die ein menschlicher Sprecher gegenüber einem unbekannten Hörer zunächst aufbauen kann, ist die Tatsache, dass das Gegenüber ebenso ein Mensch ist und somit menschlichen *Common Sense* besitzt (Krämer, 2008). Die Fähigkeit, andere Entitäten als intentionale Wesen zu begreifen, deren Verhalten von mentalen Zuständen beeinflusst ist, wurde maßgeblich von Premack & Woodruff

(1978) unter der Bezeichnung *Theorie des Geistes*[4] beschrieben.

Der Aufbau mentaler Modelle wird auch für die Mensch-Roboter-Interaktion als eine wichtige Voraussetzung angesehen. Es existieren eine Reihe von Untersuchungen darüber, wie ein künstlicher Agent ein mentales Modell seines menschlichen Kommunikationspartners aufbauen kann (Scassellati, 2002; Wachsmuth, 2008), sowie auch Untersuchungen darüber, wie ein Roboter das mentale Modell des Menschen über die Maschine beeinflusst (Kiesler, 2005). Der wichtige Aspekt für die Argumentation einer symbolischen Ebene ist, dass Kommunikation eine gemeinsame Konzeptualisierung der Welt voraussetzt, um ein gemeinsames Verständnis eines Diskursgegenstandes zu erlangen.

Eine Grundanforderung an eine erfolgreiche Mensch-Roboter-Interaktion besteht darin, dass der Roboter die vom Menschen referenzierten Objekte, Plätze, Aktionen und Ereignisse identifizieren kann (Zender et al., 2009). Gleichzeitig muss er in der Lage sein, seinerseits auf für den Menschen verständliche Weise (d.h. unter Verwendung derselben Konzepte, mittels derer der Mensch die Welt konzeptualisiert) über die Welt zu kommunizieren. Die Welt in qualitativen Konzepten zu begreifen, versetzt den Menschen in die Lage, mit komplexen Situationen umzugehen und sie zu beschreiben. Die Vagheit räumlicher Konzepte wie *vor*, *hinter*, *bei*, *links*, *rechts*, usw. ist eine Strategie, um die Welt unter Unsicherheit und trotz ihrer Komplexität zu konzeptualisieren (Johnson-Laird, 1989)[5]. Ein Roboter nimmt seine Umwelt durch seine Sensorausstattung zunächst in einer quantitativen Weise wahr. Auf dieser Ebene ist allerdings die Mensch-Roboter-Interaktion unmöglich: der Mensch hat beispielsweise ohne weiteres keine Möglichkeit, präzise metrische Daten (Winkel, Distanzen) über seine Umgebung zu erheben (vgl. Moratz & Tenbrink, 2006).

Während der Instruktion eines Service-Roboters durch einen Menschen ist zu erwarten, dass vor allem Objekte und ihre Identität im Fokus des Diskurses stehen. Essentiell dabei ist, dass sich die Kommunikationspartner auf die referentielle Identität einigen – also einen *Common Ground* darüber erreichen, welche Objekte den Diskursgegenstand bilden (Clark & Brennan, 1991). Dazu muss der Roboter zwischen seinen internen quantitativen Repräsentationen und den sprachlichen Mitteln, die Menschen verwenden, um über den Gegenstandsbereich zu reden, vermitteln. Dabei gibt es zwei grundlegende Probleme (Moratz et al., 2002): (a) Es bestehen die bereits angesprochenen Unterschiede in der Wahrnehmung zwischen Mensch und Roboter. Die Objekter-

[4] engl. *Theory of Mind*

[5] Johnson-Laird (1989, S. 201) schreibt:„Vagueness is a Solution rather than a Problem“

kennung ist noch immer ein großes Problem in der Robotik. Zwar können bekannte Objekte (solche, mit denen das System trainiert wurde) immer besser identifiziert werden, allerdings dreht es sich bei der natürlichen Kommunikation häufig um Objekte, die noch nicht bekannt sind. Im Szenario des Geometrischen Agenten beispielsweise, der einer Routeninstruktion durch eine ihm unbekannten Umgebung folgt, muss der Agent geradezu per Definition unbekannte Objekte (Landmarken) identifizieren. Dabei ist Weltwissen essentiell, um verschiedene Perzepte unter gemeinsam Konzepte zu subsumieren (verschiedene Häuser können die unterschiedlichsten perzeptuellen Eindrücke hinterlassen, sind aber unter einem gemeinsamen Konzept HAUS zu fassen). (b) Menschen generieren selten vollständige und eindeutige referentielle Ausdrücke, um auf Objekte zu verweisen.[6] Die Verbindung zwischen Sprache und der physikalischen Welt und somit die Fähigkeit, mit Sprache auf Entitäten in der Welt zu referenzieren, bildet eine wichtige Grundlage sprachlicher Kommunikation. Eines der größten Hindernisse für eine effektive Mensch-Roboter-Kommunikation liegt in der traditionellen Trennung zwischen Sprache, Sensorik und Aktorik. Aus dieser Erkenntnis entstehen immer mehr Computermodelle und Robotersysteme, die die Welt symbolischer Repräsentationen (Sprache) mit ihren Referenten in der physikalischen Welt verbinden.

3.3 Perzeptuelle Verankerung nach Coradeschi und Saffiotti

Nachdem in den vorangegangenen Abschnitten für eine symbolische Ebene in Robotikarchitekturen argumentiert wurde, werden nun Realisierungsvorschläge analysiert. Eine symbolische Wissensbasis in einem Robotersystem wird durch die Fähigkeit erweitert, Symbole geeignet und in Echtzeit mit der Signalverarbeitung auf Perzeptions- und Motorebene zu verbinden. Dazu gehören die Identifizierung symbolisch beschriebener Objekte, die symbolische Beschreibung der Wahrnehmungen und die Wahrheitswertzuweisung propositionaler Wissensbasiseinträgen anhand sensorischer Information (vgl. Hertzberg & Saffiotti, 2008).

„Perzeptuelle Verankerung“ nach Coradeschi & Saffiotti ist das Problem, die Verbindungen zwischen einer symbolischen Wissensrepräsentationsebene und der Perzeptionsebene zu etablieren und über die Zeit aufrecht zu erhalten. Präziser heißt es in (Coradeschi et al., 2001, S. 111):

[6] Das allgemeine Problem, einen nicht-ambigen referentiellen Ausdruck zu generieren, liegt in der Komplexitätsklasse NP-HART (Dale & Reiter, 1995). Menschen konstruieren häufig nicht-eindeutige referentielle Ausdrücke.

Definition 2 (Perzeptuelle Verankerung) *Anchoring is the process of creating and maintaining the correspondence between symbols and percepts that refer to the same physical objects.*

Verankerung im Allgemeinen verbindet koreferente Daten auf unterschiedlichen Ebenen innerhalb einer informationsverarbeitenden Architektur. Das Problem wurde bereits in verschiedenen Disziplinen erkannt. Es finden sich Architekturen zur Lösung des Problems aus der Computer Vision Community (Hudelot & Thonnat, 2003; Chella et al., 2004; Heintz et al., 2007), aus dem Bereich Kognitive Robotik (Hähnel et al., 1998; Shapiro & Ismail, 2003; Roy, 2005; Brenner et al., 2007; Kuipers, 2008), der multimodalen Dokumentverwaltung (Möller et al., 2009) und Psychologie (Pylyshyn, 2000, 2009).

Gemein ist allen Ansätzen neben einer subsymbolischen und einer symbolischen Ebene eine dritte, vermittelnde Ebene. Chella et al. (2004) verwendet konzeptuelle Räume nach Gärdenfors (2004) und bei Shapiro & Ismail (2003) gibt es eine Perzeptuo-Motor-Ebene, die zwischen einer Wissensrepräsentations- und einer sensori-motorischen Ebene vermittelt. In einigen dieser Ansätze werden zeigerähnliche Datenstrukturen motiviert, die die Korrespondenzen zwischen Daten auf den unterschiedlichen Ebenen explizit machen und über die Zeit aufrecht erhalten. Bei Kuipers (2008) heißt diese Datenstruktur *Tracker*, Pylyshyn (2000) nennt sie *FiNST* und bei Coradeschi & Saffiotti (2000) heißt die Datenstruktur *Anker*.

Die perzeptuelle Verankerung nach Coradeschi & Saffiotti wird im Folgenden näher betrachtet. Diese Wahl ist darin begründet, dass dieser Ansatz als erster überhaupt die Verankerung als eigenes, domänenunabhängiges Problem begreift und das Rahmenwerk einen vergleichsweise hohen Formalisierungsgrad erreicht hat.

3.3.1 Das Modell

Die hier nachvollzogene Konstruktion des Rahmenwerks orientiert sich an der Beschreibung in (Coradeschi & Saffiotti, 2000). Eine Anpassung an den Geometrischen Agenten und die Perzeption von Objekten unter Berücksichtigung ihrer geometrischen Ausdehnung folgt in Kapitel 4.

Zeitinvariante Komponenten des Modells

Das Modell geht zunächst von einem Agenten aus, der mit einem Symbolsystem und einem Perzeptionssystem ausgestattet ist:

Definition 3 (Symbolsystem) *Das Symbolsystem Σ besteht aus einer Menge $X = \{x_1, x_2, \ldots\}$ von Individuen (Konstanten), sowie einer Menge unärer Prädikatensymbolen $P = \{p_1, p_2, \ldots\}$.*

Definition 4 (Perzeptionssystem) *Das Perzeptionssystem Ξ enthält eine Menge $\Pi = \{\pi_1, \pi_2, \ldots\}$ der Perzepte und die Menge der Attribute $\Phi = \{\phi_1, \phi_2, \ldots\}$. Ein Perzept ist eine strukturierte Sammlung von Messungen, die mit einem physikalischem Objekt assoziiert sind. Ein Attribut ϕ_i ist ein messbares Merkmal mit Werten aus der Domäne $D(\phi_i)$.*

Weser & Zhang (2007) merken an, dass die Verfügbarkeit distinkter Perzepte bereits eine sehr starke Annahme ist, von der nach dieser Definition eines Perzeptionssystems ausgegangen wird. Ein reales sensorisches System liefert zunächst einen kontinuierlichen, hoch-dimensionalen, nicht-linearen und multimodalen Datenstrom. Die Abstraktion von rohen Sensordaten auf die Ebene distinkter Perzepte ist selbst ein nicht-triviales Problem. Der Einwand ist insofern berechtigt, als dass er ein Teilproblem des *Symbol Grounding* identifiziert, dessen Lösung eine Voraussetzung für die perzeptuelle Verankerung darstellt (vgl. Harnads Phasen „Ikonisierung“ und „Diskriminierung“). Das Rahmenwerk von Coradeschi & Saffiotti gibt für diese Probleme keinen allgemeinen Lösungsvorschlag an (und erhebt diesen Anspruch auch nicht). Die Definition des Problems der perzeptuellen Verankerung setzt erst auf einer Ebene an, auf der Perzepte verfügbar sind. Es setzt Objekterkennung (die Erkennung eines individuellen Objektes als Mitglied einer Kategorie) voraus und befasst sich mit Objektidentifizierung (die Erkennung eines individuellen Objektes als genau dieses spezifische Objekt).[7] Auch ist es nicht das Ziel der perzeptuellen Verankerung, ein allgemeines Verfahren zum Verankern von Eigenschaften wie „blau“ oder genereller Konzepte wie „schwierig“ anzugeben (Coradeschi & Saffiotti, 2003). Perzeptuelle Verankerung ist damit als ein Teilproblem des *Symbol Grounding Problem* zu verstehen, das sich mit der Verwaltung der Korrespondenzen zwischen perzeptuellen und symbolischen Beschreibungen von Objekten in der physikalischen Welt befasst. Ein Objekt kann auf symbolischer Ebene mittels einer Menge von Prädikaten beschrieben werden.

[7] Objekterkennung ist eine Voraussetzung für Verankerung, löst das Problem aber nicht. Es ist eine weiterführende Herausforderung, ein Objekt als „das Haus, das durch die symbolische Beschreibung σ bezeichnet ist“ zu identifizieren, als ein Objekt als Haus zu erkennen (was als solches ein schwieriges Problem ist).

Eine symbolische Beschreibung gleicht damit einem Merkmalsbündel:

Definition 5 (Symbolische Beschreibung) *Eine Symbolische Beschreibung $\sigma \in 2^P$ ist eine Menge unärer Prädikate.*

Auf perzeptueller Ebene ist ein Objekt als eine Sammlung von Attribut-Messwert-Paaren beschrieben:

Definition 6 (Perzeptuelle Signatur) *Eine Perzeptuelle Signatur $\gamma : \Phi \rightarrow D(\Phi)$ ist eine partielle Funktion von Attributen auf Attributwerte.*

Die Übersetzung von Attributwerten auf Prädikate erfolgt über eine spezielle *Grounding Relation* $g \subseteq P \times \Phi \times D(\Phi)$ mit $D(\Phi) =_{def} \bigcup_{\phi \in \Phi} D(\phi)$. Die Relation bestimmt für welche Werte aus $D(\phi)$, die für ϕ gemessen werden, das Prädikat p wahr ist. Abbildung 3.1 verdeutlicht die Abbildung von Attributwerten auf Prädikate durch g. In ihrer Implementierung verwenden Coradeschi et al. (2001) einen auf Fuzzy-Mengen basierenden Ansatz um der Vagheit linguistischer Terme (für die die Prädikate stehen) und der Unsicherheit sensorischer Messungen gerecht zu werden. Diese Sicht auf die Abbildung von Sensormesswerte auf Prädikate ist nicht ganz unproblematisch, wie das Beispiel in Abbildung 3.1 selbst zeigt: so wird eine Menge von Messwerten, die in ein bestimmtes Intervall fallen, auf das Prädikat „groß“ abgebildet. Allerdings dürfte dieses Intervall für verschiedene Klassen von Objekten unterschiedlich sein (ein großer Becher ist ein ziemlich kleines Haus). [8]

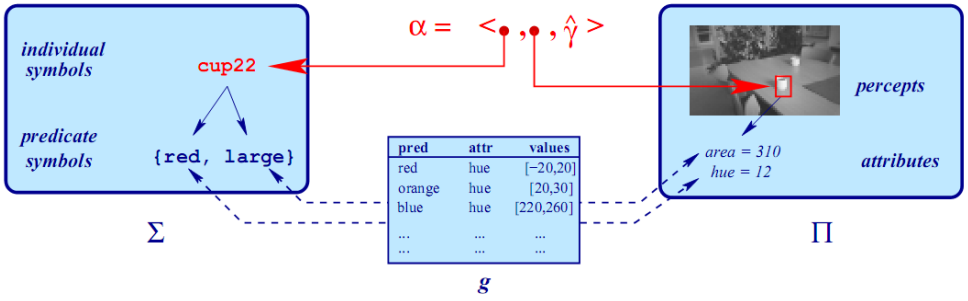


Abbildung 3.1: Perzeptuelle Verankerung (Quelle: Coradeschi & Saffiotti, 2003)

[8] Einen zu Fuzzy Mengen alternativen Ansatz zur perzeptuellen Verankerung verfolgen Chella et al. (2004) mit konzeptuellen Räumen nach Gärdenfors (2004). In (Raubal, 2004) wird am Beispiel der Wahrnehmung von Landmarken zu unterschiedlichen Tageszeiten gezeigt, wie kontextabhängige Interpretationen in konzeptuellen Räumen modelliert werden können.

Zeitvariante Komponenten des Modells

In einem perzeptuellen System werden kontinuierlich Perzepte generiert, die dem Agenten neue Information über den aktuellen Zustand der physikalischen Welt liefert. Um die Zeitvarianz der Beschreibungen eines Objektes – sowohl auf symbolischer als auch auf perzeptueller Ebene – sowie deren Korrespondez zu repräsentieren, werden dem Modell drei zeitabhängige Komponenten hinzugefügt: die *symbolische Objektbeschreibung*, die *perzeptuelle Objektbeschreibung* und der *Anker*.

Definition 7 (Symbolische Objektbeschreibung) *Eine symbolische Objektbeschreibung* $\Delta_t : X \rightarrow 2^P$ *assoziiert jedes Individuum* $x \in X$ *mit seiner symbolischen Beschreibung zum Zeitpunkt* $t \in T$.

Definition 8 (Perzeptuelle Objektbeschreibung) *Eine perzeptuelle Objektbeschreibung* $S_t : \Pi \rightarrow (\Phi \rightarrow D(\Phi))$ *assoziiert jedes Perzept* $\pi \in \Pi$ *mit seiner perzeptuellen Signatur zum Zeitpunkt t. Wenn* π *nicht im Wahrnehmungsfokus zum Zeitpunkt t ist, dann ist* $S_t(\pi)$ *für jedes* $\phi \in \Phi$ *undefiniert. Die Menge der wahrgenommen Perzepte zum Zeitpunkt* $t \in T$ *heißt* V_t.

Definition 9 (Anker) *Ein Anker ist eine partielle Funktion der Form* $\alpha : T \rightarrow X \times \Pi \times (\Phi \rightarrow D(\phi))$. *Ein Anker repräsentiert die Korrespondenz zwischen Perzepten und Symbolen über die Zeit.*

Ein Anker $\alpha(t)$ enthält zu jedem Zeitpunkt $t \in T$ ein Symbol $x \in X$, das das Objekt denotiert, ein Perzept $\pi \in \Pi$, das aus der Beobachtung des Objektes resultiert, und eine perzeptuelle Signatur γ, die die messbaren Attribute zum Perzept zusammenfasst (siehe auch Abbildung 3.1). Für die Erstellung und Verwaltung der Anker definieren Loutfi et al. (2005) drei grundlegende Operationen (vgl. Abbildung 3.3).

Acquire Der Operator Acquire generiert einen neuen Anker, wenn ein Perzept wahrgenommen wird, das keinem bereits existierenden Anker zugeordnet werden kann. Er nimmt als Parameter das Perzept π und gibt einen neuen Anker α zurück, der zum aktuellen Zeitpunkt t definiert ist. Dabei wird auf symbolischer Ebene eine neue Konstante eingeführt, die fortan das perzipierte Objekte denotiert. Dabei hat Acquire eine selektive Aufmerksamkeit auf bestimmte Objekte, die das System für derzeitig relevant hält.

Find Find sucht unter den bereits existierenden Ankern einen Anker, der auf eine symbolische Beschreibung σ passt. Wenn die symbolische Beschreibung ein Objekt beschreibt, das von Acquire nicht berücksichtigt wird, werden zusätzlich

die aktuellen Perzepte nach einem passenden Perzept durchsucht. Im Falle einer Übereinstimmung wird ein neuer Anker generiert.

Definition 10 (Übereinstimmung: partiell, vollständig) *Gegeben sei eine Beschreibung σ, ein Perzept π und eine* Grounding*-Relation g. π ist* vollständig übereinstimmend *mit σ, wenn – mit g – jedes Prädikat in σ kompatibel mit den Messwerten des korrespondierenden Attributs in der perzeptuellen Signatur von π ist. π ist* partiell übereinstimmend *mit σ, wenn σ Prädikate enthält, die in π nicht beobachtet wurden, aber jedes Prädikat aus σ, zu dem es eine korrespondierende Messung in der perzeptuellen Sigantur von π gibt, durch die Messung erfüllt wird.*

Track Track erweitert die Definition eines für $t - k$ definierten Ankers α nach t. Damit wird es möglich, dass der Anker auch dann genutzt werden kann, wenn das durch ihn denotierte Objekt nicht perzeptuell präsent ist. Der Roboter hält so ein Weltmodell über die Zeit aufrecht. Der Track-Operator wird vor allem von der Reacquire-Funktion genutzt. Zu einem früheren Zeitpunkt perzipierte Objekte können als bereits bekannte Objekte wiedererkannt werden.

3.3.2 Zielgerichtete Auflösung ambiger Situationen

Das Ziel vieler kommunikativer Akte ist, dass der Hörer die vom Sprecher gemeinten Objekte in der Welt identifiziert. Dabei wird „Identifizierung“ im Kontext perzeptueller Verankerung hauptsächlich im Sinne von Cohen (1984) verstanden: Gemeint ist eine Situation, in der die Referenten für den Hörer prinzipiell perzeptuell zugänglich sind und der Sprecher vom Hörer erwartet, das vom Sprecher gemeinte Objekt in der physikalischen Welt zu erkennen. Die Identifizierung eines Referenten setzt eine Beschreibung des Referenten voraus. Die Beschreibung kann von einem anderen Agenten produziert worden sein (im Falle von direkter Mensch-Mensch- oder Mensch-Agent/Roboter-Kommunikation), von einer externen Quelle stammen (instruierender Text, Bild, Karte) oder intern durch ein Planmodul generiert worden sein. In Datenbanken bzw. formalen Wissensbasen werden Referenten über die Schnittmengenbildung der Extensionen der Prädikate in der gegebenen Beschreibung identifiziert. Dabei wird allerdings vorausgesetzt, dass die Extensionen aufgrund der in der Wissensbasis enthaltenen Informationen berechnet werden können (Cohen, 1984). Ein Roboter hat Zugriff auf eine weitere Informationsquelle: seine Wahrnehmung der Welt. Er kann Aktionen planen, um die Extensionen von Prädikaten in der realen Welt zu bestimmen.

Dabei ist es für den Agenten von Bedeutung, welcher Art die Referenzen sind. Karlsson et al. (2008) unterscheiden zunächst definite und indefinite Referenzen. Eine Referenz ist definit, wenn sie genau ein Objekt denotiert (wie z. B. in „Dann siehst Du Haus F“). Eine indefinite Referenz denotiert eine Menge von Objekten (wie z. B. in „Bring mir eine Flasche Wasser“). Karlsson et al. (2008) identifizieren fünf Fälle und spezifizieren jeweils, wie der Roboter auf die Fälle reagieren soll (Tabelle 3.1).

	# Matches		Definite		Indefinite	
Case	Full	Partial	Result	Action	Result	Action
1	0	0	Fail	Search	Fail	Search
2	0	1+	Fail	Observe	Fail	Observe
3	1	0	OK	–	OK	–
4	1	1+	OK/Fail	–/Observe	OK	–
5	2+	Any	Conflict	–	OK	–

Tabelle 3.1: Verankerungsprobleme bei definiten und indefiniten Referenzen: Fälle und Reaktionen des Agenten (Quelle: Karlsson et al., 2008)

Fall 1 Es gibt keine Übereinstimmung zwischen der gegebenen Beschreibung und einem perzipierten Objekt. Der Agent fährt mit einem Suchverhalten fort.

Fall 2 Es gibt mehrere partielle Übereinstimmungen. Der Agent initiiert Aktionen, um mehr über die Objekte zu erfahren.

Fall 3 Es gibt genau eine vollständige Übereinstimmung. Der Agent glaubt an die Koreferenz zwischen der Beschreibung und der Perzeption des Objektes.

Fall 4 Es gibt genau eine vollständige Übereinstimmung und mehrere partielle Übereinstimmungen. Im Fall indefiniter Referenz kann sich der Agent für das vollständig übereinstimmende Objekt entscheiden. Im Fall definiter Referenz existiert die zusätzliche Option, weitere Aktionen zu initiieren, um mehr über die Objekte zu erfahren, für die partielle Übereinstimmungen gefunden wurden.

Fall 5 Es gibt mehrere Objekte, die vollständig mit der Beschreibung des gesuchten Objektes übereinstimmen. Während im indefiniten Fall der Agent einfach eines der passenden Objekte auswählen kann, liegt im definiten Fall ein Konflikt vor, der nur beispielsweise durch Rückfrage an den Sprecher mit Bitte um eine genauere Beschreibung aufgelöst werden kann.

Sprecher von Routeninstruktionen beschreiben genau eine Route entlang bestimmter Entscheidungspunkte aus der Sicht eines imaginären Wanderers (Wunderlich & Reinelt, 1982). In Routeninstruktionen sind deswegen vor allem definite Referenzen zu erwarten. Die besondere Herausforderung definiter Referenzen liegt darin, dass die Beschreibungen oft unterspezifiziert sind und somit auf mehrere Objekte passen bzw. darin, dass die perzipierbaren Attribute eines Objektes nicht ausreichen, um es anhand dieser Attribute eindeutig zu beschreiben.

Zu den in der Verankerungsliteratur noch unbehandelten Referenzierungsformen gehören Beschreibungen, die gemeinsames Wissen zwischen Sprecher und Hörer voraussetzen – wie beispielsweise in „Bring mir die Flasche Wasser, aus der ich heute morgen getrunken habe“. Auch die Fälle überspezifizierter oder fehlerhafter Referenz sind noch ausgeklammert. Ein berühmtes Beispiel für den Fall überspezifizierter Referenz ist „Die Frau, die Chamapgner trinkt“, die auch gemeint ist, wenn im Champagnerglas in Wahrheit Sekt oder Selters enthalten ist (vgl. Evans, 1982). Der Fall fehlerhafter Referenz kann insbesondere in Routeninstruktionen bei der Verwechslung von links mit rechts auftauchen oder bei falscher Erinnerung an Straßenkonstellationen. So kann beispielsweise die mit „die dritte Straße“ beschriebene Straße in Wahrheit auch die zweite Straße sein.

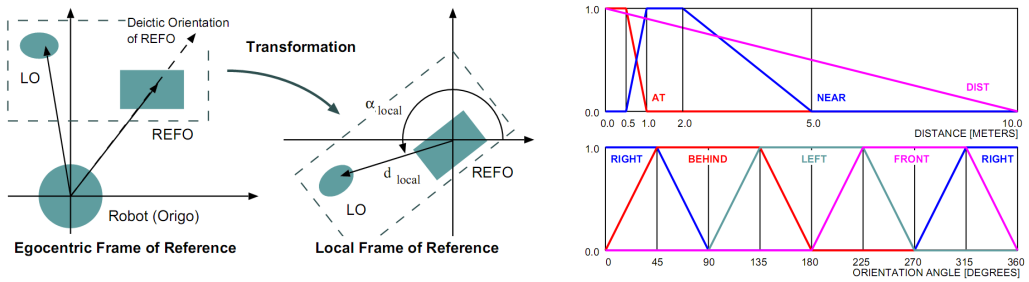
3.3.3 Behandlung räumlicher Relationen

Räumliche Relationen spielen bei der Beschreibung und bei der Identifizierung von Objekten eine bedeutende Rolle. Die Beziehung von Objekten zu anderen Objekten oder zu einem Betrachter stellt ein wichtiges Merkmal ihrer Beschreibung dar. Durch die Verwendung räumlicher Relationen können im Falle perzeptuell ununterscheidbarer Objekte[9] ambige Situationen aufgelöst werden. Beispielsweise könnte es zwei perzeptuell nicht unterscheidbare Häuser geben, die nebeneinander stehen. Diese Situation kann durch einen projektiven Term aufgelöst werden: „Dann gehst Du am *linken* Haus vorbei“. Eine weitere Möglichkeit besteht darin, ambige Situationen, vor denen der Roboter steht, mittels Klärungsdialogen zwischen Mensch und Roboter zu lösen:

Mensch: Siehst Du das hohe Haus?
Roboter: Ich sehe zwei hohe Häuser. Ist das linke oder das rechte Haus gemeint?

[9] für eine gegebene sensorische Ausstattung

Projektive Präpositionen wie *vor*, *hinter*, *links*, *rechts*, topologische Präpositionen wie *bei*, *in* und das Wort *zwischen* sind in Routeninstruktionen oft verwendete Terme. Einen Ansatz zur Berechnung räumlicher Relationen innerhalb des Rahmenwerks zur perzeptuellen Verankerung beschreiben Melchert et al. (2007). Dabei wird zunächst von einem egozentrischen Referenzrahmen ausgegangen (das heißt aus Sicht des Roboters, Abbildung 3.2(a)). Es wird also erwartet, dass der Instruierende räumliche Beziehungen zwischen Objekten relativ zur Pose des Roboters tätigt. Diese Konfiguration wurde auch in Studien mit naiven Roboterinstrukteuren (Moratz & Tenbrink, 2006) als eine für den Menschen intuitive Konfiguration identifiziert. Allerdings finden sich in sprachlichen Äußerungen im Allgemeinen auch Anweisungen bezüglich intrinsischer („Das Haus ist hinter der Kirche“ mit Ursprung „Kirche“) und absoluter („Das Haus ist nördlich der Kirche“) Referenzrahmen (Levinson, 1996), die hier unbehandelt bleiben.



(a) Egozentrischer Referenzrahmen (b) Fuzzy Akzeptanzsektoren

Abbildung 3.2: Räumliche Relationen zwischen verankerten Objekten (Quelle: Melchert et al., 2007)

Melchert et al. (2007) modellieren die topologischen Präpositionen *bei* und *nahe* sowie die projektiven Präpositionen *vor*, *hinter*, *links*, *rechts*. Dafür definieren sie zwei Funktionen a_{topo} und a_{proj}, die – gegeben ein Lokalisationsobjekt (LO) und ein Referenzobjekt (REFO) – für jede Präposition basierend auf die jeweilige Zugehörigkeitsfunktion $\mu_{topo/proj}$ (Abbildung 3.2(b)) einen Akzeptanzwert $a_{topo/proj}(LO, REFO)$ berechnen:

$$a_{topo} : (LO, REFO) \mapsto \mu_{topo}(d_{local}(LO))$$

a_{topo} berechnet sich auf Basis der euklidischen Distanz zwischen LO und REFO.

$$a_{proj} : (LO, REFO) \mapsto \mu_{dist}(d_{local}(LO)) \cdot \mu_{proj}(\alpha_{local}(LO))$$

a_{proj} zieht neben der euklidischen Distanz zwischen LO und REFO auch den Winkel α_{local} zwischen LO und REFO ein.

Die räumlichen Relationen werden als Tupel $< LO, REFO, Relation, Akzeptanzwert >$ in den Anker integriert und der FIND-Operator wird so angepasst, dass er Beschreibungen unterstützt, die räumliche Relationen enthalten (Abbildung 3.3). Nach jedem Perzeptionszyklus werden alle räumlichen Relationen zwischen allen Ankern neu berechnet.

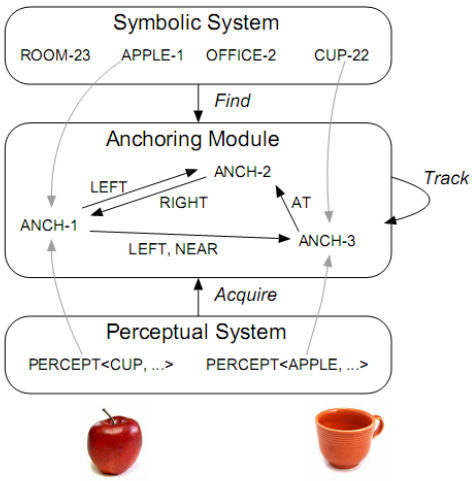


Abbildung 3.3: Räumliche Relationen zwischen Ankern (Quelle: Melchert et al., 2007)

In diesem Ansatz wird angenommen, dass die vollständigen 2D-Koordinaten-Repräsentationen der Objekte vom Roboter wahrgenommen werden, sodass aufgrund dieser Information die Zentroiden berechnen werden können (Abbildung 3.2(a)). Im Folgenden wird eine Variante der perzeptuellen Verankerung vorgeschlagen, in der diese Annahme fallengelassen wird. Das Problem, die räumlichen Relationen zwischen den Ankern während der Roboterfahrt ständig aktualisieren zu müssen, wird zudem durch eine Integration der erkannten Landmarken in eine Überblickskarte gelöst, aus der bei Bedarf die räumlichen Relationen geometrisch inferiert werden. Die Prozesse zur Identifizierung und Wiedererkennung von Objekten werden auf den Fall der Identifizierung und Wiedererkennung ausgedehnter Landmarken während der instruierten Navigation angepasst.

4 Perzeptuelle Verankerung ausgedehnter Landmarken

Während der Navigation durch eine reale Umgebung sind im Allgemeinen mehrere Wahrnehmungsereignisse mit einer Landmarke verbunden. Es können partielle Verdeckungen von Landmarken auftreten und es existieren Landmarken, die aufgrund des Ausmaßes ihrer geometrischen Ausdehnung nicht vollständig vom perzeptuellen Apparat des Agenten „auf einen Blick“ erfassbar sind. Während der Navigation im Freiland orientiert sich ein navigierender Agent an Gebäuden, Straßen, Bäumen, Flüssen und Plätzen. Solche Landmarken besitzen eine deutliche räumliche Ausdehnung.

4.1 Konzeption eines Modells

Im Folgenden wird ein Verfahren zur perzeptuellen Verankerung ausgedehnter Landmarken vorgeschlagen, das einzelne koreferente Wahrnehmungsereignisse zu einer kohärenten Repräsentation einer Landmarke zusammenfügt. Auf Grundlage des dabei entstehenden räumlichen und episodischen Gedächtnisses wird der Agent Entscheidungen bei der Koreferenzauflösung und der Aktionsplanung treffen können und die im Abschnitt 2.3 analysierten Probleme lösen.

Die Basis der Konzeption bildet der Perzeptionsknoten (vgl. Kapitel 2). Perzeptionsknoten bündeln Merkmale distinkter perzipierter Teile von Landmarken. Auf dieser Ebene ist also das Problem der Segmentierung und Diskretisierung der sensorischen Information gelöst. Des Weiteren sind quantitative Messwerte bereits in ihre qualitativen Konzeptualisierungen überführt. Das entspricht der Anwendung der *Grounding*-Relation für Prädikate bei Coradeschi & Saffiotti (2000).

Um einzelne Perzepte über die Zeit integrieren zu können, werden Perzeptionsknoten zunächst in Ereignisse gekapselt. Ereignisse sind Punkte in Raum und Zeit und bilden die Basiskomponenten des episodischen Gedächtnisses (Tulving, 1993).

Definition 11 (Perzeptionsereignis) *Ein Perzeptionsereignis ist ein Tripel bestehend aus einem Perzeptionsknoten PE^{Knoten}, einer Situation $PE^{Situation}$ und einer Pose PE^{Pose}.*

Mit dieser Definition eines Perzeptionsereignisses ist ein Perzeptionsknoten explizit in Raum und Zeit verankert. Der Agent merkt sich, wann und wo er eine Landmarke (partiell) wahrgenommen hat.

Während der Navigationsphase sind im Allgemeinen mehrere unterschiedliche Wahrnehmungsereignisse mit einer Landmarke assoziiert. Für eine kohärente Repräsentation einer Landmarke werden koreferente Perzeptionsereignisse geeignet zusammengefasst. Eine Repräsentationsmöglichkeit besteht darin, koreferente Perzeptionsereignisse über Koreferenzkanten zu einem gemeinsamen Instruktionsknoten in Beziehung zu setzen (Abbildung 4.1). Der Nachteil an dieser Lösung besteht darin, dass der Instruktionsknoten eine Voraussetzung dafür ist, dass der Agent die Landmarke memoriert und zu einem späteren Zeitpunkt wiedererkennen kann. Die Landmarkenerkennung und -wiedererkennung soll aber zunächst unabhängig von der Entscheidung sein, ob die perzipierte Landmarke in der Instruktion vorkommt.

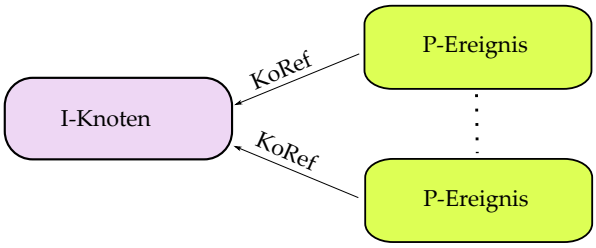


Abbildung 4.1: Koreferenzrelationen zwischen Instruktionsknoten und mehreren Perzeptionsereignissen

Wie bereits auf Seite 27 ausgeführt und in Abbildung 3.3 dargestellt, ist die Erkennung und Wiedererkennung von Objekten (ACQUIRE) ein anderer Prozess als die Identifizierung eines Objekt als „das Beschriebene“ (FIND). Aus diesem Grund wird eine weitere Ebene zwischen Perzeptionsereignis und Instruktionsknoten eingeführt: das landmarkenspezifische Perzeptionsmodell (im Folgenden auch nur „Perzeptionsmodell“ oder „P-Modell“ genannt und als solches vom „Modell aus der Perzeption“ einer Szene zu unterscheiden).

Zunächst sei das Konzept eines Perzeptions-Chunks definiert:

Definition 12 (Perzeptions-Chunk) *Die Menge PE sei die Menge aller Perzeptionsereignisse. Ein Perzeptions-Chunk PChunk* $\in \mathcal{P}(PE)$ *ist die Menge der Perzeptionsereignisse, die vom Agenten als paarweise koreferent betrachtet werden.*

Die Definition eines Perzeptions-Chunks steht in Einklang mit kognitionspsycholo-

gischen Gedächtnismodellen (Gobet et al., 2001), wonach ein Chunk einzelne Informationseinheiten über eine Entität in der Welt zusammenfasst. Ein Chunk ist eine Sammlung von Elementen, die untereinander stärker assoziiert sind als mit Elementen anderer Chunks. Der Perzeptions-Chunk bündelt also die Information zu einer Landmarke und dient als Grundlage zur Inferenz landmarkenspezifischen Wissens – insbesondere des Wissens über die Geometrie der Landmarke zu einem bestimmten Zeitpunkt. Der Perzeptions-Chunk und das daraus gefolgerte landmarkenspezifische Wissen bildet das Perzeptionsmodell.

Definition 13 (Perzeptionsmodell, Mentales Bild) *Das Perzeptionsmodell kapselt perzipiertes und inferiertes Wissen zu einer Landmarke. Es ist ein Tupel bestehend aus einem Perzeptions-Chunk und einer geometrischen Repräsentation, die im Folgenden als mentales Bild der Landmarke bezeichnet wird.*

Das mentale Bild in einem Perzeptionsmodell ergibt sich aus der Integration der geometrischen Information der einzelnen Wahrnehmungsereignisse im assoziierten Perzeptions-Chunk. Die Idee dahinter ist, dass der Agent durch direkte Interaktion mit der Umwelt über die Zeit unüberwacht eigene Repräsentationen seiner Umgebung erwirbt. Insbesondere wenn der Agent ein Roboter ist, unterscheidet sich die Wahrnehmung des Agenten beträchtlich von der Wahrnehmung eines Menschen. Somit sollte der Agent selbstständig eigene Modelle der Objekte in der Welt erlernen und diese Modelle dann während der Aktionsplanung und Aktionsausführung nutzen (Modayil & Kuipers, 2008). Des Weiteren ermöglicht das inkrementelle Verfahren dem Agenten, über seine gesamte „Lebensspanne“ aus neuen Erfahrungen zu lernen.

Eine erweiterte Repräsentation der perzeptuellen Verankerung ausgedehnter Landmarken verknüpft die einzelnen koreferenten Perzeptionsereignisse über Koreferenzbeziehungen mit einem gemeinsamen Perzeptionsmodell (Abbildung 4.2). Der Vorteil dabei besteht darin, dass der Instruktionsknoten jetzt keine notwendige Voraussetzung für die perzeptuelle Verankerung einer Landmarke mehr darstellt. Das Perzeptionsmodell ist eine symbolische, kohärente Repräsentation der Landmarke – unabhängig davon, ob ein koreferenter Instruktionsknoten existiert.

In dieser Repräsentation steckt allerdings wenig Information über den zeitlichen Verlauf der Koreferenzzuweisungen. Auf Seite der Perzeptionsereignisse, die Zeitinformation tragen, kann der Agent den zeitlichen Verlauf des Wissenserwerbs zwar nachvollziehen, allerdings kann der Agent aus dieser Repräsentation nicht schließen, wann er die Koreferenz zwischen Instruktionsknoten und Perzeptionsmodell etabliert hat.

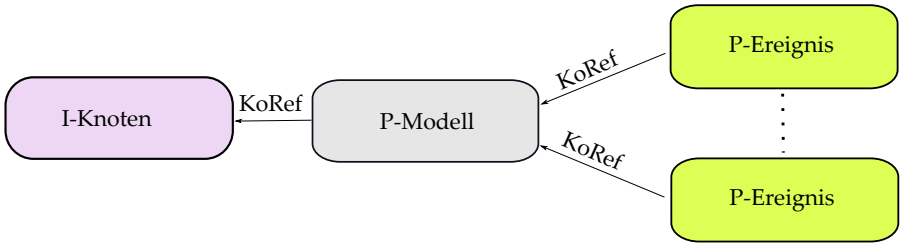


Abbildung 4.2: Repräsentation der Koreferenzbeziehungen im Anker

Auch ist noch nicht berücksichtigt, dass Koreferenzen auch revidiert werden könnten. So kann es beispielsweise für spätere Entscheidungen von Bedeutung sein, ob der Agent schon einmal an eine Koreferenz zwischen einem Instruktionsknoten I und einem Perzeptionsmodell PM geglaubt hat, um etwa zyklische Präferenzen hinsichtlich der Koreferenzbeziehungen zu vermeiden. Auch ist es denkbar, dass die Zugehörigkeit von Perzeptionsereignissen zu einem Anker revidiert werden. Dann könnte es von Nutzen für den Agenten sein zu wissen, in welcher Situation die Revision durchgeführt wurde.

Aus diesen Gründen wird eine Datenstruktur eingeführt, die die Koreferenzrelationen zwischen Perzeptionsereignissen und Perzeptionsmodellen einerseits und Instruktionsknoten und Perzeptionsmodellen andererseits über die Zeit repräsentiert:

Definition 14 (Anker einer ausgedehnten Landmarke) *Sei S die total geordnete Menge der Situationen und IK die Menge der Instuktionsknoten und PE die Menge aller Perzeptionsereignisse. Ein Anker einer ausgedehnten Landmarke ist eine Relation $Anker \subseteq S \times \mathcal{P}(PE) \times PModelle \times IK$. Der Anker repräsentiert die Korrespondenz zwischen einer Menge von Perzeptionsereignissen und einem Perzeptionsmodell und zwischen dem Perzeptionsmodell und einem Instruktionsknoten über den zeitlichen Verlauf.*

Im Anker werden Perzeptionsmodelle, die aus Wahrnehmungsereignismengen resultieren, mit korrespondierenden Modellen aus der Instruktion in Relation gesetzt, von denen der Agent glaubt, dass sie auf dieselbe physikalische Landmarke referenzieren. Es findet eine inkrementelle Aggregation koreferenter Perzeptionsereignisse statt.

Neben der bloßen Repräsentation von Koreferenzbeziehungen, bietet der Anker eine Sicht auf den zeitlichen Verlauf des Wissenserwerbs zu einer Landmarke und kann somit als Gedächtnisspur im episodischen Gedächtnis des Agenten aufgefasst werden. Während semantisches Gedächtnis Wissen über die Welt in Form von Konzepten, Individuen und Relationen enthält (vgl. Abschnitt 3.1), befähigt ein episodisches Gedächt-

nis einen Agenten dazu, sich an erlebte Ereignisse zu erinnern und sich in vergangene Situationen „zurückzuversetzen“ (Tulving, 1993). Der Agent hat nicht bloß eine mentale Repräsentation der Landmarke und weiß, an welche Koreferenzbeziehungen er glaubt, sondern er hat auch Zugriff auf die Historie des Wissenserwerbs über eine Landmarke und die damit verbundenen Koreferenzentscheidungen. Die zeitliche Entwicklung eines Ankers ist in Abbildung 4.3 dargestellt. Der Anker kann als partielle Funktion über die Menge der Situationen S betrachtet werden. In jeder Situation registriert der Agent, ob die mit dem Anker assoziierte Landmarke perzipiert wurde (vgl. ACQUIRE-Operator, Kapitel 3.3). Ist dies der Fall, enthält der Anker in der aktuellen Situation ein aktualisiertes Perzeptionsmodell. Konnte in einer Situation ein Instruktionsknoten in den Anker verankert werden, wird der Anker ebenfalls aktualisiert. Fortan ist dann der Instruktionsknoten auch koreferent zu den folgenden Perzeptionsmodellen, bis die Verankerung revidiert wird. Man beachte, dass die Generierung von Ankern als *Bottom-Up*-Prozess modelliert ist: Ein neues Perzeptionsereignis, das keinem bestehenden Anker zugeordnet werden kann, erzeugt einen neuen Anker, dessen Slot für den Instruktionsknoten zunächst leer ist.

Abbildung 4.4 veranschaulicht die Beziehungen der einzelnen Komponenten zueinander. Jede Wahrnehmung einer Landmarke erzeugt einen Perzeptionsknoten, der ein Merkmalsbündel und eine geometrische Repräsentation der Kontur des gesehenen Landmarkenteils erhält. Der Perzeptionsknoten wird daraufhin mit Raum- und Zeitinformation (Pose des Agenten und Situation) angereichert und als Perzeptionsereignis über einen Anker mit einem Perzeptionsmodell verankert. Dies entspricht der Funktion des ACQUIRE-Operators (siehe Seite 27) und stellt eine *Bottom-up*-Verankerung der Perzeptionsereignisse dar. Außerdem verwaltet der Anker die Koreferenz zwischen einem Perzeptionsmodell und einem Instruktionsknoten. Die Festellung dieser Koreferenz basiert auf der Koreferenzauflösung, wie sie in Abschnitt 2.2 beschrieben wurde. Dieser Vorgang entspricht dem FIND-Operator aus der Konzeption von Coradeschi & Saffiotti und entspricht einer *Top-Down*-Verankerung einer symbolischen Beschreibungen.

Der Ablauf der perzeptuellen Verankerung ist in Abbildung 4.5 dargestellt: 4.5(a) Der Sprecher der Routeninstruktion drückt einen Sachverhalt bezüglich einer bestimmten Landmarke aus. Der Hörer interpretiert diese Äußerung und fügt seine Interpretation einem internen Routenmodell hinzu. Dieses interne Routenmodell ist eine symbolische Beschreibung der gesuchten Landmarkenkonstellationen, relativ zu denen Aktionen definiert sind. 4.5(b) In der Navigationsphase perzipiert der Agent Landmarken in

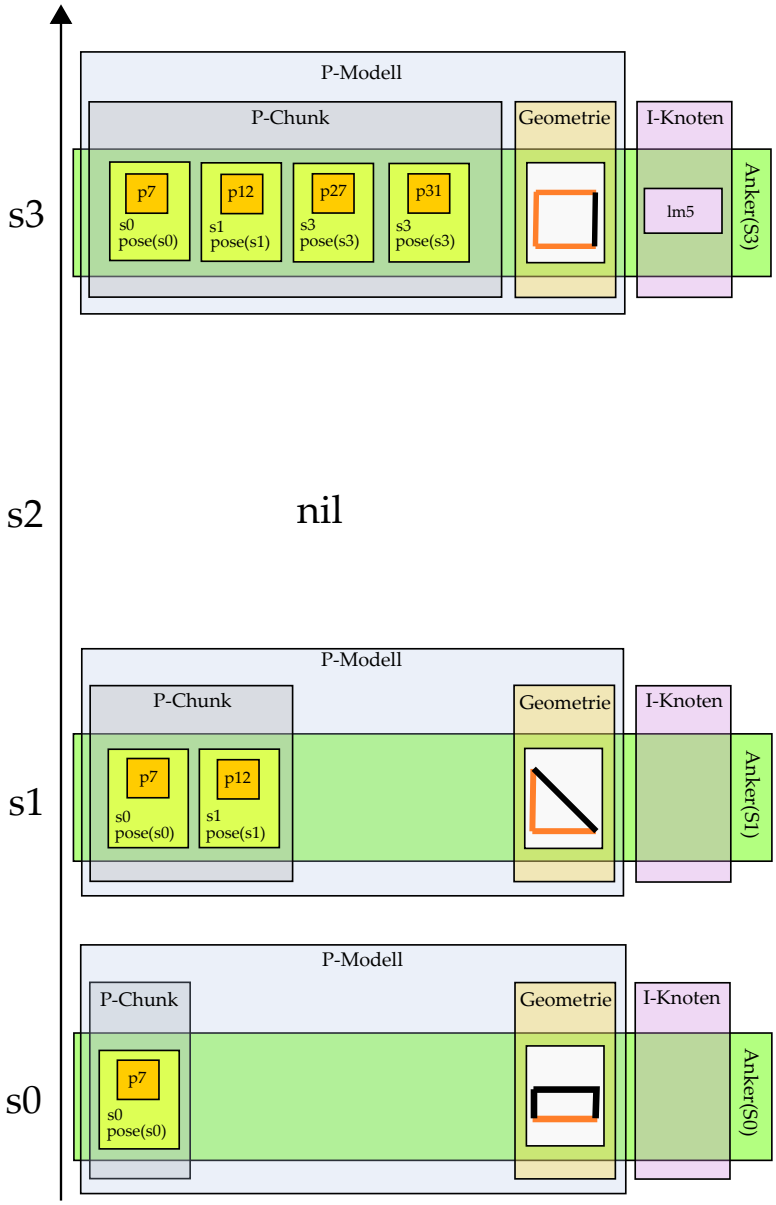


Abbildung 4.3: Entwicklung eines Ankers über die Zeit: In s0 wird die Landmarke erstmals vom Agenten wahrgenommen und ein neuer Anker wird initialisiert. In der Situation s1 wird eine weitere Seite der Landmarke perzipiert und das Perzeptionsmodell aktualisiert. Als nächstes (s2) gerät die Landmarke aus dem Fokus des Agenten (gekenntzeichnet durch *nil*). Gleich zwei Perzeptionsereignisse sind in s3 mit der Landmarke verbunden; zudem kann eine Koreferenz mit einem Instruktionsknoten hergestellt werden.

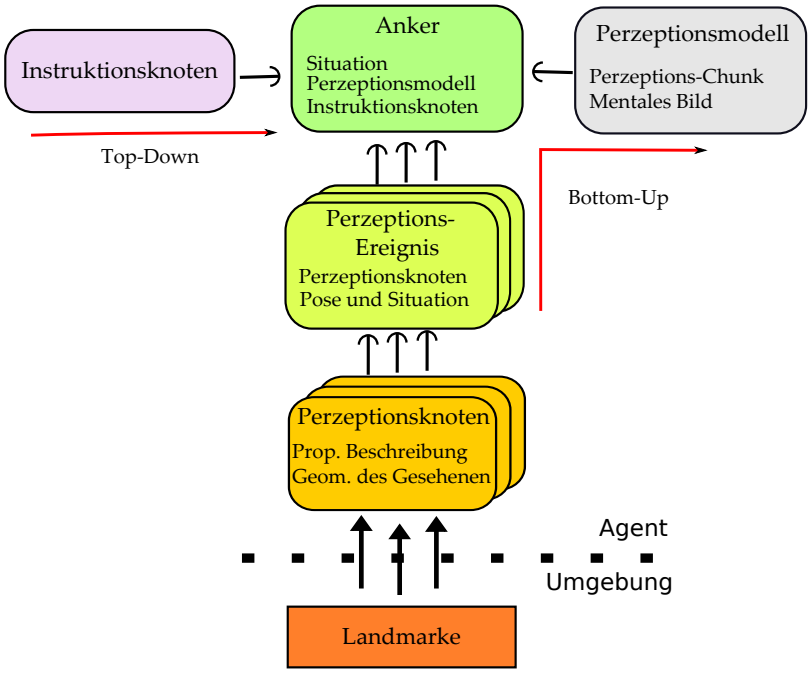
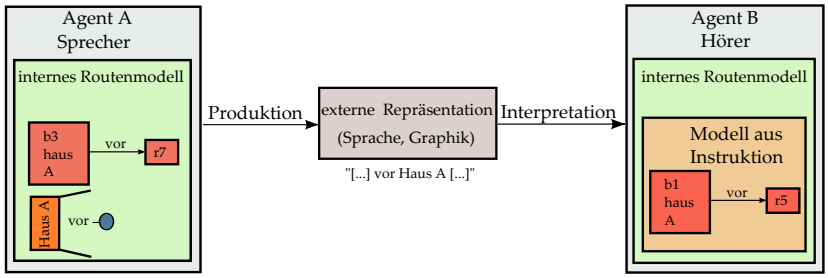


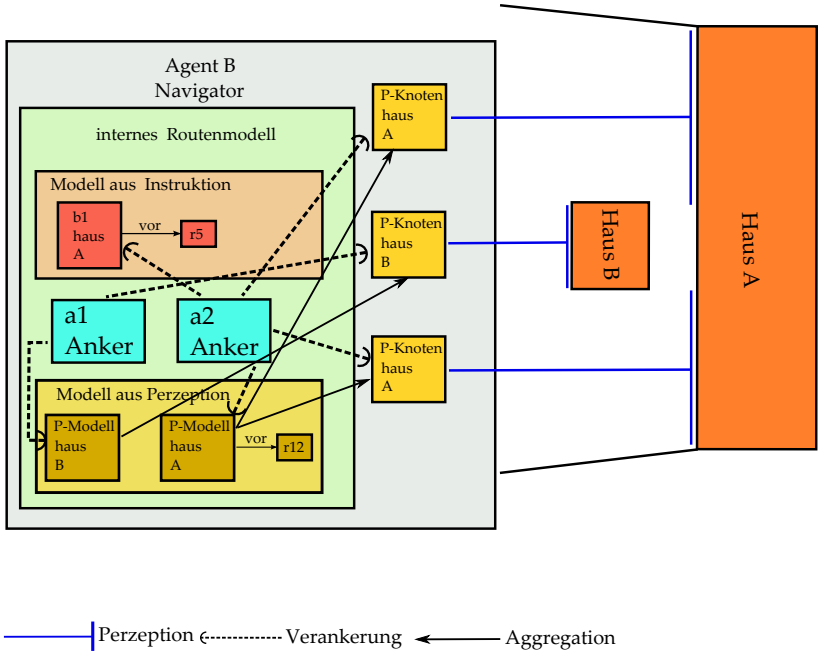
Abbildung 4.4: Perzeptuelle Verankerung einer ausgedehnten Landmarke

der Umgebung. In diesem Beispiel sind das zwei Landmarken „Haus A“ und „Haus B“. Haus A wird mittig von Haus B verdeckt. Es werden zwei Perzeptionsereignisse zu Haus A und ein Perzeptionsereignis zu Haus B induziert. Zu jedem Perzeptionsereignis wird ein Anker gesucht, dem es hinzugefügt werden kann. Ist die Suche erfolgreich, identifiziert der Agent den Referenten des Perzeptionsereignis als „schon einmal gesehen“. Misslingt die Suche hingegen, gilt der Referent des Perzeptionsereignis als eine neue, zuvor noch nicht wahrgenommene Landmarke und es wird ein neuer Anker erstellt. In dem Beispiel ist die Wiedererkennung erfolgreich und beide Perzeptionsereignisse zu Haus A werden in denselben Anker verankert. Das mit Haus B assoziierte Perzeptionsereignis wird einem eigenen Anker zugefügt. Auf Grundlage der Aggregation der Landmarkenteile von Haus A kann im nächsten Schritt die Vor-Region zu Haus A gebildet werden. Daraufhin identifiziert der Agent die räumliche Konstellation als genau die in der Routeninstruktion beschriebene Konstellation und etabliert Koreferenzbeziehungen zwischen a) dem Instruktionsknoten, der ein Haus mit einem Label „A“ beschreibt und dem Perzeptionsmodell eines Hauses das das Label „A“ trägt und b) dem Instruktionsknoten, der die Vor-Region beschreibt und die auf Grundlage des aktuellen Perzeptionsmodells von Haus A gebildete Vor-

Region. Man bemerke, dass im Gegensatz zu der Definition eines Match auf Seite 12 nicht mehr direkt Koreferenzen zwischen Instruktionsknoten und Perzeptionsknoten aufgelöst werden, sondern zwischen Instruktionsknoten und den jeweils aktuellsten Perzeptionsmodellen in den Ankern (das gilt für alle Typen von Landmarken, die der Agent verankert). Das aktuellste Perzeptionsmodell eines Ankers ist das Perzeptionsmodell, das bei der letzten Aktualisierung des Ankers erstellt wurde.



(a) Instrukionsphase



(b) Navigationsphase

Abbildung 4.5: (a) Interpretation der Routeninstruktion und Aufbau eines Modell der Route (CRIL-Graph). (b) Aufbau von Perzeptionsmodellen und Verankerung

Die Wiedererkennung von Landmarken ist ein zentraler Aspekt der perzeptuellen Verankerung und stellt eine kognitive Fähigkeit dar, die für die Navigation von Bedeutung ist. Die folgenden Abschnitte werden die Wiedererkennung ausgedehnter Landmarken thematisieren und aufzeigen, an welchen Stellen der Geometrische Agent davon profitiert.

4.2 Wiedererkennung ausgedehnter Landmarken

Damit ein mentales Modell einer Landmarke im Gedächtnis inkrementell aufgebaut werden kann, werden die einzelnen zu einer Landmarke gehörenden Wahrnehmungsereignisse in einen gemeinsamen Anker verankert. Dafür wird für jedes Perzeptionsereignis entschieden, welchem Anker es zuzuordnen ist.

Das Problem der Wiedererkennung eines zu einem früheren Zeitpunkt verankerten Objektes hat in der Verankerungsliteratur noch relativ wenig Beachtung gefunden. Dabei handelt es sich bei dieser Aufgabe um eine der drei Grundfunktionen des Verankerung-Frameworks (vgl. Abschnitt 3.3). Die Wiedererkennung eines Objektes ist immer ein Fall definiter Referenz (Karlsson et al., 2008). Bei einer Wiedererkennungsaufgabe ist ein bestimmtes Objekt gesucht, dessen perzeptuelle Signatur kompatibel mit einem zuvor generierten Anker ist. Im Falle von Uneindeutigkeiten zwischen der symbolischen Beschreibung des gesuchten Objektes und der perzeptuellen Signatur eines aktuell perzipierten Objektes planen Karlsson et al. (2008) Wahrnehmungsaktionen um weitere Informationen über das aktuell fokussierte Objekte zu sammeln und so die Situation zu disambiguieren. Damit dies erfolgreich sein kann, muss allerdings die perzeptuelle Signatur prinzipiell genügend messbare Merkmale enthalten können, um die Ambiguität aufzulösen.

Wenn die durch das Perzeptionssystem extrahierbaren Merkmale nicht ausreichen, um die Identität oder Nicht-Identität von Objekten anhand ihrer perzeptullen Signaturen festzustellen, kann häufig Weltwissen die Situation auflösen. Untersuchungen über Identifizierungsstrategien perzeptuell ununterscheidbarer Objekte beim Menschen – wie in (Santore, 2005) – geben Aufschluss über mögliche Problemlösungen. Der Geometrische Agent identifiziert mit Bäumen, Häusern und Wegen Landmarken, die nach Santore zu der Klasse der inmobilen und stationären Objekte gehören. Diese Einordnung korrespondiert mit der taxonomischen Spezifizierung von Landmarken innerhalb des Geometrischen Agenten als „perzipierbar“ und „inmobil“ (vgl. Abbildung 2.5). Der Agent kann deswegen annehmen, dass Landmarken ihre Position nicht aus

eigener Kraft ändern und dass sie auch von keiner äußeren Kraft an einen anderen Ort verlegt werden. Menschen schließen nach Santore in diesem Fall anhand der Lokation des Objektes über dessen Identität, was für die Wiedererkennung im Geometrischen Agenten zwei Implikationen zulässt: i) Wenn zu einem Zeitpunkt eine Landmarke an einer Stelle lokalisiert ist, und wenn zu einem anderen Zeitpunkt eine Landmarke an derselbe Stelle lokalisiert ist, dann handelt es sich um ein und dieselbe Landmarke – die Wahrnehmungsereignisse sind koreferent; ii) wenn zu einem Zeitpunkt eine Landmarke an einer Stelle lokalisiert ist und an dieser Stelle noch keine andere Landmarke lokalisiert wurde, dann handelt es sich um eine neue, noch nicht wahrgenommene Landmarke. Die Objekte in den Experimenten von Santore (2005) sind jedoch ausschließlich kleine Objekte, die leicht vollständig in das Blickfeld des Betrachters fallen und punktartig lokalisierbar sind. Die punktartige Lokalisierung einer beliebig ausgedehnten Landmarke (beispielsweise durch die Berechnung ihres Masseschwerpunktes), von der zudem womöglich nur wenige Teilstücke bekannt sind, ist für einen Betrachter nicht möglich. Dennoch kann die Lokation der Teilstücke Aufschluss darüber geben, ob sie zu einer gemeinsamen Landmarke gehören oder nicht. Das für die perzeptuelle Verankerung im Geometrischen Agenten vorgeschlagene Entscheidungsverfahren nutzt neben der in den Perzeptionsknoten enthaltenen propositionalen Beschreibungen (Merkmalsbündel) auch die perzipierte geometrische Information, um die Kompatibilität des Perzeptionsereignis mit einem Anker zu ermitteln.

Das Problem der Wiedererkennung ist in Abbildung 4.6 dargestellt. Für jedes neue Perzeptionsereignis muss ein Entscheidungsverfahren ermitteln, ob die aktuell perzipierte Landmarke bereits schon einmal gesehen wurde und somit ein Anker existiert, in den das Perzeptionsereignis verankert werden kann, oder ob das Perzeptionsereignis zu keinem Anker kompatibel ist und somit ein neuer Anker erstellt werden soll.

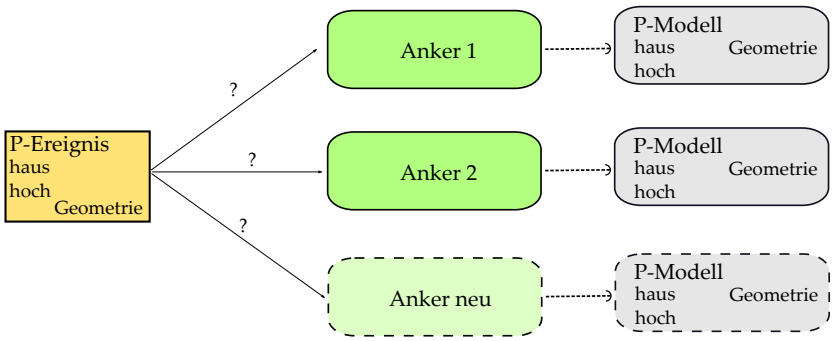


Abbildung 4.6: Das Wiedererkennungsproblem

Dazu werden die aktuellesten Perzeptionsmodelle der Anker auf Kompatibilität mit dem neuen Perzeptionsereignis untersucht. Die propositionale Information wird anhand von in Helwich (2003) vorgeschlagenen Ähnlichkeitsmaßen zwischen Knoten (hier: Perzeptionsereignis und Perzeptionsmodell) verglichen. Ist die Ähnlichkeit hinreichend hoch, ist das ein Indiz dafür, dass es sich beim Referenten des Perzeptionsereignisses um eine zuvor schon einmal wahrgenommene Landmarke handelt. Allerdings können mehrere Landmarken existieren, die durch äquivalente Merkmalsbündel repräsentiert sind (beispielsweise gibt es auf dem Informatik-Campus der Universität Hamburg mehrere hohe Gebäude).

Um Situationen aufzulösen, in denen die Information aus dem Merkmalsbündel nicht aureicht, wird die Multimodalität des Perzeptionsknotens ausgenutzt. Auf Grundlage der geometrischen Repräsentation des gesehenen Teils der Landmarkenkontur werden geometrischen Kritieren angesetzt, anhand derer entschieden werden kann, welchem der Anker das Perzeptionsereignis zuzuordnen ist (oder dass es zu keinem existierenden Anker kompatibel ist).

Definition 15 (Kompatibilität zwischen P-Ereignis und P-Modell) *Ein Perzeptionsereignis ist kompatibel zu einem Perzeptionsmodell genau dann, wenn die propositionale Beschreibung im Perzeptionsereignis mit der propositionalen Beschreibung im Perzeptionsmodell kompatibel ist und ein geometrisches Zusammenhangskriterium zwischen dem mentalen Bild des Perzeptionsmodells und der perzipierten geometrischen Form aus dem Perzeptionsereignis erfüllt ist.*

Bei der Berechnung der geometrischen Kompatibilität wird zunächst grundsätzlich zwischen solchen Landmarken unterschieden, die geometrisch durch ein geschlossenes Polygon repräsentiert werden (Häuser und Bäume) und solchen Landmarken, die geometrisch durch einen offenen Linienzug repräsentiert sind (Wege).

4.2.1 Geometrisches Zusammenhangskriterium für polygonal ausgdehnte Landmarken

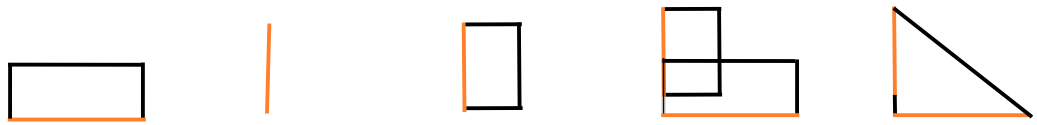


Abbildung 4.7: Überlappungskriterium

Überlappen sich das mentale Bild eines Perzeptionsmodells und die mental erweiterte Geometrie eines aktuell perzipierten Teilstückes einer polygonal ausgedehnten Landmarke, dann ist das Überlappungskriterium erfüllt.

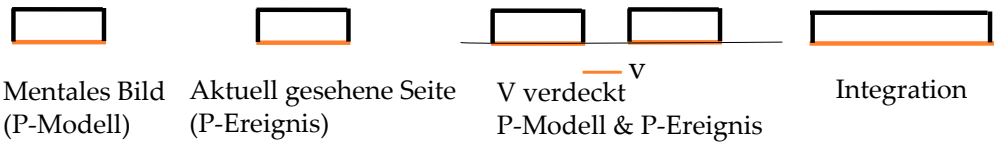


Abbildung 4.8: Verdeckungskriterium

Gibt es eine Landmarke V, die zwei mentale Bilder verdeckt, ist das Verdeckungskriterium erfüllt. Der Agent nimmt an, dass das Perzeptionsereignis eine Fortsetzung des Perzeptionsmodells ist. Dieses Kriterium ist aus psychologischen Studien zur Wahrnehmung partiell verdeckter Objekte als kohärente Einheit motiviert (u.a. Slater et al., 1990).

4.2.2 Geometrisches Zusammenhangskriterium für linear ausgedehnte Landmarken

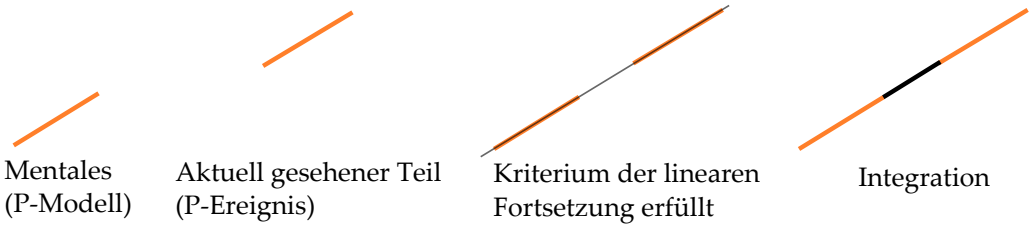
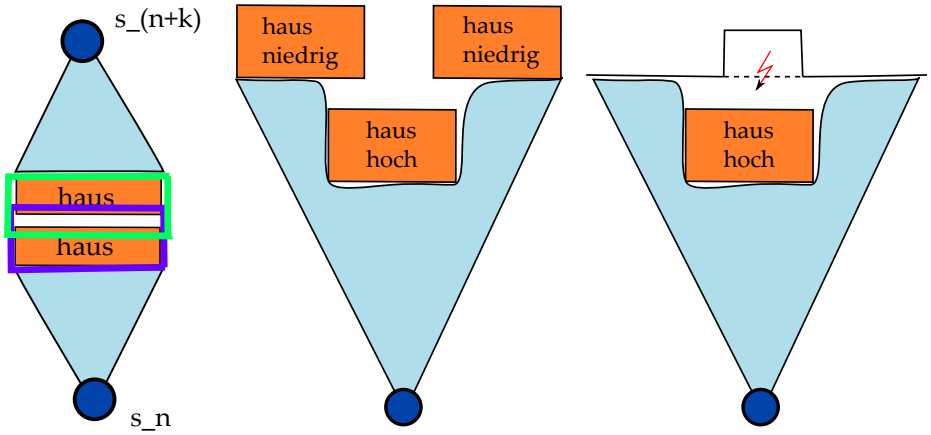


Abbildung 4.9: Lineares Fortsetzungskriterium

Liegen das mentale Bild eines Perzeptionsmodells und die Geometrie eines aktuell perzipierten Teilstückes einer linear ausgedehnten Landmarke auf einer Geraden, dann ist das Kriterium der linearen Forsetzung erfüllt.

4.2.3 Bewertung der geometrischen Zusammenhangskriterien

Die geometrischen Zusammenhangskriterien sind ein Mittel, um bei mehreren äquivalenten Merkmalsbündeln eine Entscheidung über die Zuordnung eines Perzeptionsereignisses zu einem Anker treffen zu können. Dabei werden die beiden im Perzeptionsknoten verfügbaren Repräsentationsformen komplementär genutzt. Es lassen sich



(a) Überlappung (b) Verdeckung (c) Lineare Fortsetzung

Abbildung 4.10: Problematische Konstellationen

allerdings zu allen drei Kriterien auch Fälle finden, in denen die getroffenen Entscheidungen nicht mit der tatsächlichen Konstellation in der Welt übereinstimmen. Dann kommt es dazu, dass das mentale Modell der Welt nicht mit der tatsächlichen Welt übereinstimmt. Drei Beispiele für solche Fälle sind in Abbildung 4.10 aufgezeigt. Es ist möglich, dass die geometrische Ausdehnung einer Landmarke „nach hinten“ weniger stark ist, als es dem mentale Bild des Agenten entsprechen würde (in Abbildung 4.10(a) durch die beiden fetten Rechtecke angedeutet). Dann kann es passieren, dass zwei nah beieinanderstehenden Landmarken für eine einzige Landmarke gehalten werden. Das Beispiel in Abbildung 4.10(b) zeigt einen Fall, in dem das Verdeckungskriterium angewendet wird. Allerdings verdeckt das hohe Haus lediglich die Lücke zwischen den beiden niedrigen Häusern.[1] Das dritte Beispiele – Abbildung 4.10(c) – zeigt die Problematik bei der Zuordnung von Wegstücken zu Wegen durch Anwendung des Kriteriums der linearen Fortsetzung. Zwar liegen die beiden Wegstücke links und rechts von dem Haus auf einer Gerade, allerdings nehmen hinter dem Haus die Wege einen Verlauf, der nicht mit der Annahme des Agenten korrespondiert.

Die Schlüsse, die durch die geometrischen Kriterien gezogen werden, sind abduktive Schlüsse. Der Agent weiß „wenn verschiedene Teile einer (ein und derselben) Landmarke wahrgenommen werden, dann können Überlappungen nach dem Über-

[1] Man könnte argumentieren, dass in der Realität Merkmale zwischen den beiden niedrigen Häusern existieren, die verhindern, dass die Häuser als Einheit wahrgenommen werden. Der Agent nimmt aber nur beschränkt viele Merkmale wahr und es lassen sich für jede sensorische Ausstattung Welten finden, in denen die wahrnehmbaren Merkmale nicht ausreichen, um die Unterscheidung sicher zu treffen.

lappungskriterium auftreten“, „wenn eine Landmarke partiell so verdeckt wird, dass mehrere Perzeptionsereignisse zu der Landmarke auftreten, dann trifft das Verdeckungskriterium zu“ und „wenn Wegteile einen zusammenhängenden Weg bilden, dann trifft das Kriterium der linearen Fortsetzung zu“. Der Agent schließt aus der Beobachtung, dass eines der geometrischen Kriterien zutrifft, auf die linke Seite der jeweiligen Implikation. Aus der Beobachtung – beispielsweise – , dass das Überlappungskriterium zutrifft, schließt der Agent also, dass er mehrere Teile einer (ein und derselben) Landmarke wahrgenommen hat und integriert die Wahrnehmungen in ein gemeinsames Modell.

Genauso wie separate Landmarken fälschlicherweise zusammengefasst werden, kann es auch passieren, dass Teile ein und derselben Landmarke nicht als zusammenhängend erkannt werden. Ein Beispiel ist in Abbildung 4.11 dargestellt. Die Landmarke ist sehr weit nach hinten ausgedehnt. Der Agent bringt das Wahrnehmungsereignis in Situation $s1$ nicht mit dem Wahrnehmungsereignis aus $s0$ in Zusammenhang. In Situation $s2$ gibt es dann aber Evidenz für den Zusammenhang aller drei Wahrnehmungsereignisse.

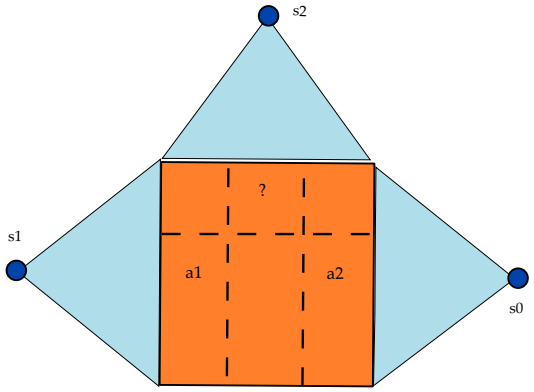
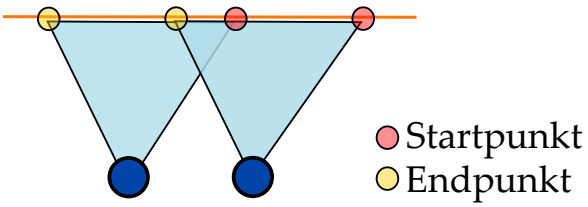


Abbildung 4.11: Zusammenhang wird spät erkannt

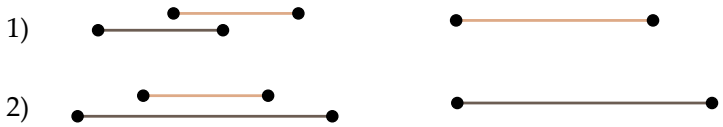
Das mentale Bild, das der Agent über eine Landmarke aufbaut, ist die jeweils beste Erklärung seiner Wahrnehmungen. Sobald aber Evidenz für die Fehlzuordnung von Perzeptionsereignissen zu Ankern vorliegt, muss der Agent die Zuordnung von Perzeptionsereignissen zu Ankern und damit sein mentales Bild der Landmarken revidieren können. Dieses Problem wird allerdings im Rahmen dieser Arbeit nicht weiter diskutiert und die Lösung späteren Arbeiten überlassen. Trotz der aufgezeigten Probleme, verbessert die Wiedererkennung die Navigationsleistung des Agenten in vielerlei Fällen.

4.3 Rekonstruktion der Geometrie einer Landmarke

Der navigierende Agent nimmt die Landmarken im Allgemeinen nur partiell wahr. Um das mentale Bild einer Landmarke zu erzeugen, ergänzt der Agent die gesehenen Teile der Landmarke um hinzugedachte Teile, so dass ein geschlossenes Polygon entsteht. Gesucht ist die geometrische Rekonstruktion, die die Perzeptionsereignisse am besten erklärt. Die gesehenen Teile der Kontur der Landmarke sind geometrisch als gerichtete Linien modelliert, die aus einer Menge gerichteter Strecken bestehen. Der Algorithmus, der hier vorgeschlagen wird, dekomponiert die Menge der Linien, die aus den Perzeptionsereignissen eines Ankers extrahiert werden, zunächst in die Menge der Strecken. Dabei kann es zu sich überlappenden Strecken kommen, wenn der Agent bestimmte Teile einer Landmarke mehrmals gesehen hat wie in Abbildung 4.12 beispielhaft dargestellt. Diese Überlappungen werden zu einer einzigen Strecke fusioniert.



(a) Überlappende Wahrnehmungen



(b) Überlappende Strecken (links), Fusion (rechts)

Abbildung 4.12: Streckenfusionierung

Daraufhin wird eine Fallunterscheidung zwischen linear ausgedehnten Landmarken (wie Wege) und polygonal ausgedehnten Landmarken (wie Häuser und Bäume) getroffen. Wege sind als Linien repräsentiert, die nur eine Strecke enthalten, wohingegen die Repräsentation der Kontur polygonal ausgedehnter Landmarken eine Menge von Strecken enthält. Im Fall linear ausgedehnter Landmarken wird aus dem Startpunkt der ersten Strecke und dem Endpunkt der letzten Strecke eine neue Linie erzeugt. Für polygonal ausgedehnte Landmarken wird mittels eines Suchverfahrens (Algorith-

mus 1)[2] eine Menge von Strecken gesucht, die zusammen mit den Strecken aus der Perzeption eine geschlossene Geometrie der Landmarke ergeben.

Algorithm 1 SucheFehlendeStrecken(Liste L)

Require: Liste L aktuell hinzugefügter Strecken.
Ensure: Liste hinzugefügter Strecken, die zusammen mit den gesehenen Strecken ein geschlossenes, einfaches Polygon bilden.

```
1: if hinzugefügte Strecken ∪ gesehene Strecken bildet ein geschlossenes Polygon then
2:   return Liste hinzugefügter Strecken L
3: end if
4: if #hinzugefügte Strecken < #gesuchte Strecken then
5:   for Alle Strecken SK aus der Liste der Kandidaten LK do
6:     if Hinzunahme von SK verletzt die Randbedingung nicht then
7:       Füge SK zur Liste L hinzu
8:       Resultat ← SucheFehlendeStrecken(L)
9:       if Resultat ist ungleich null then
10:        return Resultat
11:      end if
12:      Entferne SK aus L
13:    end if
14:  end for
15: end if
16: return null
```

Für das Suchverfahren wird ausgenutzt, dass die Strecken in dieselbe Richtung gerichtet sind – also der Startpunkt relativ zur Pose des Agenten immer rechts vom Endpunkt liegt (vgl. Abbildung 4.13(a)). Außerdem ist es unmöglich für den Agenten, die Innenseite der Kontur wahrzunehmen, so dass kein Teilstück durch zwei entgegegengesetzt gerichteten Linien repräsentiert sein kann.

Die Suche nach den fehlenden Strecken kann deswegen ohne Zwischenschritt in das Problem transformiert werden, eine Paarung zwischen Startpunkten aus $SP_{Grad<2}$ und Endpunkten aus $EP_{Grad<2}$ zu finden. $SP_{Grad<2}$ ist die Menge der Startpunkte der perzipierten Strecken, die einen Grad kleiner 2 haben ($EP_{Grad<2}$ analog dazu die Menge der

[2] Es gibt Fälle, in denen keine Rekonstruktion mit diesem Verfahren erreicht werden kann – beispielsweise, wenn erst eine einzige Seite der Landmarke perzipiert wurde. Dann wird auf Grundlage der gesehenen Linie ein Polygon mit einer festgesetzten Tiefe konstruiert.

Endpunkte mit Grad kleiner 2).[3] Insgesamt gibt es $|SP_{Grad<2}|!$ mögliche Paarungen.[4] Es soll eine Paarung so gefunden werden, dass die gesehenen Strecken zusammen mit den hinzugedachten Strecken die folgenden Randbedingungen nicht verletzen:

1) Es gibt keine Schnittpunkte zwischen Strecken außer an den Start-/Endpunkten,

2) jeder Endpunkt einer Strecke ist Startpunkt genau einer anderen Strecke,

3) insgesamt entsteht genau ein geschlossenes Polygon entsteht, und

4) keine hinzugedachte Strecke verdeckt eine gesehene Strecke.

Zunächst werden alle Strecken berechnet, die Startpunkte aus $SP_{Grad<2}$ mit Endpunkten aus $EP_{Grad<2}$ verbinden und keine gesehene Strecke schneiden. Dieser Menge bildet die Liste von Kandidaten. Dann wird sukzessive versucht, die Kandidaten der Ergebnisliste L hinzuzufügen (ab Zeile 5). Dabei darf ein neuer Kandidat keinen anderen Kandidaten in L berühren und keine gesehene Strecke verdecken (Bedingungen 1 und 4, Zeile 6). Verletzt er diese Randbedingung nicht, wird er der Liste L hinzugefügt, mit der der Algorithmus rekursiv aufgerufen wird. Hat L genau die Länge $|SP_{Grad<2}|$ erreicht, werden die Strecken aus L mit den gesehenen Strecken fusioniert und es wird geprüft, ob aus der der Vereinigung der Streckenmengen (gesehen und hinzugedacht) um genau ein geschlossenes Polygon handelt, an dem alle Start- und Endpunkte der gesehenen Strecken beteiligt sind (Bedingungen 2 und 3). Ein Beispiel für das Ergebnis einer Rekonstruktion zeigt Abbildung 4.13(b).

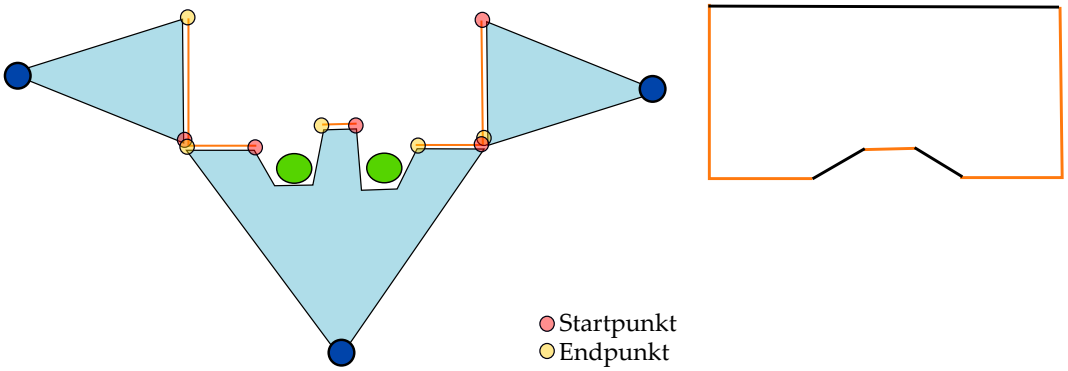
Zwei weitere Anwendungsfälle sind in Abbildungen 4.14(a) und 4.14(b) dargestellt. Sie zeigen den inkrementellen Aufbau eines mentalen Bildes durch die Aggregation der geometrischen Information koreferenter Perzeptionsereignisse. In 4.14(a) ist das Verdeckungskriterium erfüllt und in 4.14(b) trifft das Überlappungskriterium zu.

4.3.1 Relevante Arbeiten zur Linienrekonstruktion

Die Rekonstruktion kognitiv plausibler Linienzüge aus Punktmengen ist ein aktuelles Forschungsgebiet (Althaus et al., 2000; Galton, 2008). Allerdings wurde noch nicht untersucht, wie sich diese Verfahren anwenden lassen, wenn bereits Teilstrecken bekannt sind. Auch wäre es lohnenswert zu untersuchen, wie Rekonstruktionsverfah-

[3] Der Geometrische Agent repräsentiert gesehene Teile immer als Linien, niemals nur als Punkt. Deswegen ist jeder Start- bzw. Endpunkt einer Strecke mindestens vom Grad 1.

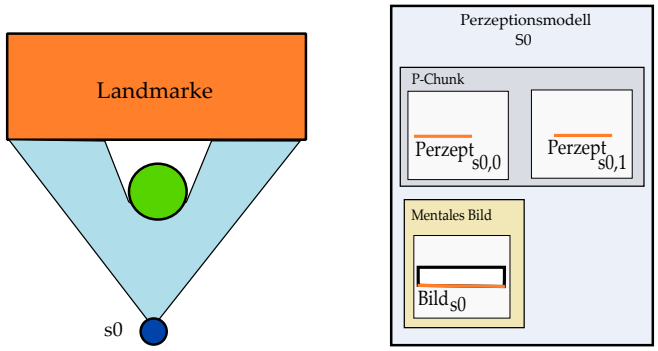
[4] $|SP_{Grad<2}| = |EP_{Grad<2}|$



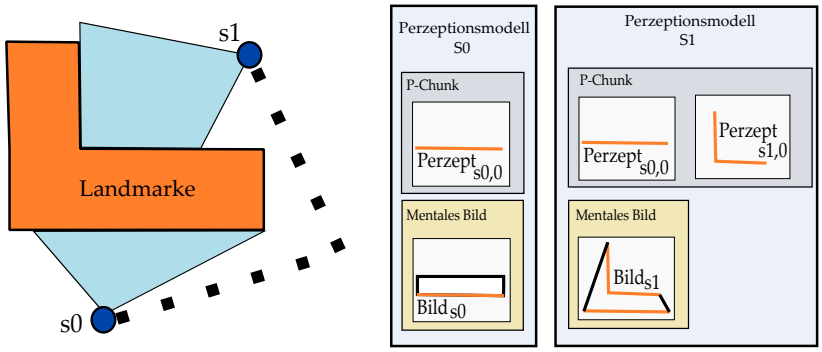
(a) Wahrnehmung verschiedener Teile der Kontur einer Landmarke

(b) Rekonstruktion

Abbildung 4.13: Rekonstruktion der Landmarkenkontur anhand unvollständiger Information aus der Wahrnehmung



(a) In einem Schritt



(b) Über mehrere Schritte

Abbildung 4.14: Aggregation koreferenter Perzeptionsereignisse. Gesehene Seiten (orange) und hinzugedachte Seite (schwarz) im P-Modell. (a) erfüllt Verdeckungskriterium, (b) erfüllt Überlappungskriterium

ren mit Hintergrundwissen verbessert werden können.[5] Wahrnehmungspsychologische Arbeiten zur Vervollständigung teilweise verdeckter Objekte beim Menschen gibt es von Kellman (2003).

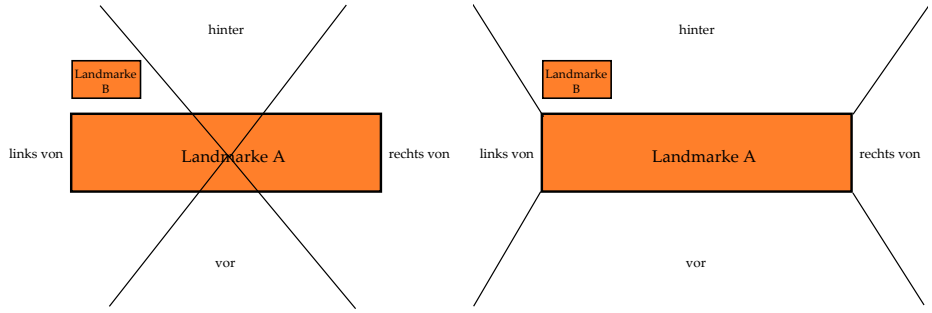
Für die Robotik beschäftigt sich der Ansatz von Modayil & Kuipers (2008) mit dem Erwerb geometrischer Repräsentation von Objekten aus Laserscandaten. Dafür nimmt der Roboter ein Objekt aus verschiedenen Perspektiven wahr und integriert die Perzepte zu einer aggregierten Geometrie. Allerdings besteht das Ziel des Geometrischen Agenten nicht darin, ein ganz exaktes Bild der Geometrie einer Landmarke zu erwerben (dafür müsste er die Landmarke vollständig umlaufen). Vielmehr sollte er auf Grundlage partieller Perzepte mit Hilfe von Vorwissen über eine prototypische Kontur von Häusern/ Bäumen/ Wegen ein möglichst gutes mentales Bild aufbauen. Für eine Anwendung in einer realen Umgebung mit realen Abstandssensoren muss das Rekonstruktionsverfahren mit verrauschten Daten korrekt umgehen. Einen Ansatz dafür findet sich in (Latecki et al., 2005).

4.3.2 Kohärentes Landmarkenmodell zur relativen Lokalisierung

Der Geometrische Agent profitiert von der Zusammenführung einzelner Perzeptionsereignisse bei der relativen Lokalisierung von Landmarken. Im Gegensatz zu dem in Abschnitt 3.3.3 skizzierten Verfahren von Melchert et al. (2007), ist die geometrische Repräsentation der vollständigen Kontur einer Landmarke nicht Teil eines Perzeptionsereignisses im Geometrischen Agenten. Für die relative Lokalisierung von Landmarken werden geometrische Repräsentationen der Landmarken vorausgesetzt. Gebräuchlich sind zweidimensionale, polygonale Repräsentationen, die die Kontur der Landmarke abbilden, oder punktartige Repräsentationen, für die der Zentroid einer polygonalen Repräsentation bestimmt wird. Beide Varianten kommen zu genaueren Lokalisierungs-Ergebnissen, je mehr Wissen über die Ausdehnung von Referenzobjekt und Lokalisationsobjekt bekannt ist.

Punktartige Repräsentationen von Landmarken besitzen formal gute inferentielle Eigenschaften, eignen sich aber weniger gut für das Schließen über die Lokalisation ausgedehnter Landmarken (vgl. Schmidtke, 2001). Insbesondere in Konstellationen, an denen Landmarken mit unterschiedlichen Größen und unregelmäßigen Proportionen auftreten, eignen sich Punkt-Repräsentationen nicht. An dieser Stelle ist anzumerken, dass der im Geometrischen Agenten gegenwärtig angewendete Ansatz zur Bildung

[5] Der Geometrische Agent weiß, dass es sich um ein Haus/einen Weg/einen Baum handelt.



(a) Punktartige Lokalisierung: Landmarke B ist *links von* Landmarke A lokalisiert

(b) Polygonale Lokalisierung: Landmarke B ist *hinter* Landmarke A lokalisiert

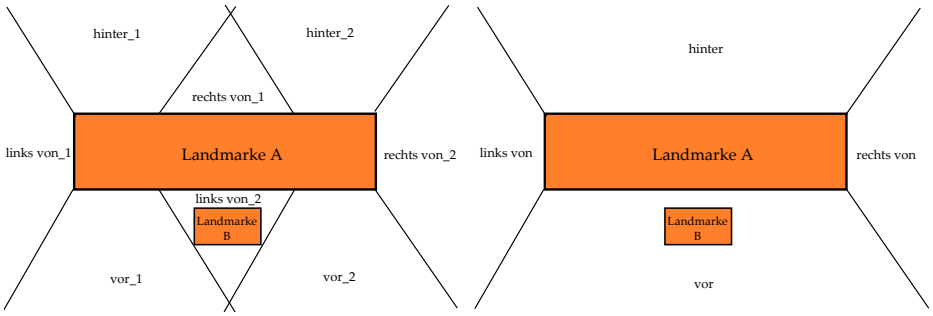
Abbildung 4.15: Punktartige und polygonale Repräsentationen zur relativen Lokalisierung

von Regionen/Akzeptanzsektoren die Geometrie der Landmarke als einziges Kriterium einbezieht. Psychologische Studien zur Raumkognition zeigen deutliche Evidenz dafür, dass der Mensch bei Lokalisierungsaufgaben neben der Geometrie auch funktionale und kontextuelle Kriterien verwendet (Carlson, 2000; Coventry et al., 2001; Richards et al., 2004). Eine angemessene geometrische Repräsentation bleibt aber eine wichtige Voraussetzung für eine plausible Regionenbildung beim räumlichen Schließen.

Abbildung 4.15 zeigt die Problematik punktartiger Repräsentationen an einem einfachen Beispiel: Landmarke B (Lokalisationsobjekt) befindet sich hinter Landmarke A (Referenzobjekt). Im Falle einer punktartigen Repräsentation des Referenzobjektes (hier der Zentroid der Landmarke A) befindet sich Landmarke B aber in der Links-Region des relativen Referenzrahmens zu Landmarke A (vgl. Abbildung 4.15(a)). In der Variante nach (Peuquet & Ci-Xiang, 1987) wird der triangulare Akzeptanzsektor relativ zur betrachteten Richtung vor- bzw. zurückgeschoben, so dass die Endpunkte der betrachteten Seite des Referenzobjektes auf den Kanten des Sektors liegen (vgl. Abbildung 4.15(b)).

Das Problem der Lokalisierung nur teilweise wahrgenommener Landmarken ist in Abbildung 4.16 dargestellt. Dieser Fall kann aufgrund partieller Verdeckungen auftreten oder, weil der Agent zu unterschiedlichen Zeitpunkten verschiedene Teile der Landmarke wahrnimmt. In dem Beispiel befindet sich Landmarke B (Lokalisationsobjekt) vor Landmarke A (Referenzobjekt) und verdeckt somit Landmarke A mittig, so dass zwei perzeptuelle Eindrücke von Landmarke A entstehen. Um inferieren zu können, dass die Vor-Relation zwischen Landmarke A und Landmarke B gilt, bedarf

es einer geeigneten Zusammenfassung beider mit Landmarke A assoziierten Wahrnehmungsereignisse, die die perzeptuelle Verankerung leistet.



(a) Lokalisierung vor der Integration (b) Lokalisierung nach der Integration

Abbildung 4.16: Relative Lokalisierung vor und nach der Integration von Perzeptionsereignissen

4.4 Kognitive Karte

Die Anker, die der Agent während der Navigation aufbaut, können zu einer überblicksartigen, mentalen Repräsentation der Umgebung zusammengefügt werden. Tolman (1948) führt den Begriff der kognitiven Karte ein, um Ergebnisse zahlreicher Experimente mit Ratten zu erklären. Seine Beobachtungen konnte er sich nicht anhand des damals vorherrschenden behavioristischen Paradigmas erklären, nach dem sich Lernen und Verhalten alleine durch die Verstärkung von Reiz-Reaktions-Verbindungen ergibt. Tolman schließt, dass Lernen räumlicher Konstellationen mit dem Aufbau mentaler Modelle einhergeht. Lynch (1960) betrachtet Städte unter dem Aspekt der *Imageability*, der beschreibt, wie gut eine Stadt kognitiv organisiert werden kann. Dafür versetzt er sich in die Perspektive eines durch die Stadt navigierenden Menschen und analysiert unter Berücksichtigung der menschlichen Wahrnehmung, welches Bild der Stadt ein Mensch erwirbt und aus welchen Entitäten dieses Bild besteht. Lynch beobachtet, dass das mentale Modell, das Menschen von einer Stadt aufbauen, abhängig von den Erfahrungen ist, die der Mensch mit der Stadt macht. Insbesondere macht die Art und Weise der Fortbewegung – ob als Autofahrer, zu Fuß, mit öffentlichen Verkehrsmitteln oder mit dem Fahrrad – einen Unterschied. Anhand einer mentalen Repräsentation der Umgebung sind Menschen in der Lage, sich in Umgebungen zu orientieren und zu navigieren, die so groß sind, dass sie nicht vollständig überblickt werden können. Ein Schlüsselelement in Lynchs Konzeption räumlicher Repräsenta-

tionen ist die Landmarke. Landmarken sind auffällige Merkmale in der Landschaft, die sich von ihrer Umgebung abheben und so Aufmerksamkeit erregen. Somit stellen sie Referenzpunkte dar, relativ zu denen der Raum mental organisiert werden kann. Anhand von Landmarken wird Überblickswissen der Umgebung generiert, anhand dessen ein Agent sich im Raum orientieren und neue Routen erschließen kann (Sorrows & Hirtle, 1999). Eine interne Realität, die sich von der externen Realität unterscheidet, bedeutet auch, dass das räumliche Wissen individuell ist. Dennoch gibt es interpersonelle Gemeinsamkeiten hinsichtlich der Konzeptualisierung des Raumes, die die Kommunikation ermöglichen. Diese Gemeinsamkeiten entstehen dadurch, dass der Prozess des Wissenserwerbs sich gleicht. Wissen über die räumlichen Konstellationen basieren auf der direkten Interaktion mit der Umgebung (Peuquet, 2002).

Der Geometrische Agent erwirbt Wissen über die Umgebung auf zwei Wegen. Zunächst liefert die Routeninstruktion eine partielle Beschreibung der Umgebung entlang der Route. Während der Navigation nimmt der Agent über seine sensorische Ausrüstung Landmarken wahr und integriert dieses Landmarkenwissen in landmarkenspezifische Perzeptionsmodelle. Diese Perzeptionsmodelle beinhalten unter anderem eine geometrische Repräsentation, die sich aus den Konturen der gesehenen Landmarkenteile und den hinzugedachten Konturstücken zusammensetzt. Die geometrischen Repräsentationen sind in ihrer Koordinatendarstellung bezüglich eines allozentrischen Referenzsystems spezifiziert. So kann die Menge der aktuellen Perzptionsknoten in den Ankern als kognitive Karte aufgefasst werden, die sowohl geometrische Repräsentationen der perzipierten Landmarken enthält als auch symbolische Konzeptualisierungen. Zudem kann der Agent sich innerhalb dieser Karte selbstlokalisieren. Die Reichhaltigkeit der kognitiven Karte hängt von der Anzahl und Art der Landmarken ab, die der Agent wahrnehmen kann. Derzeit sind dies Häuser, Bäume und Wege. Nicht verankert werden Regionen, Pfade und Entscheidungspunkte. Regionen sind aber immer relativ zu Landmarken bzw. dem Agenten selbst spezifiziert und können so geometrisch anhand der kognitiven Karte erschlossen werden. Pfade beschreiben Trajektorien, die in der gegenwärtigen Implementierung des Geometrischen Agenten ausschließlich entlang von Wegen verlaufen und können so auch anhand der in der kognitiven Karte enthaltenen Wege berechnet werden. Entscheidungspunkte liegen auf den Pfaden und sind so auch implizit in der kognitiven Karte enthalten.

Die Menge aller Anker erfüllt die Funktionen eines räumlichen Gedächtnisses, das das Erkennen und Wiedererkennen von Landmarken und Orten ermöglicht, sowie die relative Lokalisierung von Landmarken und dem Agenten selbst.

Seit zwei Jahrzehnten wird das Problem der simultanen Selbstlokalisierung und Kartierung (SLAM) in der Robotik untersucht (siehe (Thrun, 2002) für einen Überblick). Das Ziel dabei ist, dem Roboter die Kartierung einer a-priori unbekannten Umgebung zu überlassen, während er sich durch die Umgebung bewegt. Dabei muss der Roboter den Konflikt zwischen den beiden Aufgaben Selbstlokalisierung und Kartierung lösen: Zur korrekten Kartierung muss der Roboter seine Position kennen, und um seine korrekte Position zu ermitteln, muss er die Karte kennen. Das Problem wird durch die Ungenauigkeit von Odometrie und Sensorik erschwert. Das Ergebnis solch einer Kartierung der Umgebung ist zunächst meist eine zweidimensionale Repräsentation der freien, begehbaren Fläche. Mittels eines Distanzsensors nimmt der Roboter Hindernisse war, die er in ein allozentrisches Koordinatensystem einträgt. Allerdings integrieren die wenigsten Ansätze Objekterkennung. Der Roboter nimmt lediglich wahr, dass an einem Ort $< x, y >$ ein Hindernis existiert, ohne aber zu wissen, um was für ein Objekt es sich handelt. Insbesondere ist so die Integration von dynamischen, sich bewegenden Objekten unmöglich. Auch bietet diese rein metrische Ebene keine Möglichkeit zur Kommunikation über den Raum.

Die durch die Kartierung entstehenden metrischen Karten sind zunächst einmal zu genau um als Modell für menschliche kognitive Karten zu gelten. Quantitative Angaben über Entfernungen und relative Lokalisierungen sind beim Menschen systematisch fehlerhaft (Tversky, 1993). Tversky lehnt deswegen den Begriff der kognitiven Karte ab und schlägt vor, von kognitiven Kollagen zu reden. Das Wissen über den Raum gleicht demnach eher einem Flickenteppich als einer Karte. Trotz der psychologischen Unplausibilität einer exakten metrischen Ebene in mentalen Repräsentationen des Raumes, ist eine metrische Ebene für die direkte Interaktion eines Roboters mit der physikalischen Umgebung unabdingbar. Diese quantitativen Repräsentationen des Raumes sind aber dem Robotersystem intern und eignen sich nicht zur Mensch-Roboter-Interaktion, da Menschen nicht ohne weiteres in der Lage sind, quantitative räumliche Beschreibungen auszuwerten oder zu produzieren.

Ein Ansatz, der quantitatives und qualitatives Wissen über die Umgebung berücksichtigt, ist die von Kuipers et al. vorgeschlagene *Spatial Semantic Hierachy* (SSH). Die SSH ist sowohl aus den Erkenntnissen kognitiver Karten beim Menschen motiviert und soll gleichzeitig als Framework zur Exploration und Kartierung der Umgebung durch Roboter dienen. Auf der untersten Ebene – der Kontrollebene – ist die direkte sensorimotorische Koppelung zwischen Agent und Umgebung modelliert. Die Kausalebene abstrahiert von der Kontrollebene, indem diskrete Zustandsübergänge repräsentiert

sind. Aus diesen Zustand-Aktion-Zustand-Repräsentationen wird auf einer nächsten Ebene die Topologie der Umgebung abduziert. Auf dieser Ebene gibt es eine Ontologie über Orte, Pfade und Regionen. Schließlich werden die lokalen metrischen Informationen zu einer globalen metrischen Karte zusammengefügt. Beeson et al. (2007) schlagen auf Basis der SSH mit der *Hybrid Spatial Semantic Hierachy* (HSSH) eine hierarchische Repräsentation von *Large-Scale*-Umgebungen vor, in der die Umgebung auf vier Ebenen repräsentiert ist: Zwei lokale Ebenen (metrische lokale Ebene und symbolische lokale Ebene) repräsentieren den aktuell wahrnehmbaren Ausschnitt der Umgebung in einer metrischen und einer symbolischen Form. Die globalen Ebenen (metrische globale Ebene und symbolische globale Ebene) integrieren die lokalen Repräsentationen in eine globale Repräsentation. Eine weitere durch die SSH motivierte Realisierung eines Robotersystems zum räumlichen Schließen über Innenraumumgebungen mit dem Ziel der Mensch-Roboter-Interaktion findet sich in (Zender et al., 2008). Die Anwendbarkeit perzeptueller Verankerung für die Erstellung „semantischer Karten“ demonstrieren Galindo et al. (2008) und zeigen die Vorteile des Ansatzes bei der Mensch-Roboter-Interaktion und bei der Aktionsplanung mittels hierachischer Plansystemen.

Die kognitive Karte des Geometrischen Agenten (Abbildung 4.17) besteht aus der Aggregation der erkannten Landmarken. Sie hat sowohl eine metrische Ebene (bzgl. eines Weltkoordinatensystems) als auch eine semantische Ebene. Auf einer dritten Ebene verbinden die Anker metrisches Wissen, semantisches Wissen und episodisches Wissen miteinander. Die metrische Ebene führt die gesehenen Konturen zusammen und ermöglicht geometrisches Schließen. Auf semantischer Ebene ist Weltwissen in Form einer Taxonomie spezifiziert (ein Haus ist ein Gebäude, ein Gebäude ist eine Landmarke, ein Gebäude ist kein Baum, eine Landmarke ist ein wahrnehmbares und inmobiles Objekt, ...). Im Anker wird das Wissen aus der Perzeption (das Merkmalbündel, das die Merkmale aus der Objekterkennung enthält) und aus der Instruktion (das Merkmalbündel, das die Merkmale aus der Routeninstruktion enthält) zusammengeführt. Darüber hinaus kann über den Anker die Wissenserwerbshistorie ermittelt werden, die dem Agenten Aufschluss über den zeitlichen Verlauf der Wissenszustände bezüglich der ihm bekannten Landmarken gibt. Damit sind raum-zeitliche *Reasoning-Services* möglich, die von unterschiedlichen Komponenten in der Architektur des Geometrischen Agenten angefragt werden können. Die kognitive Karte stellt eine Lösung zum unter Abschnitt 2.3 in Abbildung 2.6(a) analysierten Problem dar: Der Agent kann mit ihrer Hilfe auch perzeptuell nicht-präsente Landmarken in seine Reasoning-Prozesse einbeziehen.

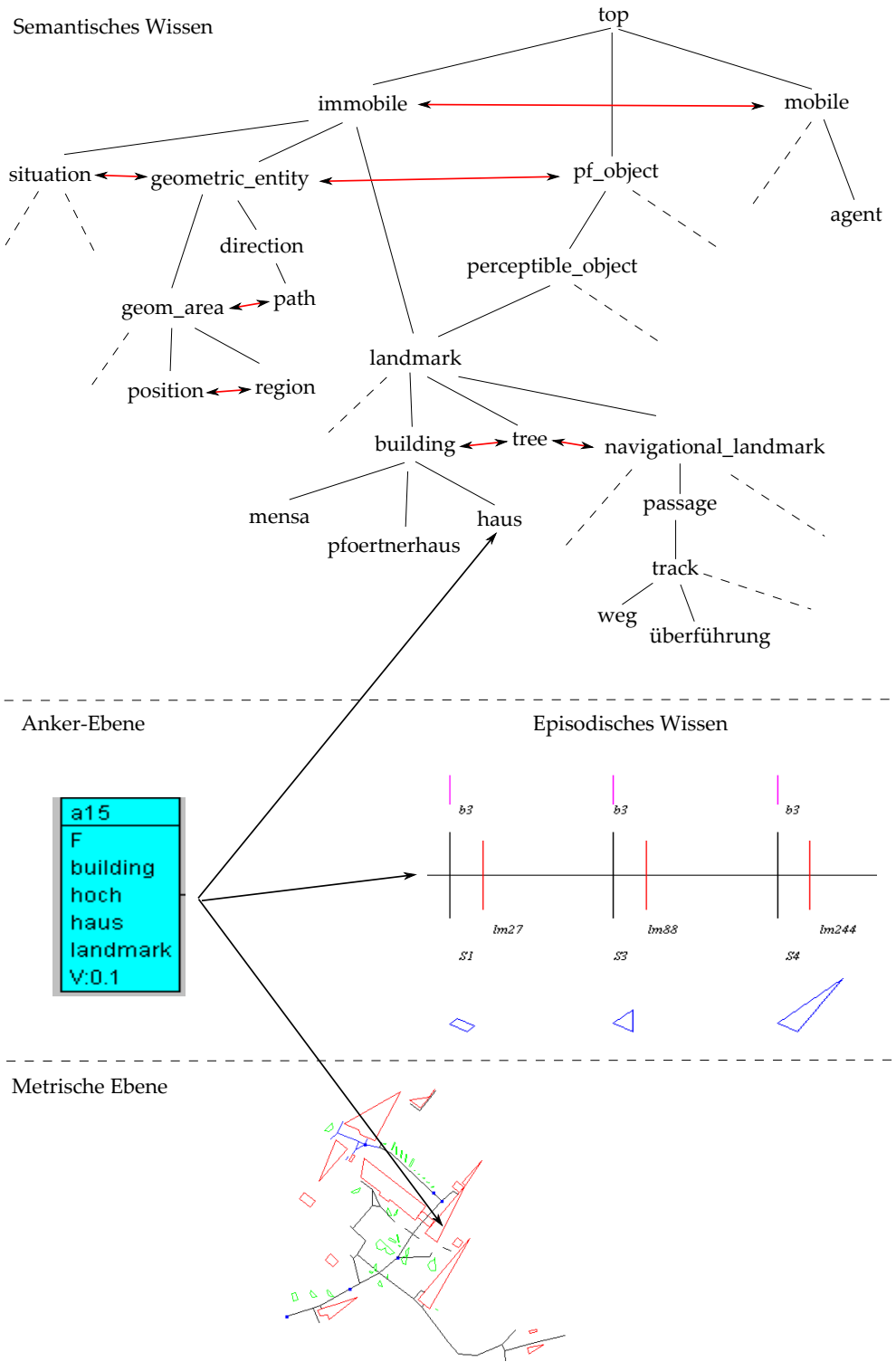


Abbildung 4.17: Die Ebenen der kognitive Karte des Geometrischen Agenten

4.5 Selektive Ankeraktivierung

Die Aggregation einzelner koreferenter Perzeptionsereignisse zu einem Perzeptionsmodell einer Landmarke hat für den Prozess der Koreferenzauflösung zwischen Perzeption und Instruktion zunächst Vorteile. Es kann in einer Situation mehrere Perzeptionsereignisse zu einer Landmarke geben, aber immer nur genau einen Anker. Mit der Anzahl der vom Agenten identifizierten und memorierten Landmarken wächst allerdings auch die Anzahl der Anker, die sowohl für die Wiedererkennung als auch für die Koreferenzauflösung zwischen Instruktion und Perzeption in Betracht gezogen werden müssen. Um auch in landmarkenreichen Umgebungen noch effizient Schließen zu können, braucht der Agent einen Mechanismus zur selektiven Aktivierung von Gedächtnisinhalten. Aus diesem Grund wird das Konzept eines Filters eingeführt, der kontextabhängig bestimmt, welche Anker in einer aktuellen Situation als mögliche Kandidaten zur Wiedererkennung und Koreferenzauflösung in Betracht gezogen werden – an welche Landmarken sich der Agent also „erinnert“.

Die beiden betrachteten Kontextparameter sind „Raum“ und „Zeit“. Der Zeitfilter wählt aus der Menge aller Anker A_{alle} diejenigen Anker $A_{aktiviert} \subseteq A_{alle}$ aus, die im Zeitraum $[s-h, s]$ aktualisiert wurden, wobei s die aktuelle Situation bezeichnet und h einen definierten Zeithorizont: $A_{aktiviert} = \{\alpha_i \mid last\_seen(a_i) \geq s-h\}$. Der Raumfilter wählt die Anker $A_{aktiviert} \subseteq A_{alle}$ aus, deren aktuelles mentale Bild nicht weiter als r von der aktuellen Pose *pose* entfernt ist oder deren jüngstes Perzeptionsereignis aus der aktuellen Situation stammt und die Landmarke zum Anker somit im perzeptuellen Fokus steht: $A_{aktiviert} = \{\alpha_i \mid d(\alpha_i^{Geom}, pose^{origin}) \leq r \vee imFokus(\alpha_i)\}$.

Untersuchungen darüber, wie die Parameter der Filter optimiert werden können, stehen aus. Denkbar ist eine vom kontextabhängige Zuweisung von Werten für h und r. In der gegenwärtigen Implementierung werden die Parameter manuell gesetzt. Auch bleibt zu prüfen, wie die beiden Filter zusammenwirken können (z. B. der Agent hat eine Landmarke länger nicht mehr gesehen, weiß aber, dass sie in der Nähe ist).

4.6 Ankerbasierte Koreferenzauswahl

Die Filterung des Ankerbestandes für die Koreferenzauflösung verringert die Anzahl der zu betrachteten Anker, verhindert aber nicht, dass der Agent unter mehreren Match-Kandidaten auswählen muss (vgl. Abschnitt 2.2). Zwischen den Kandida-

ten kann im Allgemeinen keine totale Ordnung aufgestellt werden, so dass es häufig gleichrangige Kandidaten gibt, aus denen der Agent sich für einen Kandidaten entscheiden muss. Aus der Information, die in den Ankern enthalten sind, können Kriterien zur Präferierung von Match-Kandidaten abgeleitet werden.

4.6.1 Ausschluss von Match-Kandidaten

In jeder Situation gibt es genau ein aktuellstes Perzeptionsmodell einer verankerten Landmarke. Das ist das Perzeptionsmodell, das bei der letzten Aktualisierung des Ankers erstellt wurde. Da ein Perzeptionsmodell zu genau einem Anker gehört, genau eine Landmarke Referent eines Perzeptionsmodells ist und auch ein Instruktionsknoten genau eine Landmarke repräsentiert, kann ein Instruktionsknoten nicht gleichzeitig in zwei Ankern verankert sein. Wenn also ein Anker A_n einen Instruktionsknoten I enthält, dann ist jeder Match ungültig, der eine Koreferenz zwischen einem Perzeptionsmodell eines Ankers $A_m \neq A_n$ und I enthält (Abbildung 4.18).

Abbildung 4.18: Ein Instruktionsknoten kann nur in einem Anker enthalten sein.

Bereits während der Sprachverarbeitung soll sichergestellt werden, dass koreferente Referenzen auf eine Landmarke in der Routeninstruktion zu einem einzigen Instruktionsknoten im I-Graphen zusammengefasst werden. Dann kann (mit den o.g. Voraussetzungen) in einer Situation ein Anker höchstens einen Istruktionsknoten enthalten. Wenn also ein Anker A bereits einen Instruktionsknoten I_n enthält, dann ist jeder Match ungültig, der eine Koreferenz zwischen dem aktuellen Perzeptionsmodell in A und einem Instruktionsknoten $I_m \neq I_n$ enthält (Abbildung 4.19).

Abbildung 4.19: Ein Anker kann nur einen Instruktionsknoten enthalten.

Mit diesen beiden Einschränkungen reduziert sich die Anzahl der möglichen Match-Kandidaten, unter denen sich der Agent für einen entscheiden muss. Außerdem ist damit ein Mittel geschaffen, mit dem inkonsistente Koreferenzbeziehungen vermieden werden können, in denen nicht-koreferente Perzeptionsereignisse mit einem Instruktionsknoten koreferenziert werden (vgl. Abbildung 2.7 auf Seite 15).

4.6.2 Priorisierung von Match-Kandidaten

Die vom Prozess der Koreferenzauflösung generierten Match-Kandidaten können mit Hilfe der kognitiven Karte des Agenten kontextabhängig priorisiert werden. Dafür wird das Wissen aus der Historie der Anker über den räumlichen und zeitlichen Verlauf der Route genutzt.

Abbildung 4.20 zeigt, wie die zeitliche Information genutzt werden kann: Der Agent befindet sich hier in der Situation $S2$ und soll laut aktueller Anweisung aus der Routeninstruktion ein hohes Haus sehen. Er sieht aktuell eine Landmarke, woraufhin der Anker $a2$ zu dieser Landmarke aktualisiert wird (oder – falls noch kein solcher existiert – erstellt wird). Die Landmarke passt auf die Beschreibung der Landmarke in der Routeninstruktion und somit ist das aktuelle Perzeptionsmodell von $a2$ ein Kandidat für die Koreferenz mit $b1$. Aber der Instruktionsknoten $b1$ könnte auch im Anker $a1$ verankert werden, der ebenfalls zu einem hohen Haus gehört, das der Agent in der vorherigen Situation $S1$ wahrgenommen hat. Die Koreferenzauflösung generiert zunächst beide möglichen Match-Kandidaten. Unter Verwendung der zeitlichen Priorisierung entscheidet sich der Agent für die Verankerung vom Instruktionsknoten $b1$ mit dem Anker $a2$, weil $S2$ aktueller als $S1$ ist.

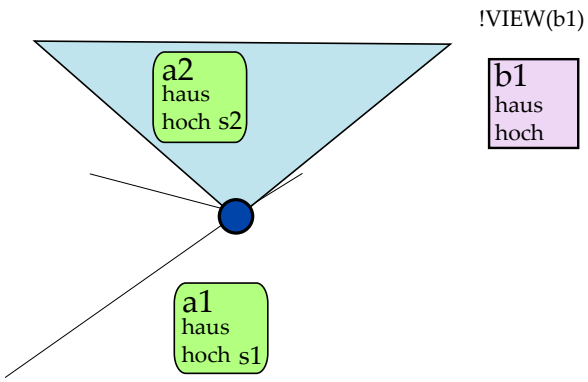


Abbildung 4.20: „Dann siehst Du ein hohes Haus“, Zeitliche Priorisierung

Auch Wissen über Entfernungen kann zur Auswahl eines Match-Kandidaten genutzt werden. Im Beispiel in Abbildung 4.21 besteht die Anweisung an den Agenten wiederum darin, ein hohes Haus zu sehen. In der aktuellen Situation sieht der Agent zwei hohe Häuser, wobei ein hohes Haus räumlich näher zum Agenten ist (verankert in Anker $a1$) als ein anderes (verankert in Anker $a2$). Die Entscheidung für die Verankerung von $b1$ in Anker $a1$ wird auf Basis des Kriteriums räumlicher Nähe der Verankerung in Anker $a2$ vorgezogen.

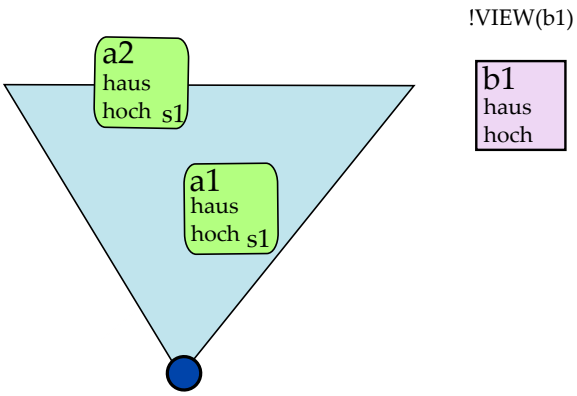


Abbildung 4.21: „Dann siehst Du ein hohes Haus“, Räumliche Priorisierung

Die Ideen hinter diesen Priorisierungen ist, dass der instruierte Agent zunächst einmal davon ausgeht, dass er erfolgreich der Route folgt (bis er eventuell explizit das Gegenteil feststellt). Darum ist es rational anzunehmen, dass die aktuell sichtbaren Landmarken eine höhere Relevanz für den Fortlauf der Route haben, als jene Landmarken, die der Agent länger nicht mehr wahrgenommen hat. Bei der Priorisierung näherer Landmarken gegenüber weiter entfernten Landmarken geht der Agent davon aus, dass in der Routeninstruktion explizit erwähnt wäre, wenn nicht die nächstgelegene Landmarke gemeint wäre, auf die die Beschreibung aus der Routeninstruktion passt (zum Beispiel durch die Phrase „das hintere hohe Haus“ statt „das hohe Haus“).

Die Entscheidungsstrategien garantieren nicht die referentielle Korrektheit der Koreferenzzuweisungen. Beide Entscheidungsstrategien könnten zu falschen Ergebnissen führen. Es ist durchaus denkbar, dass der Instruierende das Haus meinte, an dem der Agent eine Situation zuvor vorbeigelaufen ist. Ebenso ist es möglich, dass im zweiten Beispiel das weiter entfernt stehende Haus vom Instruierenden gemeint war. Es existiert keine Möglichkeit, die Korrektheit der Koreferenzen, für die sich der Agent entscheidet, formal zu beweisen. Mit dem Wissen aus den Ankern hat der Agent aber ein Mittel, um kontextabhängig möglichst plausible Entscheidungen zu treffen.

4.7 Ankerbasierte Aktionsausplanung

Navigation ist koordinierte, zielorientierte Bewegung durch eine Umgebung. Neben der Bewegungausführung (Lokomotion), schließt sie insbesondere auch die Bewegungsplanung (Wegfindung) ein (Montello, 2005). In Abschnitt 2.3 wurde das fehlende Überblickswissen des Geometrischen Agenten über seine Umgebung als ein Problem für die Wegfindung identifiziert. Im Folgen wird anhand von Beispielen demonstriert, wie die kognitive Karte des Agenten zur Aktionsplanung genutzt werden kann. Die kognitive Karte ermöglicht dem Agenten, auch Wissen über Landmarken und räumliche Konfigurationen einzubeziehen, die aktuell nicht perzeptuell präsent sind. Durch die Anker kann die Aufmerksamkeit auf bestimmte relevante Aspekte gelenkt werden. Die Bedeutung räumlichen Wissens besteht für den Geometrischen Agenten ganz konkret darin, es für die Planung von Aktionen zur Befolgung der Routeninstruktion zu nutzen. Auf Grundlage erworbenen Wissens über die Umgebung kann der Agent Hypothesen über die Lokation gesuchter Landmarken aufstellen und pragmatische Handlungsstrategien befolgen.

In Abschnitt 3.3.2 wurden Handlungsstrategien für die perzeptuelle Verankerung bei partieller Informationslage diskutiert (Seite 29). Diese Handlungsstategien zur perzeptuellen Verankerung bei unvollständiger Informationslage werden auf den Fall partieller Koreferenzierungen zwischen CRIL-Graphen angewendet. Darüber hinaus wird ein weiterer Fall behandelt, in dem ein Instruktionsknoten verankert werden kann, obwohl der Referent perzeptuell nicht präsent ist.

In der gegenwärtigen Implementierung des Aktionsmoduls im Geometrischen Agenten sind für den Fall, dass keine Koreferenz zwischen dem Instruktionsknoten, der das Argument der aktuellen Aktion ist, und einem Perzeptionsknoten besteht, Überbrückungsaktionen (Bittkowski, 2005) definiert. Der Agent dreht sich nach links und nach rechts und folgt schließlich einem Pfad. Diese Aktionen stellen die Handlungsfähigkeit des Agenten sicher. So kann er weitere Information über seine Umgebung sammeln, um doch noch die im Aktionsplan spezifizierte Aktion erfolgreich ausführen zu können. Die Überbrückungsaktionen sind uninformierte Aktionen, die solange in einer fest vorgegebenen Sequenz abgearbeitet werden, bis erfolgreich ein Match gefunden werden kann. Es gibt allerdings Situationen, in denen das Wissen aus den Ankern genutzt werden kann, um gezielte Aktionen zur Informationsbeschaffung auszuführen. Die Aktionen des Geometrischen Agenten werden entsprechend um eine weitere Ebene erweitert, um in solchen Situationen gezielte Aktionen zur Informati-

onsgewinnung auszuführen und nur im Zustand absoluter Orientierungslosigkeit auf die Überbrückungsaktionen zurückgreifen zu müssen.

4.7.1 Erweiterung der Aktion VIEW

VIEW weist dem Agenten an, nach der Landmarke lm aus der Instruktion Ausschau zu halten, die als Parameter der Aktion übergeben wurde. Die Aktion ist erfolgreich abgeschlossen, wenn eine Landmarke perzeptuell präsent ist, die hinreichend passend zur Beschreibung von lm ist. Es werden drei Fälle unterschieden:

Fall 1 Der Instruktionsknoten der gesuchten Landmarke ist erfolgreich mit einem Anker α verankert und das letzte Perzeptionsereignis zu α fand in der aktuellen Situation $s_{aktuell}$ statt. In diesem Fall befindet sich gegenwärtig eine Landmarke im Sichtfeld des Agenten, die hinreichend gut auf die Beschreibung aus der Instruktion passt – die Aktion ist erfolgreich ausgeführt.

Fall 2 Der Instruktionsknoten der gesuchten Landmarke lm ist erfolgreich mit einem Anker α verankert und das letzte Perzeptionsereignis zu α fand nicht zum Zeitpunkt $s_{aktuell}$ statt. Der Agent hat eine passende Landmarke also schon einmal in einer vergangenen Situation $s_i \prec s_{aktuell}$ wahrgenommen. In einem solchen Fall hat der Agent bereits ein mentales Bild der Landmarke und ihrer Lokation im Gedächtnis. Diese Information kann er nun nutzen, indem er einen Pfad w entlang geht, dessen Endpunkt $\mathrm{fpt}(w)$ der erinnerte Lokation l der Landmarke am nächsten ist: $\arg\min_{w \in W} d(\mathrm{fpt}(w), l) = w_{min} \rightarrow !FOLLOW(w_{min})$.[6] Diese Heuristik nimmt an, dass die gesuchte Landmarke am ehesten sichtbar wird, wenn sich der Agent ihr entgegen bewegt. Eine weitere Handlungsmöglichkeit bestünde darin, dass der Agent seine Orientierung auf die gesuchte Landmarke ausrichtet (wenn keine Information darüber existiert, dass die gesuchte Landmarke von der Pose aus, die sich aus der Drehung ergäbe, verdeckt ist).

Fall 3 Weder Fall 1 noch Fall 2 treffen zu. Der Agent ist orientierungslos und führt eine der Überbrückungsaktion aus.

[6] Bei „$\rightarrow$“ handelt es sich um einen Produktionspfeil. Die Funktion $d(x, y)$ berechnet die Distanz zwischen den geometrischen Entitäten x und y.

4.7.2 Erweiterung der Aktion GO

GO veranlasst den Agenten dazu, entlang eines Pfades w zu gehen, dessen Startpunkt $\text{spt}(w)$ der momentanen Position des Agenten entspricht und dessen Endpunkt $\text{fpt}(w)$ der Haltepunkt der Aktion bildet. Auch für die Aktion GO werden drei Fälle unterschieden.

Fall 1 Es kann eine Koreferenz zwischen dem gesuchten Pfad und einem gesehenen Pfad hergestellt werden. Der Agent geht diesen Pfad entlang.

Fall 2 Es kann keine Koreferenz zwischen dem Instruktionsknoten des gesuchten Pfades und einem gesehenen Pfad hergestellt werden, aber es existieren Instruktionsknoten, die bereits verankert sind und mit dem gesuchten Instruktionsknoten über eine relevante Relation in Beziehung stehen. Der Agent hat also bereits einen partiellen Match zwischen dem Instruktionsgraphen und dem Perzeptionsgraphen festgestellt. Beispielsweise könnte ein Weg gesucht sein, der zwischen zwei Häusern A und B führt. Haus A konnte in einer vorherigen Situation während der Agent die Instruktion „und dann siehst Du Haus A“ bereits identifiziert werden, aber das Haus B ist durch andere Landmarken verdeckt, so dass die Zwischen-Region bezüglich Haus A und Haus B nicht gebildet werden kann. Das Wissen um die Lokation von Haus A kann nun genutzt werden, um sich Haus A zu nähern und so die Chance zu erhöhen, dass auch Haus B sichtbar wird und die Zwischen-Region gebildet werden kann.

Für diesen Fall können Bedingung-Aktions-Regeln in der folgenden Form in einem regelbasierten System spezifiziert werden[7]:

$$\left(\bigwedge_{i=0}^{n} \text{rel}_i(lm_{j_i}, lm_{k_i})\right) \wedge \text{verankert}(lm_l, \alpha) \wedge \text{wählePfad}(w) \rightarrow !\text{GO}(w) \tag{4.1}$$

Dabei bildet die linke Seite der Regel die Bedingung für die Anwendung der Aktion auf der rechten Seite der Regel. Die Relationen rel_i sind Relationen zwischen Instruk-

[7] In der gegenwärtigen Implementierung befinden sich die Regeln nicht in einer Wissensbasis und es existiert kein Regelinterpreter, der die Regeln verarbeitet. Die hier aufgeführten Regeln haben exemplarischen Charakter und sind fest in den Programmtext des Aktionsmoduls eingebettet.

tionsknoten, das Prädikat verankert/2 prüft, ob ein Instruktionsknoten lm_l in einem Anker α verankert ist. wählePfad(w) wählt den gehbaren Pfad für die GO-Aktion aus, der in der Situation der beste Pfad zu sein scheint. Bei folgenden Regeln handelt es sich um beispielhafte Exemplare dieses Regeltyps:

$$\text{fpt}(p,a) \wedge \text{loc}(a,b) \wedge \text{at}(b,c) \wedge \text{verankert}(c,\alpha) \wedge \arg\min_{w \in W} d(\text{fpt}(w), \alpha^{Geom}) = w_{min} \rightarrow !\text{GO}(w_{min}) \tag{4.2}$$

Gesucht ist ein Pfad (Referent von p), dessen Endpunkt (bezeichnet durch a) in der Bei-Region – denotiert durch b – der Landmarke (c) liegen soll. Der Instruktionsknoten c ist bereits verankert. Dann wird der Pfad ausgewählt, dessen Endpunkt dem mentalen Bild α^{Geom} am nächsten ist (gibt es mehrer Pfade, deren Endpunkte die minimale Entfernung zur gesuchten Landmarke haben, wird zufällig einer unter ihnen ausgewählt). Der Pfad w_{min} wird dabei nicht mit p verankert. Er wird ausgewählt, um näher an die Landmarke heranzukommen.

$$\text{ploc}(p,a) \wedge \text{between}(a,b,c) \wedge \text{verankert}(b,\alpha) \wedge \arg\min_{w \in W} d(fpt(w), \alpha^{Geom}) = w_{min} \rightarrow !\text{GO}(w_{min}) \tag{4.3}$$

Gesucht ist ein Pfad (p), der in der Zwischen-Region (a) zweier Landmarken liegen soll, die durch die Instruktionsknoten b und c bezeichnet sind. Einer der beiden Instruktionsknoten (hier: b) konnte bereits in den Anker α verankert werden. Es wird ein Pfad ausgewählt, dessen Endpunkt dem mentalen Bild α^{Geom} am nächsten ist. Auch hier wird w_{min} nicht mit p verankert, sondern w_{min} wird gewählt, um näher an die durch b denotierte Landmarke heranzukommen. Damit ist die Hoffnung verbunden, dass der gesuchte Pfad sichtbar wird.

Ein alternativer Ansatz zu den Bedingung-Aktions-Regeln wäre eine Formulierung als Planungsproblem wie in (Karlsson et al., 2008). Dafür müsste der Ist-Zustand (inklusive der Hypothesen über noch unbeobachtete Teile der Umgebung) geeignet modelliert werden, um einen Plan zu einem Zielzustand zu generieren, in dem der Zielknoten verankert ist. Ein solcher Ansatz ist außerhalb des Fokus dieser Arbeit.

Fall 3 Weder Fall 1 noch Fall 2 treffen zu. Der Agent ist orientierungslos und führt eine der Überbrückungsaktion aus.

5 Experimentelle Evaluation

Das Modell zur perzeptuellen Verankerung ausgedehnter Landmarken im Geometrischen Agenten wird im Folgenden anhand zweier Simulationsexperimente überprüft. Im zweiten Teil des Kapitels wird eine Anbindung des Geometrischen Agenten an eine Roboterplattform skizziert. Zwei Experimente demonstrieren, dass der Geometrische Agent als Komponente zur natürlichsprachlichen Instruktion, zum räumlichen Schließen, zur perzeptuellen Verankerung und Steuerung realer Roboter dienen kann. Der dafür notwendige Weg vom Sensormesswert zum Perzeptionsknoten wird beschrieben.

5.1 Simulationsexperimente

Simulationsumgebungen ermöglichen die Verifikation von Modellen in kontrollierten Umgebungen. Im Modellbildungsprozess für kognitive Robotiksysteme liegen die Vorteile einer Simulation der Modelle in der Abstraktion von Problemstellungen, die mit der Sensorverarbeitung und der Bewegungsausführung zusammenhängen. Das in dieser Arbeit entwickelte Modell stellt zunächst Annahmen über die perzeptuellen Fähigkeiten des Agenten und über die Beschaffenheit der Welt auf:

Allozentrisches Referenzsystem Perzeptionsknoten enthalten eine geometrische Repräsentation des Gesehenen. Der Agent hat Zugriff auf die Weltkoordinaten dieser Repräsentation.

Idealisiertes Warhnehmungsmodell Die Sensorik des Agenten ist idealisiert hinsichtlich der Genauigkeit der Messungen. Es gibt kein Rauschen, keine Messwertausreißer, keine Unsicherheit bezüglich der räumlichen Lage gemessener Werte.[1] Der Agent erkennt Landmarken als Mitglied einer Klasse (Häuser, Bäume, Wege) und identifiziert ihre Merkmale (idealisierte Landmarkenerkennung).

[1] Eine Schwierigkeit von Sonarsensoren besteht beispielsweise darin, dass der Öffnungswinkel sehr weit ist (hier: 30 Grad) und so über die Position der Reflektionsquelle zunächst nur sehr unsicher geschlossen werden kann.

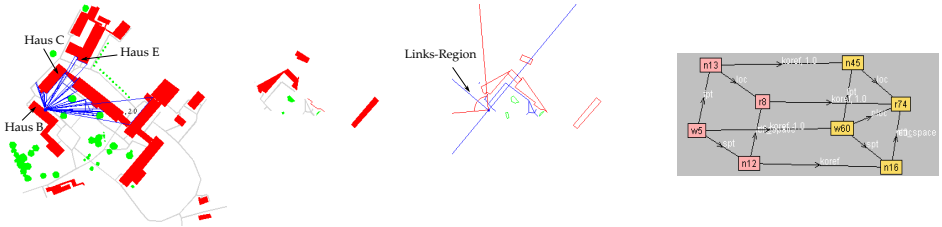
Idealisiertes Bewegungsmodell Der Agent kennt seine Pose bezüglich des Weltkoordinatensystems. Zusammen mit dem idealisierten Wahrnehmungsmodell stellt die Simulationsumgebung eine Lösung des SLAM-Problems (s.o.) bereits bereit.

Statische Welt Es gibt in der Welt des Agenten keine Objekte, die ihre Position ändern. Das bedeutet, dass alle Landmarkenkonstallationen, auf die sich der Sprecher bei der Produktion der Routeninstruktion bezieht, auch noch existieren. Es werden keine Häuser verlegt oder umgebaut, keine Bäume verpflanzt oder gefällt, und keine Wege umgeleitet oder durch Baustellen blockiert.

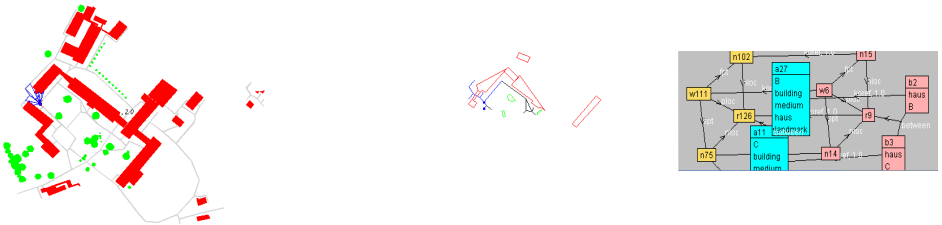
Die Ergebnisse zweier Fallstudien in der Simulationsumgebung sind in Abbildungen 5.1 und 5.3 dargestellt. Die Grafiken zeigen jeweils (von links nach rechts) i) die Übersichtskarte der Umgebung, ii) die aktuelle Perzeption des Agenten (das „Netzhautbild"), iii) die kogntive Karte und iv) die Koreferenzbeziehung zwischen Instruktion und Perzeption. Die Bedeutung der Koreferenzkanten zwischen Instruktionsknoten (lila) und Ankerknoten (blau) sind zu lesen als: „Der Instruktionsknoten ist koreferent zum aktuellen Perzeptionsmodell im Anker".

Im ersten Beispiel startet der Agent (blauer Punkt) bei der Mensa mit dem Ziel, zu Haus E zu gelangen. In der Situation, die in 5.1(d) dargestellt ist, sollte gemäß der Routeninstruktion das Ziel Haus E sichtbar sein. Das ist aber nicht der Fall. Allerdings findet der Agent unter den Ankern einen Anker, in dem der Instruktionsknoten zu Haus E verankert werden kann. Es handelt sich dabei tatsächlich um den Anker zu Haus E, das er bereits von der Mensa aus gesehen hatte (Abbildung 5.1(a)). Fall 2 der VIEW-Aktion ist gegeben. Der Agent wählt gezielt den Pfad aus, der ihn am nächsten an die erinnerte Lage von Haus E heranführt. Die zeitliche Entwicklung des Ankers zu Haus E ist in Abbildung 5.2 dargestellt.

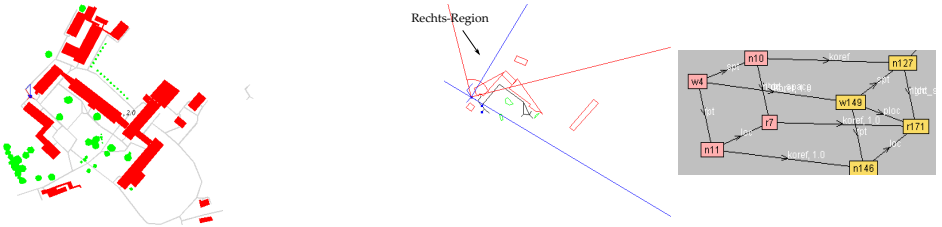
Das zweite Beispiel veranschaulicht die Anwendung der ankerbasierten Aktionen bei partieller Match-Information zwischen Instruktion und Perzeption. Die erste Anweisung in der Instruktion lässt den Agenten nach einem Haus F Ausschau halten. Unter den Ankern findet sich ein passender Anker für den Instruktionsknoten $b3$, der Haus F beschreibt – der Instruktionsknoten ist fortan über den Anker $a15$ mit einem Perzeptionsmodell zu Haus F verankert. Die folgende Instruktion ist relativ zu Haus F spezifiziert (Abbildung 5.3(b)): Der Agent soll zwischen dem soeben identifizierten Haus F und einem Haus D durchgehen. Der Weg, der zwischen Haus F und Haus D hindurchführt, ist allerdings verdeckt. Der Agent entscheidet sich zweimalig zur Anwendung von Bedingungs-Aktions-Regel 4.3 (Seite 65). Er wählt einen Pfad, der ihn



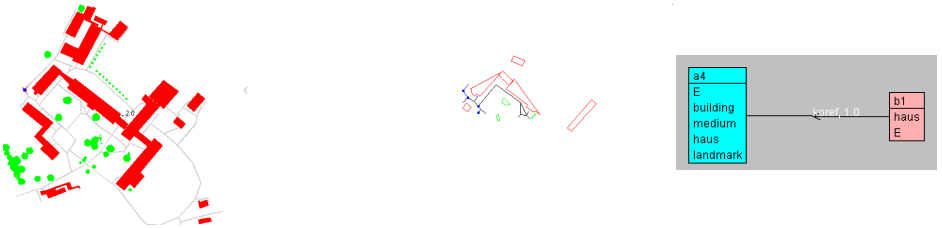
(a) „Erst gehst du links.“: Agent findet einen Pfad in die Linksregion.



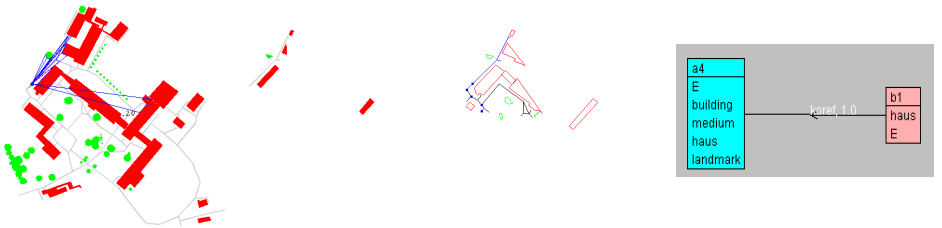
(b) „Dann gehst du zwischen Haus B und Haus C durch.“: Agent findet einen Pfad durch die Zwischenregion von Haus B und Haus C.



(c) „Dann gehst du rechts.“: Agent findet einen Pfad in die Rechtsregion.



(d) „Dann siehst du Haus E.“: Agent sieht Haus E nicht, hat Haus E aber bereits in Situation S_1 gesehen. Agent geht den Pfad, dessen Endpunkt der zuletzt erinnerten Lokalisation von Haus E am nächsten ist.



(e) „Dann siehst du Haus E.“: Agent sieht Haus E.

Abbildung 5.1: Erst gehst du links. Dann gehst du zwischen Haus B und Haus C durch. Dann gehst du rechts. Dann siehst du Haus E.

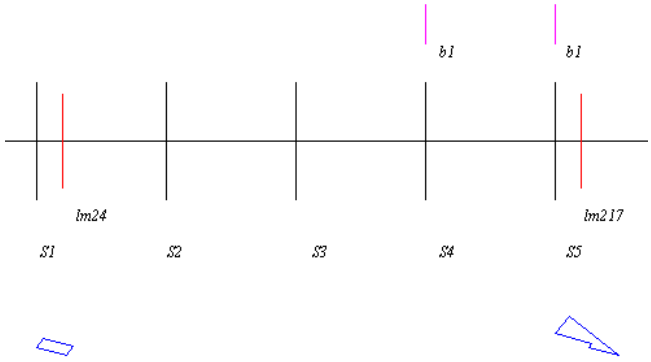
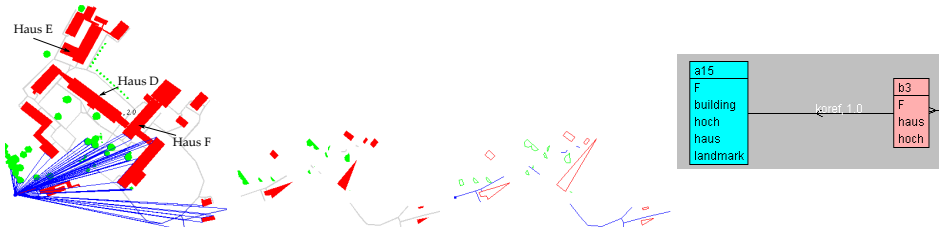
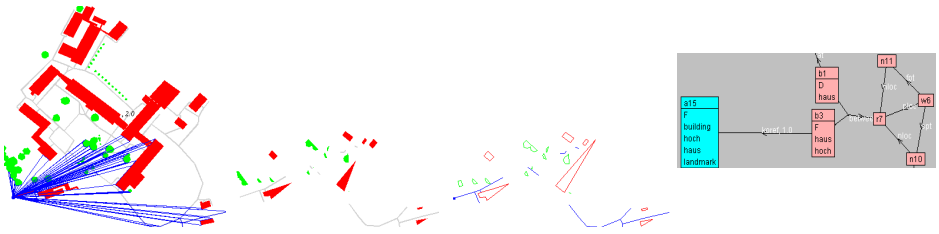


Abbildung 5.2: Der Anker zu Haus E: In der Situation S_1 wird ein Teil von Haus E erstmals wahrgenommen (Perzeptionsknoten $lm24$). Das blaue Polygon entspricht dem mentalen Bild der Landmarke Haus E zum Zeitpunkt S_1, das bis Situation S_5 unverändert bleibt. In Situation S_4 kann der Instruktionsknoten $b1$, der tatsächlich die Landmarke Haus E denotiert, dem Anker zu Haus E zugeordnet werden, ohne dass das Haus E sichtbar ist (keine Perzeptionsknoten). In Situation S_5 schließlich gerät ein anderer Teil von Haus E in das Sichtfeld des Agenten (Perzeptionsknoten $lm217$). Das mentale Bild wird entsprechend aktualisiert.

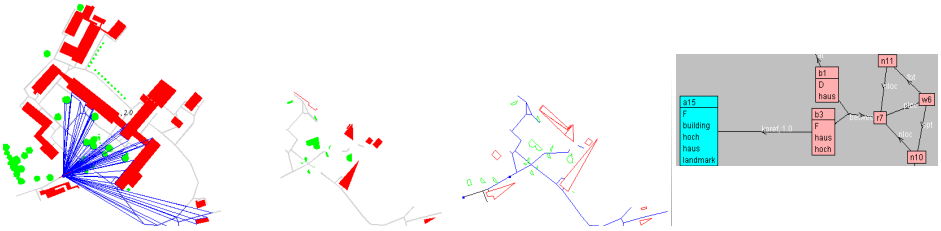
näher an die bekannte Landmarke Haus F heranbringt, bis der gesuchte Pfad identifiziert werden kann (Abbildung 5.3(d)).



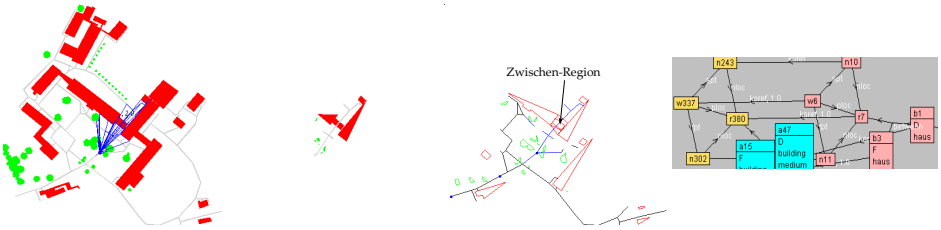
(a) „Du siehst das hohe Haus F.“: Agent sieht Haus F.



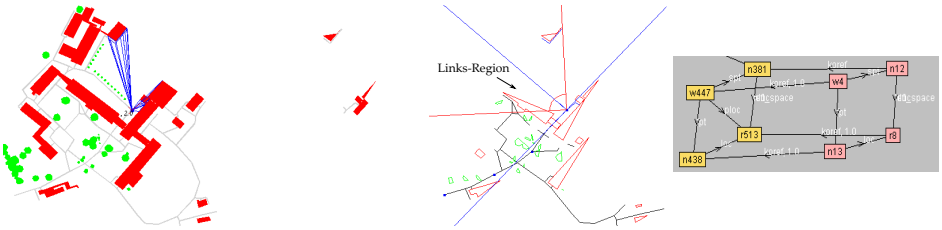
(b) „Gehe zwischen Haus F und Haus D durch.“: Agent sieht keinen Weg, der durch die Zwischenregion von Haus F und Haus D führt. Aber er kennt Haus F. Somit entscheidet er sich dazu, einen Pfad zu gehen, der ihn näher an Haus F führt.



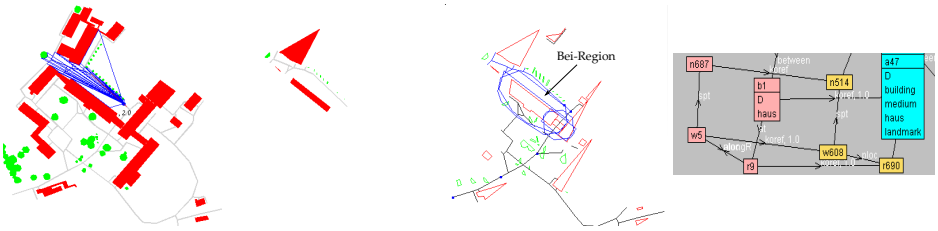
(c) „Gehe zwischen Haus F und Haus D durch.“: Agent sieht auch von hier aus noch nicht den gesuchten Weg. Agent verhält sich wie in 5.3(b)



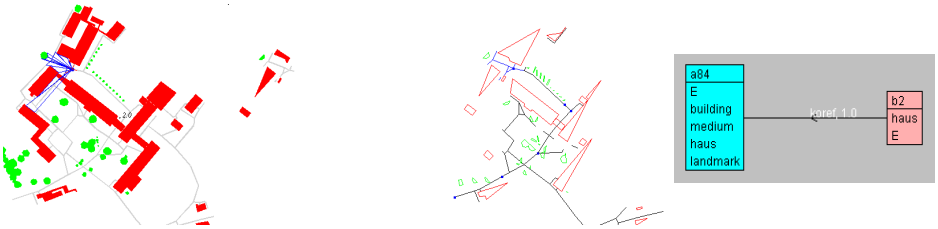
(d) „Gehe zwischen Haus F und Haus D durch.“: Agent sieht nun einen Weg zwischen Haus F und Haus D und geht den Pfad durch die Zwischenregion.



(e) „ Gehe dann nach links.“: Agent wählt einen Pfad, der ihn in die Linksregion bringt.



(f) „An Haus D gehst du entlang. . . “: Agent wählt den Pfad entlang Haus D.



(g) „. . . bis du Haus E siehst.“: Agent sieht Haus E. (Hier gibt es zwei gleichrangige Match-Kandidaten, aus denen der referentiell korrekte Match aufgrund der Priorisierung zeitlich naher Anker ausgewählt wird.)

Abbildung 5.3: Du siehst das hohe Haus F. Gehe zwischen Haus F und Haus D durch. Gehe dann nach links. An Haus D gehst du entlang bis du Haus E siehst.

5.2 Roboterexperimente

Der Geometrische Agent wurde bisher als Simulationsumgebung zur Untersuchung von Fragestellungen zur Raumkognition während der instruierten Navigation vorgestellt. Für die Evaluation des entwickelten Modells zur perzeptuellen Verankerung ausgedehnter Landmarken in der realen Welt wird das Simulationsmodul um eine Schnittstelle zu realen Robotern ergänzt.

Als Experimentierplattformen wird ein LEGO™ Mindstorms NXT eingesetzt, der mit einem Sonarsensor ausgestattet ist (Abbildung 5.4).

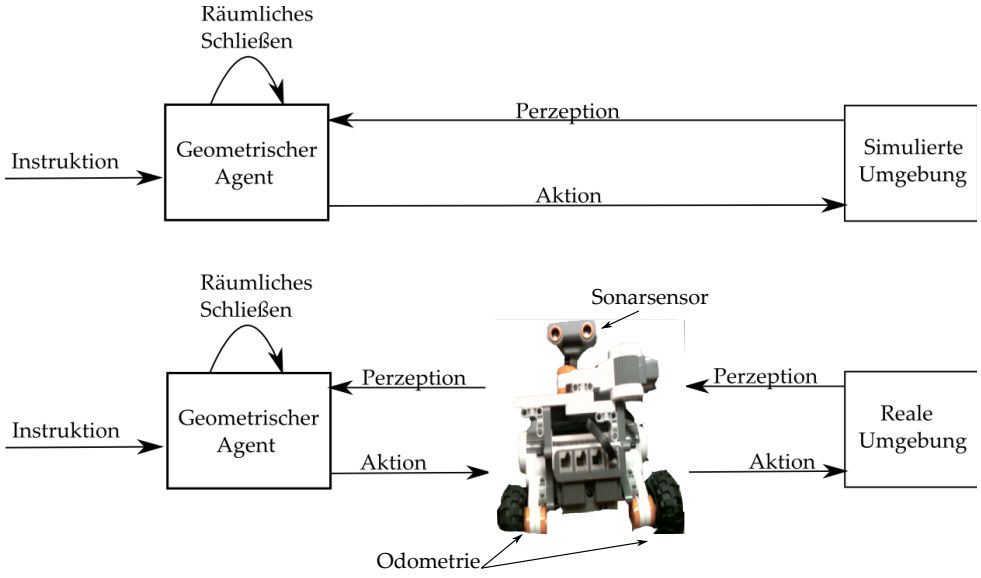


Abbildung 5.4: Simulationsumgebung durch realen Roboter ersetzen

5.2.1 Realisierung einer Schnittstelle zwischen dem Geometrischen Agenten und einem Roboter

Die softwaretechnische Realisierung der Schnittstelle zwischen dem Geometrischen Agenten und einem physikalischen Roboter ist als UML-Diagramm in Abbildung 5.5 dargestellt. Zunächst wird eine Implementierung der abstrakten Klasse `Simulation` in der Klasse `PhysicalRobot` vorgenommen. Objekte der Klasse `PhysicalRobot` delegieren die Wahrnehmungs- und Bewegungsaktionen des Geometrischen Agenten an ein Objekt einer Klasse, die das Interface `RealRobotInterface` implementiert. Das Interface umfasst Methoden für die Kommunikation zwischen einem Host-Rechner und

einem Roboter. Klassen, die das Interface implementieren, implementieren die plattformspezifische Realisierung der Sensordatenverarbeitung und Motoransteuerung einer bestimmten mobilen Plattform. Auf diese Weise sind *High-Level*-Aktionen wie „drehe Dich ein wenig gegen den Uhrzeigersinn“ mit der realen Aktorik des Roboters verankert. Diese Architektur macht es sehr einfach, weitere mobile Plattformen zu integrieren. Die einzige Voraussetzung besteht darin, dass die im `RealRobotInterface` spezifizierten Methoden implementiert werden.

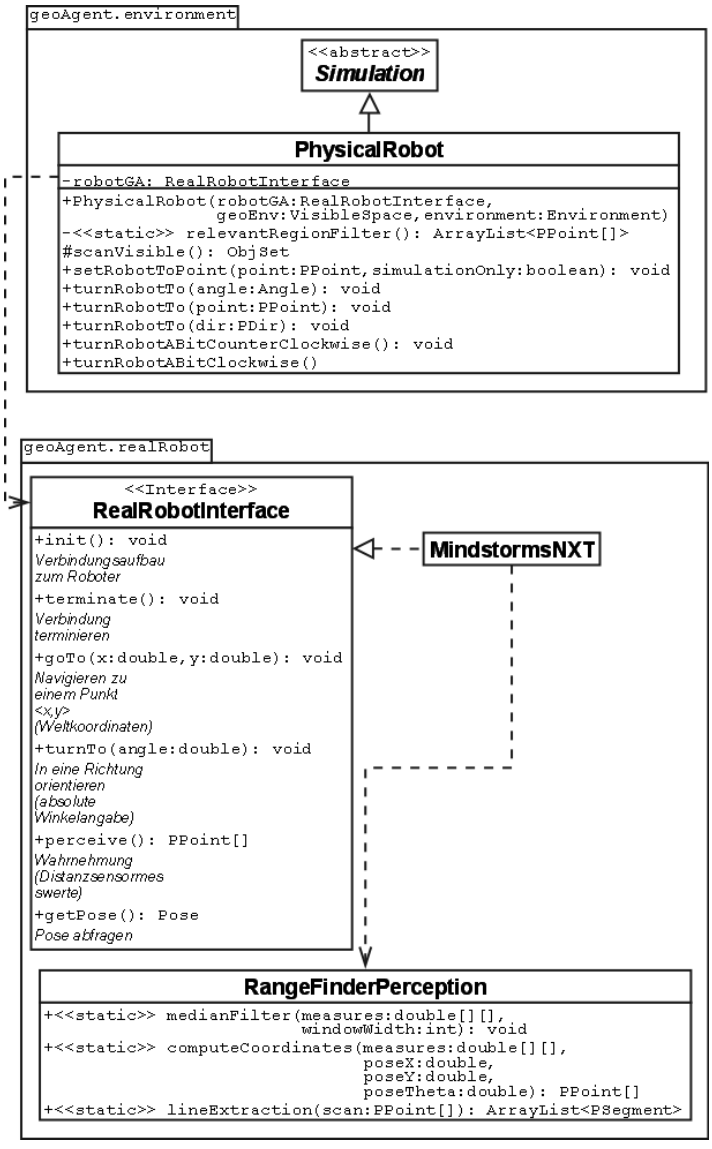


Abbildung 5.5: UML-Klassendiagramm: Anbindung des GA an einen realen Roboter

Transformation zwischen Simulationswelt und physikalischer Welt

Eine Voraussetzung für eine Evaluation des Modells in der realen Welt ist die Erkennung von Landmarken. Objekterkennung ist ein noch weitestgehend ungelöstes Problem in der Computer Vision und der Robotik im Allgemeinen und als solches nicht Gegenstand der Arbeit. Entsprechend wird die Arbeitsumgebung angepasst. Der Roboter bewegt sich in einer Welt, deren größe a-priori festgesetzt wird. Diese Welt ist eine Abbildung der Simulationswelt des Geometrischen Agenten in eine $n \times m$ cm große Welt (in den Experimenten wurde $n = m = 150$ festgelegt). Innerhalb dieser Welt kann sich der Roboter absolut selbstlokalisieren – d.h. seine Pose $(xpose, ypose, \theta)^T$ in Weltkoordinaten ermitteln. Die x-/y-Koordinaten in der simulierten Umgebung sind auf Werte zwischen 0 und 1 normiert. Entsprechend werden die Koordinaten bei der Transformation zwischen simulierter Welt und physikalischer Welt mit Faktor n (für die x-Koordinate) und m (für die y-Koordinate) multipliziert bzw. bei der Transformation von physikalischen Koordinaten auf Simulationsweltkoordinaten dividiert. Die Orientierung des Agenten wird gegebenfalls auch angeglichen, wenn das Roboterreferenzsystem sich von dem Referenzsystem des Geometrischen Agenten unterscheidet (Abbildung 5.6).

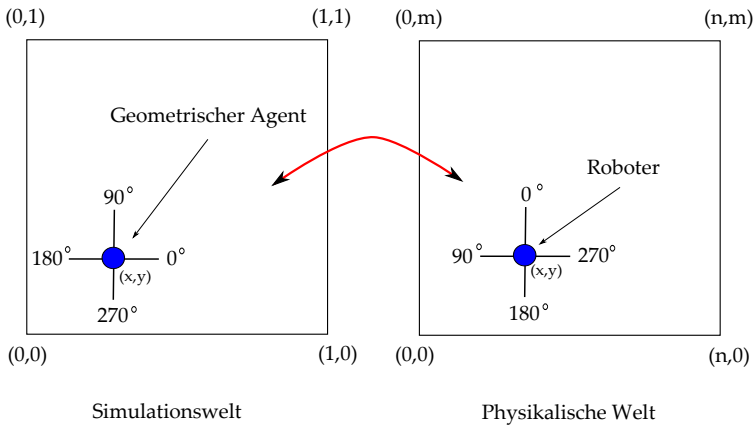


Abbildung 5.6: Transformation zwischen den Koordinatensystemen und den egozentrischen Referenzsystemen

Von Sensordaten zum Perzeptionsknoten

Die sensorische Ausstattung des Roboters besteht aus einem einzigen Sonarsensor, der die Distanz zwischen dem Roboter und dem nächstgelegenen Hindernis misst. Gemes-

sen wird um 180 Grad in 15-Grad-Schritten. Das Ergebnis eines Sonarmessungsdurchlaug ist eine Menge von Punkten, die in Polarkoodinatendarstellung (β_i, d_i) repräsentiert sind. Bei Sensormesswerten kommt es häufiger zu Ausreißern. Deswegen wird auf den Polarkoordinaten zunächst ein Medianfilter zur Filterung von Ausreißern angewendet (Formel 5.1).[2] Der Medianfilter ist robust gegen Messwertausreißern und besitzt die Eigenschaft, gemessene Kanten zu wahren. Ein Beispiel ist in Abbildung 5.7 dargestellt.

$$Y(n) = MEDIAN[d_{n-k}, \ldots, d_n, \ldots, d_{n+k}] \tag{5.1}$$

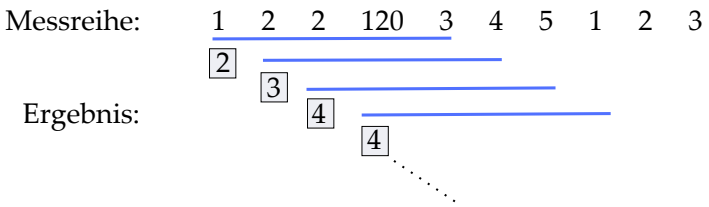


Abbildung 5.7: Medianfilter auf Sonarmessungen

Im nächsten Schritt werden die Polarkoordinaten in ihre entsprechende kartesische Koordinatendarstellung umgerechnet (vgl. Gutmann, 2000):

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} xpose \\ ypose \end{pmatrix} + \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} r_i \cos\beta_i \\ r_i \sin\beta_i \end{pmatrix} \tag{5.2}$$

Um zu ermitteln, welche Messpunkte welcher Landmarke zugeordnet werden können, werden Regionen $\{R_{HausA}, R_{HausB}, R_{HausC}, \ldots\}$ definiert, die den zu erkennenen Landmarken zugeordnet sind. Ein Punkt p gehört zu einer Region R_{lm} gdw. p in R_{lm} liegt: $p \; \iota \; R_{lm}$.[3]

Aus dem Ergebnis der beiden Filterungen werden durch Anwendung eines Split-And-Merge-Linienextraktionsverfahrens Linien durch die Punktwolke gelegt (vgl. Nguyen et al., 2007). Diese Linien bilden das Äquivalent zur geometrischen Repräsentation der gesehenen Teile von Landmarken der Simulationswelt. Zusammen mit der propositionalen Merkmalsinformationen zur perzipierten Landmarke, wird ein Perzeptionsknoten erstellt, der Repräsentation der gesehenen Kontur (das Ergebnis aus der

[2] Eine Fenstergröße von 5 ($k = 2$) erwies sich als geeignet.

[3] ι steht für den Inzidenzoperator aus der Inzidenzgeometrie. Sprich:„Der Punkt p liegt in der Region R“

Linienextraktion) enthält.

Der Ansatz vereinfacht das Landmarkenerkennungsproblem, als dass die Attribute „Objektkategorie der Landmarke“ (Haus, Baum, Weg), „Label“ und „Höhe“ nicht sensorisch erfasst werden. Dies würde einen Sensorverarbeitungsaufwand bedeuten, der über diese Arbeit hinaus geht. Was aber über die Sensorik perzipiert wird, ist die Geometrie gesehenen Landmarkenkonturen und ihre Lokation innerhalb der Welt. Wege können vom Roboter mittels Sonarsensoren nicht wahrgenommen werden. Die Wahrnehmung von Wegen ist aus diesem Grund simuliert, indem immer alle Perzeptionsknoten erstellt und in den aktuellen Perzeptionsgraphen integriert werden (Abbildung 5.8). Eine weitere Vereinfachung besteht durch die geringe Größe der Welt. In einer weitläufigen Umgebung muss der Roboter die Ungenauigkeit der Odometriemessungen ausgleichen, die in den hier gezeigten Experimenten noch nicht zu stark ins Gewicht fallen.

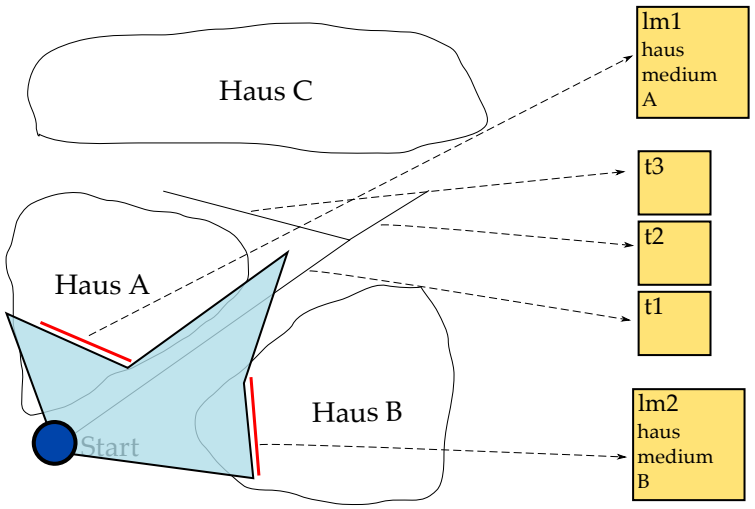


Abbildung 5.8: Perzeptionsereignisse in der angepassten Arbeitsumgebung

5.2.2 Evaluation anhand zweier Fallstudien

Zur Evaluation der Schnittstelle zwischen dem Geometrischen Agenten und dem Roboter werden zwei Instruktionen ausgeführt:

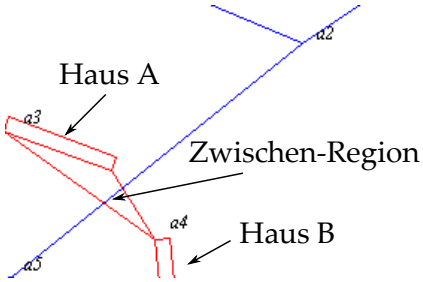
1. „Du gehst zwischen Haus A und Haus B durch. Dann siehst Du Haus C.“
2. „Du gehst hinter Haus A. Dann siehst Du Haus C.“

Beide Instruktionen können vom Roboter erfolgreich ausgeführt werden. Abbildung 5.9 zeigt den Ablauf des ersten Beispiels. Der Roboter ist zunächst in seiner Startpose auf Haus A ausgerichtet (Abbildung 5.9(a)). Die Grafik unter dem Foto zeigt den geometrischen Anteil der kognitiven Karte des Roboters nach dem ersten Perzeptionszyklus. Der Roboter hat Haus A und Haus B identifiziert und konnte die Zwischenregion erfolgreich bilden. Das dritte Bild in der Spalte zeigt die CRIL-Repräsentation der Verankerung der Instruktionsknoten mit den Ankern von Haus A und Haus B. Der Roboter wählt einen Pfad durch die Zwischen-Region. Von der neuen Pose aus (Abbildung 5.9(b)) sieht und ankert er Haus C. Die Routeninstruktion ist erfolgreich befolgt.

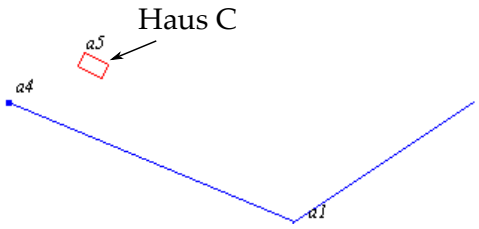
Abbildung 5.10 stellt die Ausführung von Beispiel 2 dar. Der Roboter ist zunächst in seiner Startpose auf Haus A ausgerichtet (Abbildung 5.10(a)). Nach einem ersten Perzeptionszyklus hat der Roboter Haus A identifiziert und konnte den entsprechenden Instruktionsknoten im Haus-A-Anker verankern. Auf Grundlage des mentalen Bildes von Haus A bildet der Roboter die Hinter-Region der Landmarke. Der Roboter wählt den Pfad in die Hinter-Region und kann der neuen Pose aus (Abbildung 5.10(b)) Haus C sehen.

Foto

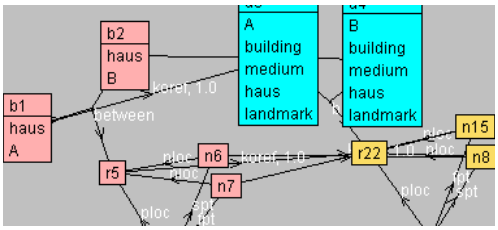
Foto



Kognitive Karte

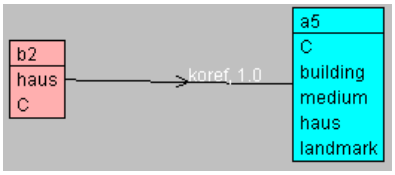


Kognitive Karte



CRIL

(a) Startpose

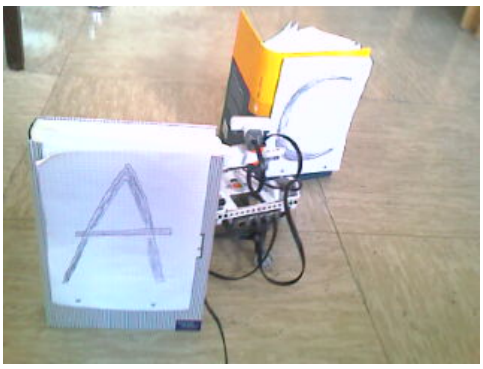


CRIL

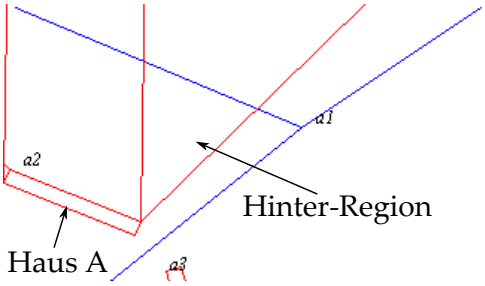
(b) Endpose

Abbildung 5.9: Roboter-Fallstudie I: „Du gehst zwischen Haus A und Haus B durch. Dann siehst Du Haus C.“

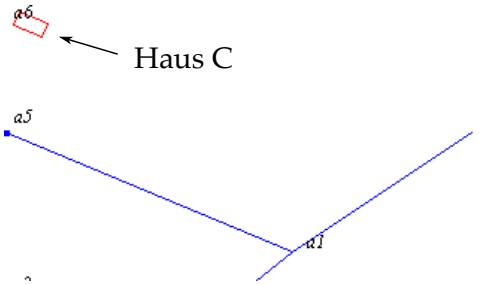
Foto



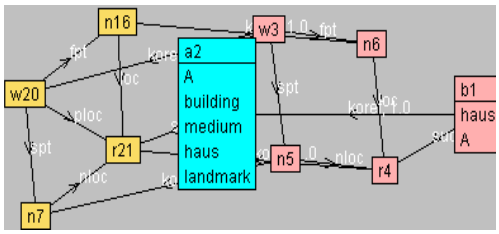
Foto



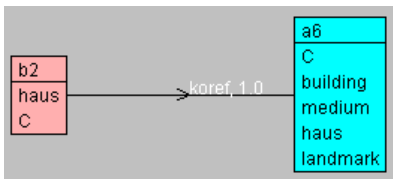
Kognitive Karte



Kognitive Karte



CRIL



CRIL

(a) Startpose

(b) Endpose

Abbildung 5.10: Roboter-Fallstudie II: „Du gehst hinter Haus A. Dann siehst Du Haus C.“

6 Schlussbetrachtung

Es konnte ein Rahmenwerk zur perzeptuellen Verankerung ausgedehnter Landmarken entwickelt werden, in dem die Lösung des Problems für den Fall der instruierten Navigation vorangetrieben werden kann. Das Rahmenwerk umfasst die Teilbereiche Wiedererkennung von Landmarken, geometrische Aggregation, Kartierung, lokale Aktionsplanung, ankergestützte Gedächtnisaktivierung und Entscheidungsfindung bei der Koreferenzauflösung. Alle Teilkomponenten werfen für sich betrachtet weitere Fragestellungen auf, aus denen aufbauende Arbeiten motiviert sein können.

6.1 Ausblick

Einige der möglichen Szenarien, in denen der Geometrische Agent von dem Verankerungsverfahren in Zukunft profitieren könnte oder Ideen zur perzeptuellen Verankerung von räumlichen Wissens werden im Folgenden skizziert.

6.1.1 Aktionen überwachen

Ein Problem des Geometrischen Agenten während der Navigationsphase ist, dass er sich des Erfolges von Aktionen nicht versichert. Ein Beispiel ist in Abbildung 6.1 dargestellt. In $s0$ nimmt der Agent eine Seite der Landmarke wahr, in deren Hinter-Region zu kommen sein Ziel sei. Der Agent erweitert die Geometrie der gesehenen Seite zu einem Polygon, das dem mentalen Bild des Agenten von der Landmarke zum Zeitpunkt $s0$ entspricht. Auf Grundlage dieses mentalen Bildes bildet der Agent die Hinter-Region. Er findet auch einen Pfad in diese Region und hakt die Anweisung, hinter der Landmarke zu stehen, als erfolgreich erledigt ab. Nun ist der Agent in der Lage, die Landmarke wiederzuerkennen, das mentale Bild durch die neuen Perzeptionsereignisse in $s1$ zu aktualisieren und festzustellen, dass er sich noch nicht hinter der Landmarke befindet. Die Schwierigkeit besteht darin, die zuvor getroffene Entscheidung über die Koreferenz der gebildeten Region mit der gesuchten Hinter-Region zu revidieren. Dann kann mit Rückgriff auf das Ankerwissen auf Basis der vorherigen

Pose und des neuen Wissens über die Geometrie der Landmarke die Hinter-Region erneut gebildet und ein Pfad in die Region gefunden werden.

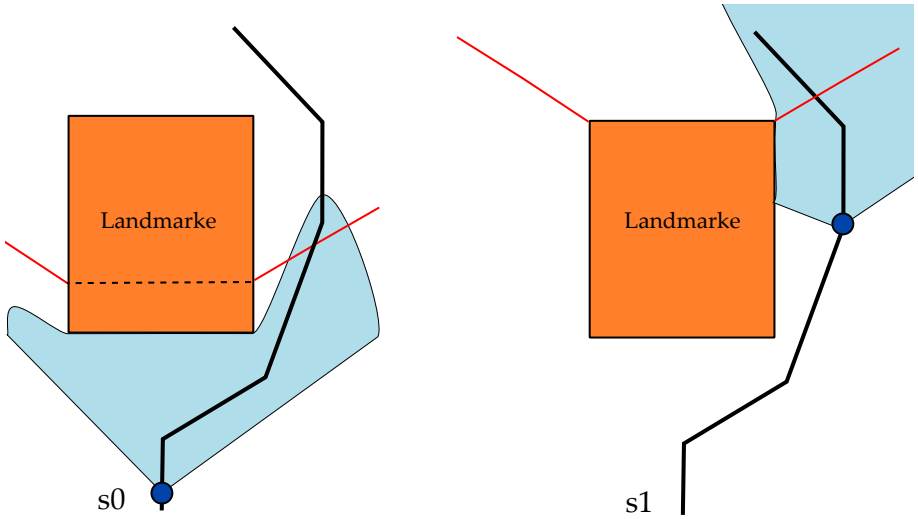


Abbildung 6.1: Monitoring von Aktionseffekten

6.1.2 Abstraktion

Durch das Verfahren der perzeptuellen Verankerung im Geometrischen Agenten erwirbt der Agent sehr schnell eine äußerst detaillierte Karte seiner Umgebung. Sein Wissen über die räumlichen Konstellationen ist sehr feinkörnig, enthält alle zeitlichen Informationen über den Ablauf seiner Route und der Lokation perzipierter Landmarkenteile. In landmarkenreichen Umgebungen kann so der Aufwand für das räumliche Schließen schnell sehr hoch werden.

Menschen besitzen die Fähigkeit, aufgabenspezifisch von irrelevanten Details zu abstrahieren. In den Kognitionswissenschaften wird vielfach argumentiert, dass ein erfolgreiches kognitives System Repräsentationen auf verschiedenen Granularitätsebenen benutzt und Schlussmechanismen sich von der Art der Aufgabe abhängigen effektivsten Ebene bedienen (Habel et al., 1995; Timpf, 1999). Insbesondere Wegfindung wurde als hierarchischer Prozess identifiziert (Timpf & Kuhn, 2003). Verschiedene Wegfindungsaufgaben verlangen verschiedene Detailstufen räumlichen Wissens. Eine erste Einschränkung dessen, was der Agent sich merkt, ist durch die Entscheidung darüber getroffen, welche Objekte und Attribute vom Verankerungsverfahren berücksichtigt werden. Darüber hinaus kann über eine Ontologie eine hierarchische Abstrak-

tion der perzeptuell erworbenen Information erreicht werden. Dies würde für den Agenten eine reichere Konzeptualisierung der Welt bedeuten und ihm somit die Möglichkeit geben, auf höherer Ebene über die Umwelt zu räsonieren und zu kommunizieren. Die bekanntesten Abstraktionmechanismen sind Generalisierung (Taxonomie) und Aggregation (Partomonie).

Ein Vorschlag zur hierarchischen Verankerung zeigt Abbildung 6.2 am Beispiel des Wissens über die räumlichen Beziehungen zwischen Teilen der Universität Hamburg. Durch die Beschreibung der partonomischen Beziehungen zwischen den verschieden granulären Landmarken gewinnt der Agent eine reichhaltigere räumliche Konzeptualisierung der Welt. Die bisher betrachteten Anker stellen die unterste Stufe der Ankerhierarchie dar. Die direkten Anker repräsentieren die perzipierbaren Anteile des Raumes (im Fall des Geometrischen Agenten sind das Häuser, Wege und Bäume). Es gibt kein Perzepte – und gleichsam keine Perzeptionsknoten –, deren Referent der Campus als solches ist (oder die Universität Hamburg). Dennoch erwirbt der Agent durch die Wahrnehmung einzelner Landmarken auf dem Informatik-Campus ein mentales Bild des Informatik-Campus selbst. Auf diese Art ist das Konzept „Informatik-Campus“ indirekt in der Perzeption des Agenten verankert.

Perzeptuelle Verankerung auf Basis einer Partonomie erfüllt grundlegende Vorteile: Der Agent konzeptualisiert die Welt anhand einer für die Mensch-Roboter Interaktion geeigneten Terminologie. Die Blätter sind derart gewählt, dass sie den perzeptuellen Kompetenzen des spezifischen Agentensystems entsprechen (also jene Teile der Welt spezifizieren, die vom Agenten erkannt werden können). Die Partonomie ermöglicht die Abstraktion dieser perzipierbaren Teile auf abstraktere Konzepte, die durch komplexe Anker (indirekt) verankert sind. Auf diesen Ebenen sind hierachische Planung, Kommunikation, Lokalisierung usw. durchführbar, wobei nur die für den gegenwärtigen Nutzen notwendigen räumliche Information betrachtet werden muss. Beispielsweise werden für die Frage, ob der Informatik-Campus nördlich vom Hauptcampus lediglich die Anker der beiden Campus aktiviert.

Einen ersten Ansatz zur perzeptuellen Verankerung komplexer Objekte wurde von Fritsch et al. (2003) verfolgt. Der komplexe Anker, der einen Menschen referenzierte, wird durch die Fusion zweier primitiver Anker gebildet, deren Referenten die Beine des Menschen (Perzept: Laserscannerdaten) respektive der Kopf des Menschen (Perzept: Kamerabild) sind.

Auch in den Geowissenschaften werden gegenwärtig auf Grundlage von Partonomien

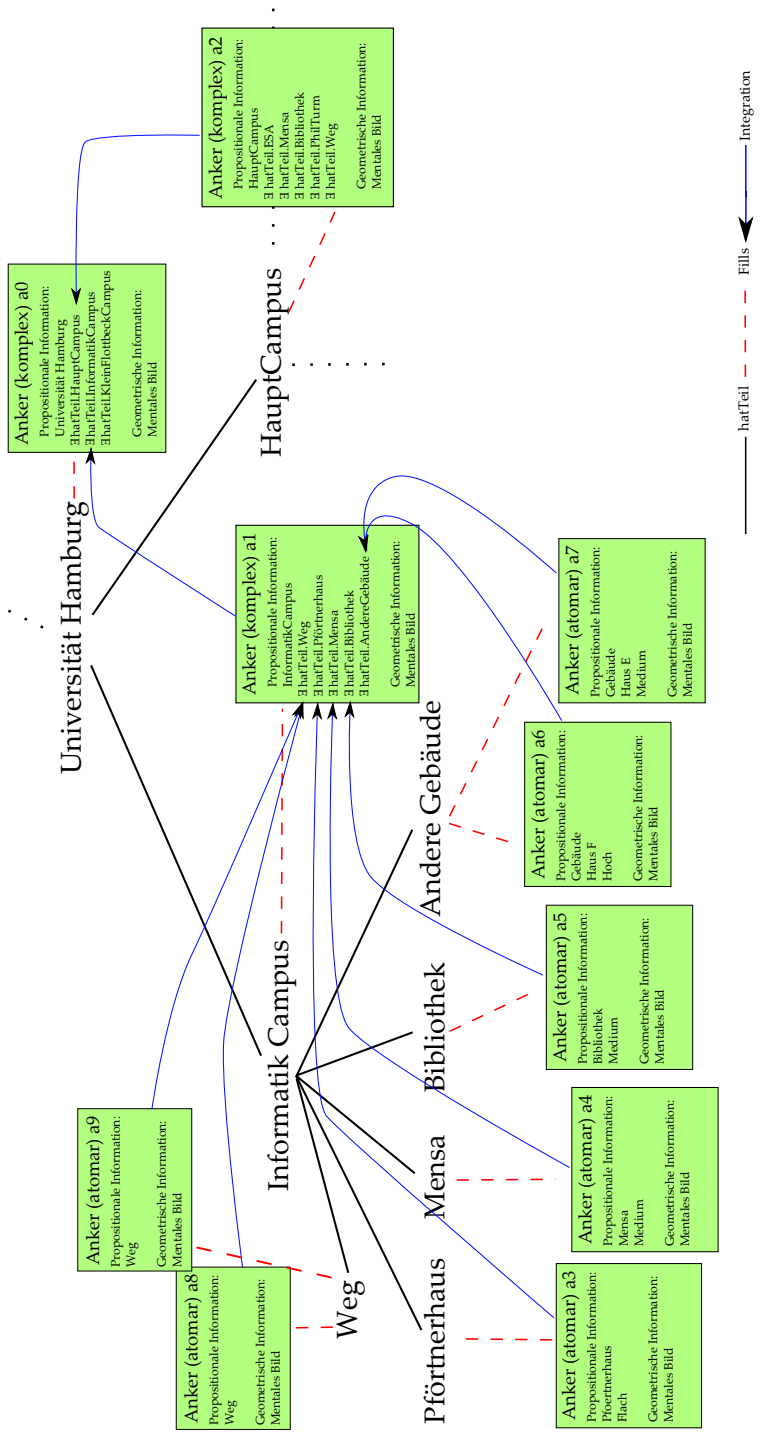


Abbildung 6.2: Hierarchische Verankerung

Modelle zur Generierung von kognitiv plausiblen Karten auf verschiedenen Granularitätsebenen untersucht, die sich mit der menschlichen Wahrnehmung der Welt decken sollen (Chaudrhy & Mackaness, 2007).

6.1.3 Lebenszyklus der Anker

Da die Anzahl der Anker und damit der Aufwand mit der Zeit wächst, wurde in der Literatur zur perzeptuellen Verankerung die Frage nach dem Ende des Lebenszyklus eines Ankers aufgeworfen (Loutfi et al., 2005). Eine grundlegende Frage auch für den Geometrischen Agenten besteht darin, wann ein Anker für irrelevant erachtet und aus der Wissensbasis gelöscht werden kann. Loutfi und Kollegen schlagen einen Wert vor, der die verbleibende Lebenszeit des Ankers repräsentiert. Die Lebenszeit fällt ab, während der Referent perzeptuell nicht präsent ist und erneuert sich, wenn der Referent wieder in den Aufmerksamkeitsfokus gerät. Die Lösung wirft aber sofort die nächste Frage nach der Funktion auf, nach der sich die Lebenszeit eines Ankers verändert; und ob sie sich für jedes Objekt gleichermaßen verändert. Ein weiterer Vorschlag besteht darin, dass Anker gelöscht werden, deren Referenten aus der Welt entfernt wurden. Der Geometrische Agent könnte (gegeben eine dynamische Welt) beispielsweise feststellen, dass eine Landmarke nicht mehr existiert (Abriss eines Hauses). Allerdings gibt es in Instruktionen auch Referenzen auf Landmarken, die nicht mehr existieren („Da, wo mal das Haus XY stand.“) (Sorrows & Hirtle, 1999).

Ein weiteres Problem besteht darin, dass die Historie eines Ankers mit jeder Wiedererkennung anwächst. Mechanismen zur Identifizierung unwichtiger episodischer Information könnten zur Löschung von Historieneinträgen genutzt werden. Beispielsweise ist die Information zweier aufeinander folgenden Situationen S_n und S_{n+1}, in denen eine Landmarke lediglich wiedererkannt wurde, redundant: In S_{n+1} ist die Information über das Perzeptionsereignis in S_n enthalten. Dies ist aber nicht der Fall, wenn ein neues Perzeptionsereignis aus einem Revisionsprozess resultiert, der ein Perzeptionsereignis von einem Anker zu einem anderen Anker transferiert. Die Verankerung eines Instruktionsknoten ist außerdem immer ein Ereignis, das nicht vergessen werden sollte.

6.2 Zusammenfassung

Der Einzug der symbolischen Künstliche Intelligenz in die Robotik bedeutet einen Fortschritt für die kognitive Leistungsfähigkeit der Roboter und für ihre Einsatzfähigkeit auch in nicht-technischen Umgebungen. Die Robotik profitiert von einer symbolischen Ebene durch die Fortschritte im Ontologiedesign, dem maschinellen Planen und von einer Vielzahl relevanter kognitionswissenschaftlicher Arbeiten aus den Problemfeldern Gedächtnis, Raumkognition und Linguistik. Eine Voraussetzung für die Integration symbolischer Künstlicher Intelligenz und Robotik ist die korrekte Verknüpfung von Symbolen und Perzepten, die dieselben physikalischen Objekte in der Welt bezeichnen. Es wurde ein Überblick über die *Symbol-Grounding*-Debatte gegeben und die Relevanz des Problems – insbesondere für die Mensch-Roboter-Interaktion – hervorgehoben.

Während der instruierten Navigation löst ein Agent das Verankerungsproblem: Er identifiziert die in der Routeninstruktion beschriebenen räumlichen Konstellationen in der realen Welt. Dabei nimmt der Agent zu verschiedenen Zeitpunkten unterschiedliche Teile der Landmarken von unterschiedlichen Perspektiven aus wahr. Unter diesem Aspekt konnte – aufbauend auf der Analyse eines bestehenden Rahmenwerks zur perzeptuellen Verankerung – eine Lösung zur perzeptuellen Verankerung ausgedehnter Landmarken gefunden werden. Es konnten zwei grundlegende Prozesse bei der perzeptuellen Verankerung von Landmarken identifiziert werden: Der erste Prozess ist verantwortlich für die Erkennung und Wiedererkennung von Landmarken. Er verläuft *bottom-up* und ordnet Wahrnehmungsereignisse mentalen landmarkenspezifischen Modellen zu. Bei diesem Prozess sind zwei unterschiedliche repräsentationelle Modalitäten beteiligt: die Merkmalsbündel, die symbolische Repräsentationen der perzipierten Merkmale einer Landmarke enthalten und analoge, geometrische Repräsentationen des gesehenen Teilstücks der Landmarke. Geometrische Zusammenhangskriterien erwiesen sich als einen Ansatz, um bei der Wiedererkennung Landmarken zu unterscheiden, die allein auf der Ebene von Merkmalsbündeln nicht unterscheidbar sind. Der zweite Prozess verläuft *top-down*. Er löst das Problem der Identifizierung einer Landmarke (oder einer beliebig komplexen räumlichen Konstellation) als koreferent zu einer Beschreibung aus der Routeninstruktion.

Das entstandene Modell zur perzeptuellen Verankerung im Geometrischen Agenten umfasst Komponenten zur Wiedererkennung von Landmarken, ein Verfahren zur inkrementellen Rekonstruktion der Landmarkengeometrien und ihrer Aggregation zu

einer kogitiven Karte der Umgebung, die metrisches, semantisches und episodisches Wissen verknüpft. Mit Filtern zur selektiven Aufmerksamkeit konnte ein Mechanismus gefunden werden, der die Performanz des Systems sicherstellen kann. Die Anker bilden eine wertvolle Informationsquelle, auf die der Agent für raumzeitliche Schlussmechanismen zurückgreift. Dies konnte für die lokale Aktionsplanung, für die relative Lokalisierung von Landmarken und für die Verankerung selbst demonstriert werden. Zu den Problemen, die eingangs unter Abschnitt 2.3 diskutiert wurden, konnten Lösungen gefunden werden. Das Modell zur perzeptuellen Verankerung ausgedehnter Landmarken erwies sich auch in Experimenten mit physikalischen Robotersystemen als erfolgreich anwendbar. Schließlich konnte gezeigt werden, dass der Geometrische Agent zur Mensch-Roboter-Interaktion und *High-Level*-Robotersteuerung eingesetzt werden kann. Innerhalb des entstandenen Rahmenwerks können sich zukünftige Arbeiten zu den zahlreichen Teilproblemen des Verankerungsproblems anschließen.

Literaturverzeichnis

Althaus, E., Mehlhorn, K., & Schirra, S. (2000). Experiments on curve reconstruction. In *Proceedings of the 2nd Workshop on Algorithm Engineering and Experiments*, (pp. 103–114).

Baader, F., & Nutt, W. (2003). Basic description logics. In F. Baader, D. Calvanese, D. McGuinness, D. Nardi, & P. F. Patel-Schneider (Hrsg.) *The Description Logic Handbook: Theory, Implementation and Application*. Cambridge University Press.

Bateman, J. A., & Farrar, S. (2005). Modelling models of robot navigation using formal spatial ontology. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, & T. Barkowsky (Hrsg.) *Spatial Cognition IV: Reasoning, Action, Interaction. International Conference Spatial Cognition 2004, Frauenchiemsee, Germany, October 2004, Proceedings*, (pp. 366–389). Berlin, Heidelberg: Springer.

Beeson, P., MacMahon, M., Modayil, J., Murarka, A., Kuipers, B., & Stankiewicz, B. (2007). Integrating multiple representations of spatial knowledge for mapping, navigation, and communication. In *Proceedings of the Symposium on Interaction Challenges for Intelligent Assistants*, AAAI Spring Symposium Series. Stanford, California.

Bittkowski, N. (2005). *Aktionsplanung und -steuerung unter Unsicherheit bei der Navigation eines Geometrischen Agenten mit Hilfe von Wegbeschreibungen*. Diplomarbeit, FB Informatik, Universität Hamburg, Hamburg.

Brenner, M., Hawes, N., Kelleher, J., & Wyatt, J. (2007). Mediating between qualitative and quantitative representations for task-orientated human-robot interaction. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*. Hyderabad, India.

Brooks, R. (1985). A robust layered control system for a mobile robot. *A. I. Memo*, *864*, 1–25.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*(1&2), 3–15.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*(1-3), 139–159.

Bugmann, G., Klein, E., Lauria, S., & Kyriacou, T. (2004). Corpus-based robotics: A route instruction example. In *Proceedings of IAS-8*, (pp. 96–103).

Carlson, L. (2000). Object use and object location: The effect of function on spatial relations. In E. van der Zee, & U. Nikanne (Hrsg.) *Cognitive Interfaces: Constraints on linking cognitive information*, (pp. 94–115). Oxford University Press.

Chalmers, D. J. (1992). Subsymbolic computation and the chinese room. In J. Dinsmore (Hg.) *The Symbolic and Connectionist Paradigms: Closing the Gap*, (pp. 25–52). Lawrence Erlbaum.

Chaudrhy, O., & Mackaness, W. A. (2007). Utilising partonomic information in the creation of hierarchical geographies. In *10th ICA Workshop on Generalisation and Multiple Representation*. Moscow, Russia.

Chella, A., Coradeschi, S., Frixione, M., & Saffiotti, A. (2004). Perceptual anchoring via conceptual spaces. In *Proceedings of the AAAI-04 Workshop on Anchoring Symbols to Sensor Data*. Menlo Park, CA: AAAI Press.

Clark, H. H., & Brennan, S. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Hrsg.) *Perspectives on Socially Shared Cognition*, (pp. 127–149). Washington DC: American Psychological Association.

Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. *Computational Linguistics*, *10*(2), 97–146.

Copleston, S. N., & Bugmann, G. (2008). Personal robot user expectations. Tech. Rep., Robotics and Intelligent Systems, University of Plymouth.
URL `http://www.tech.plym.ac.uk/soc/staff/GuidBugm/pub/UserExpectationsSurvey_CoplestonBugmann_TR0801.pdf`

Coradeschi, S., Driankov, D., Karlsson, L., & Saffiotti, A. (2001). Fuzzy anchoring. In *In The 10th IEEE Conference on Fuzzy Systems*, (pp. 111–114).

Coradeschi, S., & Loutfi, A. (2008). A review of past and future trends in perceptual anchoring. In P. Fritzche (Hg.) *Tools in Artificial Intelligence*, Kapitel 15. I-Tech Education and Publishing.

Coradeschi, S., & Saffiotti, A. (2000). Anchoring symbols to sensor data: preliminary report. In *Proceedings of the 17th AAAI Conference*, (pp. 129–135). Menlo Park, CA: AAAI Press.

Coradeschi, S., & Saffiotti, A. (2003). An introduction to the anchoring problem. *Robotics and Autonomous Systems*, *43*(2–3), 85–96.

Coventry, K. R., Prat-Sala, M., & Richards, L. V. (2001). The interplay between geometry and function in the comprehension of „over“, „under““, „above“ and „below“. *Journal of Memory and Language*, *44*, 376–398.

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233–263.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, *16*, 409–458.

Elmogy, M., Habel, C., & Zhang, J. (2008). Robot topological map generation from formal route instructions. In Y. Lespérance, G. Lakemeyer, J. Peters, & F. Pirri (Hrsg.) *Proceedings 6th International Cognitive Robotics Workshop*, (pp. 60–67).

Elmogy, M., Habel, C., & Zhang, J. (2009). Spatial language for route-based humanoid robot navigation. In *Proceedings of the 4th International Conference on Spatial Cognition (ICSC2009)*. Roma, Italy.

Engel, A. K. (2006). Neuronale Grundlagen der Merkmalsintegration. In H.-O. Karnath, & P. Thier (Hrsg.) *Neuropsychologie*, Kapitel 5, (pp. 55–65). Springer, Heidelberg.

Evans, G. (1982). *The Varieties of Reference*. Oxford University Press.

Fodor, J. A. (1980). *The Language of Thought*. Harvard University Press.

Frank, B., Stachniss, C., Schmedding, R., Burgard, W., & Teschner, M. (2009). Real-world robot navigation amongst deformable obstacles. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Kobe, Japan.

Fritsch, J., Kleinehagenbrock, M., Lang, S., Plötz, T., Fink, G. A., & Sagerer, G. (2003). Multi-modal anchoring for human-robot interaction. *Robotics and Autonomous Systems*, *43*(2-3), 133–147.

Galindo, C., Fernandezmadrigal, J., Gonzalez, J., & Saffiotti, A. (2008). Robot task planning using semantic maps. *Robotics and Autonomous Systems*, *56*(11), 955–966.

Galton, A. (2008). Pareto-optimality of cognitively preferred polygonal hulls for dot patterns. In C. Freksa, N. Newcombe, P. Gärdenfors, & S. Wölfl (Hrsg.) *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, (pp. 409–425).

Gharpure, C., & Kulyukin, V. (2008). Robot-assisted shopping for the blind: issues in spatial cognition and product selection. *Intelligent Service Robotics*, *1*(3), 237–251.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *TRENDS in Cognitive Sciences*, *5*(6), 236–243.

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. The MIT Press.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*, 199–220.

Gutmann, J. S. (2000). *Robuste Navigation autonomer mobiler Systeme*. Dissertation, Universität Freiburg.

Habel, C., Pribbenow, S., & Simmons, G. (1995). Partonomies and depictions: A hybrid approach. In *Diagrammatic Reasoning: Computational and Cognitive Perspectives*. AAAI/MIT Press.

Hähnel, D., Burgard, W., & Lakemeyer, G. (1998). Golex - bridging the gap between logic (GOLOG) and a real robot. In *KI '98: Proceedings of the 22nd Annual German Conference on Artificial Intelligence*, (pp. 165–176). London, UK: Springer-Verlag.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*, 335–346.

Harnad, S. (2003). The symbol grounding problem. In *Encyclopedia of Cognitive Science*. Nature Publishing Group/Macmillan.

Heintz, F., Rudol, P., & Doherty, P. (2007). Bridging the sense-reasoning gap using dyknow: A knowledge processing middleware framework. In *KI '07: Proceedings of the 30th annual German conference on Advances in Artificial Intelligence*, (pp. 460–463). Berlin, Heidelberg: Springer-Verlag.

Helwich, J. H. (2003). *Graphenbasierte Navigation eines Geometrischen Agenten: Integration von Perzeption und Instruktion*. Diplomarbeit, FB Informatik, Universität Hamburg, Hamburg.

Hertzberg, J., & Chatila, R. (2008). AI reasoning methods for robotics. In B. Siciliano, & O. Khatib (Hrsg.) *Springer Handbook of Robotics*, (pp. 207–223). Springer.

Hertzberg, J., & Saffiotti, A. (2008). Using semantic knowledge in robotics (editorial). *Robotics and Autonomous Systems*, *11*, 875–877.

Hudelot, C., & Thonnat, M. (2003). A cognitive vision platform for automatic recognition of natural complex objects. In *ICTAI '03: Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Hg.) *Foundations of Cognitive Science*. MIT Press.

Karlsson, L., Bouguerra, A., Broxvall, M., Coradeschi, S., & Saffiotti, A. (2008). To secure an anchor – a recovery planning approach to ambiguity in perceptual anchoring. *AI Communications*, *21*(1), 1–14.

Kellman, P. (2003). Perceptual processes that create objects from fragments. In *Proceedings of the International Joint Conference on Neural Networks*.

Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *Proceedings of the IEEE International Workshop on Robots and Human Interactive Communication (ROMAN)*.

Krämer, N. C. (2008). Theory of mind as a theoretical prerequisite to model communication with virtual humans. In I. Wachsmuth, & G. Knoblich (Hrsg.) *Modeling Communication with Robots and Virtual Humans*, vol. 4930 of *Lecture Notes in Computer Science*, (pp. 222–240). Springer.

Kuipers, B. (2008). Drinking from the firehose of experience. *Artificial Intelligence in Medicine*, *44*, 155–170.

Kuipers, B., Browning, R., Gribble, B., Hewett, M., & Remolina, E. (2000). The spatial semantic hierarchy. *Artificial Intelligence*, *119*, 191–233.

Lakoff, G. (1990). *Women, Fire, and Dangerous Things*. University Of Chicago Press.

Latecki, L. J., Lakämper, R., Sun, X., & Wolter, D. (2005). Geometric robot mapping. In E. Andres, G. Damiand, & P. Lienhardt (Hrsg.) *Proceedings of the 12th International Conference on Discrete Geometry for Computer Imagery (DGCI)*, vol. 3429 of *Lecture Notes in Computer Science*, (pp. 11–22). Poitiers, France: Springer.

Levesque, H., & Lakemeyer, G. (2008). Cognitive robotics. In F. van Harmelen, V. Lifschitz, & B. Porter (Hrsg.) *Handbook of Knowledge Representation*. Elsevier.

Levinson, S. C. (1996). Frames of reference and molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Hrsg.) *Language and Space*, Kapitel 4, (pp. 109–169). MIT-Press.

Lidoris, G., Rohrmüller, F., Wollherr, D., & Buss, M. (2009). The autonomous city explorer (ace) project - mobile robot navigation in highly populated urban environments. In *2009 IEEE International Conference on Robotics and Automation*. Kobe, Japan.

Loutfi, A., Coradeschi, S., Daoutis, M., & Melchert, J. (2008). Using knowledge representation for perceptual anchoring in a robotic system. *International Journal on Artificial Intelligence Tools*, *14*(1).

Loutfi, A., Coradeschi, S., & Saffiotti, A. (2005). Maintaining coherent perceptual information using anchoring. In *Proceedings of the 19th IJCAI Conference*. Edinburgh, UK.

Lynch, K. (1960). *The Image of the City*. The MIT Press.

Melchert, J., Coradeschi, S., & Loutfi, A. (2007). Spatial relations for perceptual anchoring. In *Proceedings of the AISB Annual Convention. Newcastle*.

Möller, M., Sintek, M., Buitelaar, P., Mukherjee, S., Zhou, X. S., & Freund, J. (2009). Scalable medical image understanding by fusing cross-modal object recognition with formal domain semantics. In H. G. A. Fred, J. Filipe (Hg.) *Best papers of BIOSTEC 2008*, Communications in Computer and Information Science, (pp. 390–401). Springer-Verlag.

Modayil, J., & Kuipers, B. (2008). The initial development of object knowledge by a learning robot. *Robotics and Autonomous Systems*, *56*(11), 879–890.

Montello, D. R. (2005). Navigation. In P. Shah, & A. Miyake (Hrsg.) *The Cambridge Handbook of Visuospatial Thinking*, (pp. 257 – 294). Cambridge University Press.

Montello, D. R., Waller, D., Hegarty, M., & Richardson, A. E. (2004). Spatial memory of real environments, virtual environments, and maps. In G. L. Allen (Hg.) *Human spatial memory: Remembering where*, (pp. 251 – 285). Mahwah, NJ: Lawrence Erlbaum Associates.

Moratz, R., & Tenbrink, T. (2006). Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial Cognition and Computation*, *6*(1), 63–106.

Moratz, R., Tenbrink, T., Bateman, J., & Fischer, K. (2002). Spatial knowledge representation for human-robot interaction. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Hrsg.) *Spatial Cognition III*, (pp. 263–286). Springer.

Newell, A., & Simon, H. A. (1975). Computer science as empirical inquiry: Symbols and search. *Communication of the ACM*, *19*(3), 113–126.

Nguyen, V., Gächter, S., Martinelli, A., Tomatis, N., & Siegwart, R. (2007). A comparison of line extraction algorithms using 2d range data for indoor mobile robotics. *Autonomous Robots*, *23*(2), 97–111.

Nilsson, N. J. (1984). Shakey the robot. Tech. Rep. 323, Artificial Intelligence Center, SRI International, Melon Park, CA.
URL `www.cs.sfu.ca/ vaughan/teaching/889/papers/shakey.pdf`

Oestreicher, L., & Severinson-Eklundh, K. (2006). User expectations on human-robot co-operation. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication: RO-Man*. Hatfield, UK.

Peuquet, D. J. (2002). *Representations of Space and Time*. New York: Guilford.

Peuquet, D. J., & Ci-Xiang, Z. (1987). An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane. *Pattern Recognition*, *20*(1), 65–74.

Premack, D., & Woodruff, G. (1978). Does a chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*, 515–526.

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Science*, *4*(5), 197–207.

Pylyshyn, Z. W. (2009). Perception, representation and the world: The finst that binds. In D. Dedrick, & L. M. Trick (Hrsg.) *Computation, Cognition and Pylyshyn*. Cambridge, MA: MIT Press.

Raubal, M. (2004). Formalizing conceptual spaces. In A. Varzi, & L. Vieu (Hrsg.) *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS*

2004), vol. 114 of *Frontiers in Artificial Intelligence and Applications*, (pp. 153–164). Amsterdam, NL: IOS Press.

Richards, L. V., Coventry, K. R., & Clibbens, J. (2004). Where's the orange? geometric and extra-geometric factors in children's talk of spatial relations. *Journal of Child Language*, *31*, 153–175.

Roy, D. (2005). Semiotic schemas: a framework for grounding language in action and perception. *Artificial Intelligence*, *167*(1-2), 170–205.

Santore, J. (2005). *Identifying Perceptually Indistinguishable Objects*. Dissertation, Department of Computer Science and Engineering, University at Buffalo, The State University of New York.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, *12*, 13–24.

Schmidtke, H. R. (2001). The house is north of the river: Relative localization of extended objects. In *COSIT 2001: Proceedings of the International Conference on Spatial Information Theory*, (pp. 415–430). London, UK: Springer-Verlag.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417–457.

Shapiro, S. C., & Ismail, H. O. (2003). Anchoring in a grounded layered architecture with integrated reasoning. *Robotics and Autonomous Systems*, *43*(2-3), 97–108.

Shi, H., & Tenbrink, T. (2009). Telling rolland where to go: HRI dialogues on route navigation. In J. B. Kenny R. Conventry, Thora Tenbrink (Hg.) *Spatial Language and Dialogue (Explorations in Language and Space)*, (pp. 177–216). Oxford University Press.

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior And Development*, *13*, 33–49.

Sloman, A. (2009). Some requirements for human-like robots: Why the recent overemphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Hrsg.) *Creating Brain-like Intelligence*, (pp. 248–277). Berlin: Springer-Verlag.

Sorrows, M. E., & Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces. In *COSIT '99: Proceedings of the International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, (pp. 37–50). London, UK: Springer-Verlag.

Taddeo, M., & Floridi, L. (2005). Solving the symbol grounding problem: a critical review of fifteen years of research. *Journal of Experimental & Theoretical Artificial Intelligence*, *17*(4), 419–445.

Thrun, S. (2002). Robotic mapping: A survey. In G. Lakemeyer, & B. Nebel (Hrsg.) *Exploring Artificial Intelligence in the New Millennium*. Morgan Kaufmann.

Timpf, S. (1999). Abstraction, levels of detail, and hierarchies in map series. In C. Freksa, & D. M. Mark (Hrsg.) *COSIT*, vol. 1661 of *Lecture Notes in Computer Science*, (pp. 125–139). Springer.

Timpf, S., & Kuhn, W. (2003). Granularity transformations in wayfinding. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Hrsg.) *Spatial Cognition III*, vol. 2685 of *Lecture Notes in Computer Science*, (pp. 77–88). Springer.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208.

Tschander, L., Schmidtke, H. R., Eschenbach, C., Habel, C., & Kulik, L. (2003). A geometric agent following route instructions. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Hrsg.) *Spatial Cognition III*, vol. 2685 of *Lecture Notes in Computer Science*, (pp. 89–111). Springer.

Tulving, E. (1993). What is Episodic Memory? *Current Directions in Psychological Science*, *3*, 67–70.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank, & I. Campari (Hrsg.) *Spatial Information Theory: A Theoretical Basis for GIS*, (pp. 14–24). Berlin: Springer Verlag.

Wachsmuth, I. (2008). 'I, max' communicating with an artificial agent. In *Modeling Communication with Robots and Virtual Humans*. Springer.

Weser, M., & Zhang, J. (2007). Proactive multimodal perception for feature based anchoring of complex objects. In *Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, (pp. 1075–1080). Sanya, China.

Wittgenstein, L. (1953). *Philosophische Untersuchungen*.

Wunderlich, D., & Reinelt, R. (1982). How to get there from here. In R. J. Jarvella, & W. Klein (Hrsg.) *Speech, Place, and Action*, (pp. 183–201).

Zender, H., Kruijff, G.-J. M., & Kruijff-Korbayová, I. (2009). A situated context model for resolution and generation of referring expressions. In *12th European Workshop on Natural Language Generation (ENLG 2009)*. Athens, Greece.

Zender, H., Mozos, O. M., Jensfelt, P., Kruijff, G.-J. M., & Burgard, W. (2008). Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, *56*(6), 493–502.

Eidesstattliche Erklärung

Ich versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Departments Informatik einverstanden.

Buxtehude, 10. Juli 2009

_____

Felix Lindner