Some Aspects of Scouting Smart Environments

Jörg Baus, Andreas Butz, Antonio Krüger, Marco Lohse, Cyrille Breihof
Collaborative Research Center 378
University of Saarbrücken
PO-Box 151150
66041 Saarbrücken
Germany
[baus,butz,krueger,mlohse,breihof]@cs.uni-sb.de
http://w5.cs.uni-sb.de/~[baus,butz,krüger,mlohse,breihof]

Abstract

We discuss a virtual presenter in the form of a 3D avatar whose task it is to show people a way through a (3D model of a) public building. Our thesis is that when looking for way descriptions, seeing such a walkthrough guided by a virtual scout is easier to remember than a purely textual description of the way or the kind of annotated floor plans used widely in today's public buildings. Furthermore, we discuss some technical aspects of how such a scout can be constructed and what underlying data and processes are needed for the automated generation of guided way descriptions. One of our main goals is to give the avatar the possibility to react to her spatial environment in an appropriate matter, e.g. pointing to relevant objects and following a path that can be easily memorized by the viewer. The walkthroughs described here will be used within a building navigation system that also includes navigation via a handheld device, (such as a PDA,) once the user has left the information booth. Both the 3D visualization at the info booth and the more sketch-like presentation on the PDA are generated from the same data and by the same system, and thus can refer to each other for a much more consistent overall appearance.

Introduction

Today's your lucky day. You have just arrived in a city you've never been to before at an airport you don't know. You want to catch your connecting flight and buy some international newspapers on the way. If you're lucky, you end up standing in front of a huge environment map mounted to the wall, where your actual position is marked as a red spot. Now you have to find a newspaper store and your destination gate on the map, either by "brute force" searching over the map or by looking them up alphabetically in the legend. From the legend you will get hints, such as "A1" and "B5", which are grid cells on the map, where you have to look further for your destination.

Meanwhile time doesn't stand still and you have to catch that connecting flight. OK, back to the map! The newspaper store and gate are found. Now you have to search for a way from the red spot to those points. After you have found it, memorize it! You can use different techniques: Start looking for landmarks, salient objects in the environment on the way, especially at decision points where you have to turn, memorizing the sequence of turns, ... For scientists, this is "building a cognitive map" of the environment, but don't forget that flight you ought to catch! Hurry up a little!

As you haste along the hallways, your slowly fading cognitive map helps you to orient yourself in the environment. You finally manage to get your newspaper and reach your destination gate just before boarding is closed. If you were not so lucky, you might have forgotten which way to turn after leaving the store and you'll have to ask someone for advice or start looking for a map again. A peek at your wristwatch doesn't help either...

Enough of this! Now let's take a step into a parallel universe, where after leaving the plane, you walk right up to an information booth and pull out your electronic organizer. Its infrared port connects to the booth and within a second a gentle voice says: "Hello Mrs. Smith, I see you want to catch your connecting flight on gate F-17." (The organizer knew this from an infrared port in your airline's onboard multimedia system, in case you wonder...) "That'll probably take you five minutes, but you have twenty minutes left. Is there something else I can do for you along your way?" From the shopping menu just popped up on the screen you select "newspapers" and the voice says: "OK, I'll show you the shortest way. Just follow me!"

Next thing you see on the booth is a 3D view of the hallway you're actually standing in, matching almost exactly what you're seeing when you look up. Your scout starts walking along the hallway and as your eyes follow her, the route she's taking gives you a direct visual experience of the route you're supposed to take yourself. The camera might cut back and forth between an overview position and the scout's view. On the way your scout will turn her head and, pointing to her left, will say: "Here's your newspaper store, and, by the way,

restrooms are just next door...", whereupon a door in the 3D view is highlighted. After she has reached the connecting gate and shortly briefed you on the further use of your organizer, you start your own walk to the store. After choosing your favorite newspaper and a quick excursion to the restrooms, of course you have forgotten which way to turn. Those huge airports are still confusing you, but with a quick look at your organizer you're back on the track and assured that you're well on time. As you reach the gate, boarding is just about to start...

Please bear with us for this lengthy introduction, but we believe that it summarizes many of the aspects relevant to the successful design of building information and navigation systems. In this paper we will focus on how to construct a virtual scout and a fitting environment in order to generate guided walkthroughs in the aforementioned way.

The REAL stuff

In the REAL project we investigate the development of a resource adaptive navigational help system with the capacity to adapt to various restricted resources (cognitive or technical) during the generation of multimodal route descriptions. The resource adaptive generation of graphics/animation for different classes of output devices implies dealing with limited technical resources on one hand and taking into account a user's limited cognitive resources to decode and understand the presentation on the other. The upper end of the scale is represented by a 3D walkthrough guided by a virtual scout accompanied by spoken or written text and meta graphics complementing each other, while the lower end is represented by simple sketches of the environment and arrows indicating the direction.

For this paper we'd like to concentrate on the generation of the 3D visualization. It combines an animation of a virtual presenter with accompanying text and meta-graphics in order to describe the route itself and relevant parts of the environment at the same time (see Figure1). You can watch this presentation from different viewpoints, zoom out to get an overview or zoom in to continue the presentation from the virtual presenter's view of the scene. In case of time pressure, the presentation will speed up and the system will reduce the amount of details presented in the virtual walk-through. This can be achieved for example by abstraction techniques (Krüger 1999) that generate visually clearer and simpler route sketches.

This also helps to avoid giving a delayed passenger very long and detailed directions which are very hard to memorize and to follow correctly in the remaining time. In addition to stationary information booths we are also considering information presentation for mobile systems, such as PDAs with limited technical resources. At the information booth you will obtain the display application for your PDA over an infrared connection, make choices where you want to be guided, or what kind of information you are additionally interested in.

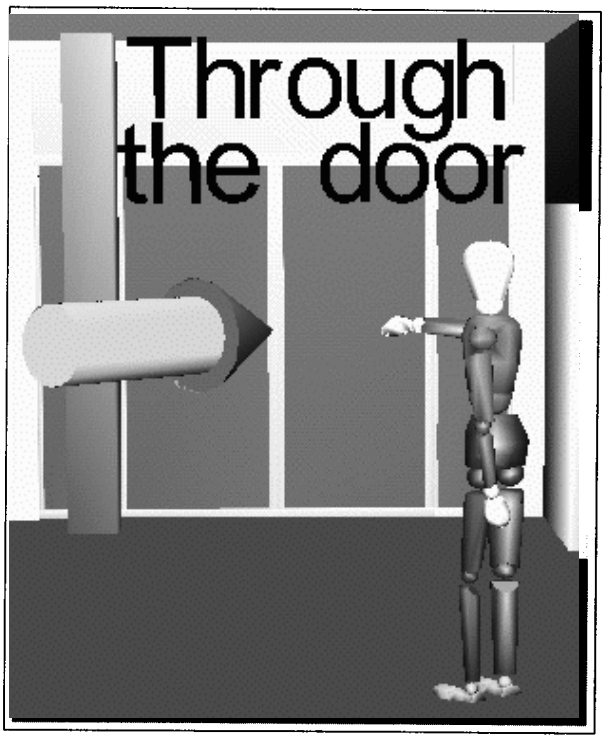

Figure 1: A snapshot of a virtual walkthrough

The presentation has to be tailored to these devices' limited display capacities by generating route descriptions in a simple sketch-like form. After leaving the booth the PDA filters the corresponding information out of a data stream broadcast by infrared transmitters spread throughout the building. The underlying protocol we developed guarantees a fast availability of simple or abstract information, while more details are accumulated as the user stays within the range of one transmitter (Baus, Butz, & Krüger 1999). A passenger in a hurry might only see an arrow pointing in the direction of her destination, while someone pausing at a place will soon have an environment map available with additional information about shops or restrooms.

The construction of a presentation involves at least the following steps: determination of new information, determination of a presentation structure, conveying of the information/presentation. In addition, the information has to be presented taking the maximum benefit of particular strengths of each presentation medium, taking into account the information to be presented as well as the current environment/context. We'd like to focus especially on

- limited resources on both sides of the communication channel, that are technical limitations of the sender (computer) and cognitive limitations of the receiver (user).
- variable resource restrictions that are unknown in ad-

vance and are coming up during the presentation.

In this paper we discuss the technical framework that enables us to deal with both of these issues. Instead of planning all actions of the virtual scout down to every detail, only an abstract plan of the presentation is prepared (e.g. reflecting the path and the resource limitations that are known in advance). The final presentation is then derived by a kind of simulation of an incremental route description, performed by the virtual presenter as she's moving along her path scouting the surroundings. For this purpose we suggest to represent the relevant task knowledge as annotations to the objects of the domain. The presenter is made sensitive to the smart environment around her, thus getting only the relevant information depending on her task and position.

In the following sections we will concentrate on the relevant knowledge in the scenario at hand and how this knowledge can be incorporated into the environment.

Building smart environments

In the given visualizations we have to cope with various time constraints and large and complex environments. Can we represent the environment in a way in which well designed objects and spaces can help us in performing the task (comparable to our real world)?

> Even human beings rely upon the real world to tell them what to do and why; well-designed objects and spaces can help us in performing daily tasks by containing clues to their meanings and operations. (Doyle & Hayes-Roth 1997).

We expect to find clues that give us the information we need as we need it. Instantly, they make us limited experts in the domain. In a properly designed world knowledge is immediately retrievable, and it's only necessary to interpret the world. A good design facilitates the immediate understanding, easing the user's cognitive load. In order to design such a world in the virtual we will show how to represent and establish reference frames that enable us to talk about the space the avatar is in. Furthermore we will discuss the general properties of objects in the domain, including buildings and street networks as these play an important role in route descriptions. Finally we will explain in more detail how the avatar presents route descriptions and information about her environment.

Reference frames in route descriptions

When talking about space we always need to establish a reference frame. A reference frame in our computational model is specified by three characteristics:

1. The origin of the coordinate system (independent of the kind of coordinate system).
2. The handedness of the coordinate system specifying the relations between axes.

Throughout this paper we will use the following terms, when we talk about different frames of reference (*FOR*) (for a more detailed discussion see (Klatzky 1998)):

- *allocentric/extrinsic*: A FOR with a fixed coordinate system imposed by external factors such as a geographic space.
- *egocentric*: This is the agent's FOR. The location of points is defined with respect to an intrinsic axis of oneself's (or one's avatar's) orientation and physical configuration. When we move, this is the FOR we take with us in every step and its allocentric location changes constantly.
- *intrinsic*: In this FOR inherent properties of the reference object (topology, size, shape) determine the coordinate system. Buildings (or building sections, such as shops or offices) have an intrinsic or prominent front, usually where the entrance is. The agent's egocentric FOR is a special case of an intrinsic FOR, where the reference object is the agent. In addition, assuming the agent is in front of a building, the building's intrinsic FOR can be used as an allocentric one with its own reference direction.

Route descriptions should help the addressee in constructing a cognitive map of the environment. They describe the way to follow including landmarks, regions, and mention spatial relations between objects in the current environment. Verbalization is given step by step and the addressee undertakes a virtual journey through the environment (see (Schmauks 1998; Maass & Schmauks 1998)).

In our case we'd like to visualize the virtual journey, i.e. a course of motion in a given environment. In the visualization we focus on this course of motion from a mostly egocentric frame of reference. Elements in the environment are localized relative to the agent's position or relative to each other from a point of view called route perspective (Tversky 1993). This kind of perspective is helpful to convey route-knowledge, knowledge about path segments and landmarks. For the transfer of survey knowledge, information about regions or the structure of the environment e.g., in order to help the addressee taking a short cut, often another point of view is chosen. Elements of the scene are referred to in an allocentric FOR corresponding to a survey perspective of the environment (see (Tversky 1993; Herrmann *et al.* 1997)). In this case we have to choose a suitable viewpoint to look at the scene, for example a "birds-eye" or "top-down" view on a map. In addition to the animation of a virtual walkthrough, we want to animate the virtual presenter, as a guide to help users in understanding and navigating the environment. A virtual human presenter is largely enriching the repertoire of available presentation options which can be used to effectively communicate information (André & Rist 1996). The presenter is used to guide the user through the presentation, attract the user's attention and convey additional conversational and emotional signals. For the description of a software agent able

to generate multimodal incremental route descriptions without a virtual presenter and the ability to choose different viewpoint/camera positions see (Maass 1996; Maass, Baus, & Paul 1995)

General object properties

Our representation of the environment consists of a hierarchy of objects. Objects can be divided into two groups: mobile and static. An object representation consists of three parts: geometry, appearance and annotations. At the very least an object must have a geometry, an appearance and one annotation, which is its unique name. The geometry and appearance representation is object centered and hierarchical. By hierarchical we mean that objects can be decomposed into parts and subparts. Geometry and appearance will be used to visualize the environment. Annotations are the clues that "give us the information we need as we need it". An object might have as many annotations as we need to provide meaningful content in different tasks. The list of annotations may contain:

- The object's location in world coordinates, an allocentric frame of reference
- References to the object's subparts, given by the subobjects' unique names
- Linguistic/textual information, useful for verbal descriptions of the objects, such as: "the white house"
- Information on how to depict the object in 2D graphics, e.g. an iconic representation of the entity for the visualization of those entities in a 2D graphics.
- Information how to highlight the object
- Information about the different coordinate systems attached to the object. Objects can have different coordinate systems suitable for different localization tasks, such as: "stand in front of the book shelf. The book is located in the upper left corner."
- functional annotations, e.g. functions to compute different coordinates from the object's 3D-model, e.g. center, 2D bounding box, 3D bounding box. These concepts can be used in the computation of different spatial relations (for a closer look see (Gapp 1994)). In fact it could be any query to a knowledge or database you need in your application in order to provide meaningful content.
- *Scheduling bounds.* Scheduling bounds specify a spatial boundary in which a behavior can take place. The region within this boundary is called a *scheduling region.* A behavior is not active unless the *activation volume* (the scheduling bound of the avatar) intersects with the behavior's scheduling region. If there is an intersection, the behavior becomes "alive" or enabled.

An enabled behavior can also receive stimuli and respond to those stimuli performing certain actions. Stimuli are used to schedule/trigger any kind of action described in the behavior as long as the behavior is enabled. Behaviors are useful for information retrieval, interaction and animation. They provide a link between a stimulus and an action described by the behavior. Furthermore, combinations of stimuli and combinations of actions are possible. One scheduling bound can trigger different behaviors or different scheduling bounds can schedule one behavior. Let's look at a behavior that only needs to be scheduled if the viewer is nearby. It should be enabled if the avatar's activation region intersects with the behavior's scheduling bounds. On the other hand behavior bounding enables costly behaviors to be skipped if the viewer isn't nearby. In other words, if there is no one in the forest to see a tree falling, it does not need to fall.

Representing street networks

Following (Tversky & Lee 1998) a path of motion can be divided into segments. Each segment consists of four parts belonging to different categories: starting point, reorientation(direction), path/progression and ending point. Paths are continuous. The ending point of one segment serves as a starting point for the next. Segments correspond to changes of direction in the route and routes can be schematized as a network of nodes and links. It is straightforward to represent a network of streets (or hallways, for that matter) as an annotated graph consisting of edges and nodes, edges representing streets and nodes crossings. Street items, streets and crossings are static objects. For streets the annotations contain:

- References to the neighboring street items at both ends, another street segment or a crossing
- References to objects located at the street item (e.g., buildings)
- A function to compute the middle line of the street using the item's geometry specification and its location in world coordinates. This allows us to move along the street on the middle line.
- A scheduling bound

The annotations of a crossing include:

- References to the neighbors, usually streets
- References to buildings at the crossing
- A functions to compute a connection from the middle line of attached streets to the center of the crossing
- A scheduling bound

Using a search algorithm on graphs will give us a path form Point A to Point B as a list of edges and nodes. Together with the functional annotations of street items, we are able to derive a continuous path. Such a path can be transformed to a first description of a route: forward, right, forward ... right, forward, halt. Most people will find such a description hard to memorize, but it might be suitable to take with you on a piece of paper or your PDA. On the PDA we can

visualize such a description by arrows pointing in the direction to go. At the information booth you would see a raw visualization of the course of motion, without buildings. To incorporate additional information in such a description we need information about objects along the course of motion, which might be used as landmarks.

Representing buildings

Buildings (building sections, etc.) are static objects. In addition to their obligatory representation (geometry, appearance and object name) they include the complete list of general annotations for objects (see above). This is because buildings may be very complex and have hierarchical structure. They may consist of many subparts, references to subparts of the building, references to rooms in the building, which may have references to their furniture and so on. They also contain different coordinate systems and information (in objects coordinates), how these coordinate systems are attached to the building, e.g., specifying the building's prominent/intrinsic front and information on how coordinate systems are scaled. This latter information depends on the object's size and the context of the environment. For example the meaning of the word "near" describes different regions depending on the object's size. "Near the empire state building" denotes a different region than "the electron is near the nucleus". Another example concerning context: Your nearest neighbor in Australia might be 300 miles away from you, while in Manhattan this is usually not the case. For a detailed discussion of these concepts see (Gapp 1994). With this information about objects along a path from A to B we are able to:

1. Determine the geometries we need to display (or load in the first place). In the case of a large and complex environment this obviously helps saving computational resources.
2. Present a "top-down" view on only the path in question (including its buildings) on a PDA and at the information booth.
3. Visualize a virtual walkthrough with buildings in the environment.
4. Give verbal or textual descriptions using the linguistic/textual annotations
5. Use all the functional annotations of the relevant objects, particularly query knowledge/databases

In the next section we describe the representation of the virtual presenter and how we use the aforementioned scheduling bounds to make the virtual presenter "aware" of her current environment.

The virtual presenter

The virtual presenter differs from all other objects. First it is represented as a mobile object. Also, in situations in which we have to speed up the presentation of the virtual walkthrough to a level at which the animation of the presenter isn't tractable or useful anymore, we can omit her visualization, but only her visualization. The invisible virtual presenter herself stays alive, still offering her abilities to scout the environment. The virtual scout with her egocentric frame of reference is shown in figure 2. Using this coordinate system to es-

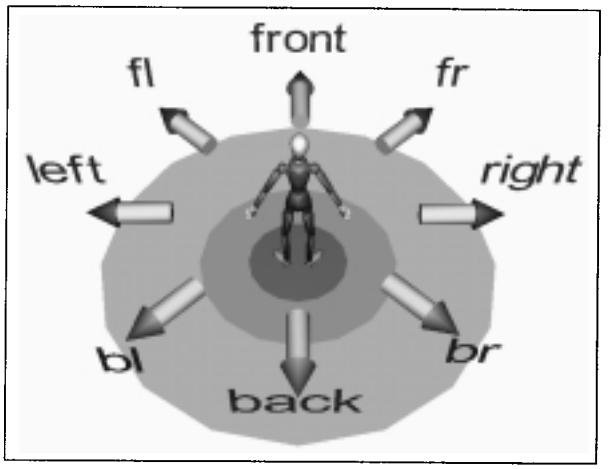

Figure 2: The virtual presenter and the coordinate system establishing the egocentric frame of reference

tablish the egocentric frame of reference we are able to obtain a qualitative representation of space[1]. This coordinate system differentiates four distance relations: here, near, far, very far (from dark grey to white in the picture), where each successive range reaches twice as far as the previous one. Furthermore it differentiates between eight primary directions. For each object referred to with respect to this coordinate system the object's distance and directions to the origin of this system will be discretized in 4 levels of distances and 8 values for direction. This is a transformation from a quantitative representation of an object's location to a qualitative representation. For example ("building–134" 6.0 45) becomes ("building–134" far front-right) which can be easily transformed into a description using building–134's linguistic annotations: "There is a white house left in front of you." When we transform quantitative representations into qualitative ones and verbalize them, we loose information. The localization will be vague. This loss of information can be compensated for by highlighting or labeling the object in the walkthrough. Perhaps we can point at it. Pointing could be defined as a behavior of the virtual scout. The scout herself has to meet several requirements. According to her functional roles in the presentation, she must be "familiar" with a range of presentation gestures and rhetorical body postures (e.g. standing upright) and

[1]For a discussion of different representations of space see (Freksa, Habel, & Wender 1998).

should adopt a reasonable and lively behavior without being distracting or obtrusive. Here we propose a high level declarative specification of the presenter's top level behaviors: follow–path(list–of–street–items), show–object(building). These behaviors can be used to automatically generate the virtual presenter's animation. For a purely synthetic actor (i.e. not controlled by a human,) the system must generate the right sequence and combination of parameters to create the desired movements. In order to achieve this, we propose a hierarchical structure of behaviors where our top level behaviors are assembled from other behaviors.

- follow-path(list–of–street–items): a navigational behavior which enables the presenter to follow a path computed by a search algorithm. It is scheduled by the scheduling bounds of street items. This behavior should make the virtual presenter's movement appear smooth, continuous and collision free. To achieve this natural movement the virtual presenter's navigational behavior generates a spline that interpolates the discretized path from the virtual presenter's current location in the environment through a list of successive control points to the target destination. The computation of control points can be done with the help of the street items' functional annotations. The navigational behavior consists of different other behaviors, for example behaviors for leg movement.
- show-objects(building): This behavior consists of a look-at(building) and/or a point-at(building) behavior. These behaviors will be scheduled by the object's scheduling bounds. The point-at behavior is built from different other behaviors to move the presenters arm. These behaviors enable the scout to point and/or glance at an object. The point to look and/or point at can be computed using the appropriate function from the object's functional annotations.

It should be stated that all the actions constituting the different behaviors could be done in parallel and at different speeds. Knowing minimal and maximal speeds for the different behaviors/actions allows us to give the virtual scout the command: follow-path(from–a–to–b) and the scout will perform the task alone without system intervention, by querying the smart environment. A show-object behavior without a parameter will also be implemented. This involves searching for landmarks suitable for route descriptions and incorporating them in the description. How to solve this special task is described in detail in (Maass 1996).

Conclusions and future work

In this paper we have presented some aspects of building a resource adaptive 3D route description system. We particularly focused on details about how to represent spatial configurations of the domain that are important for the task. The actions performed by the virtual presenter (e.g., moving along the path, pointing at objects and turning her head) are planned on an abstract level. The final presentation is determined by the limitations (e.g. time restrictions) coming up during the presentation. Thus the presentation will be tailored to the context without replanning it.

The methods used to install the objects' frames of reference and to derive information about simple spatial prepositions are derived from former empirical work that has been carried out in this field. In order to apply more complex prepositions (e.g. path relations, such as "along") to our scenario, we have recently carried out experiments that are still under evaluation. Hopefully the results will help to improve the selection of the avatar's path, so that a viewer can memorize the shown information easier.

Another subject of further investigation is the question, down to which level the planner should specify the description of the presentation. One answer might be the design of a mulitresolution planer that provides different elaborations of one presentation plan amongst which the system can choose the one that fits the given restrictions best.

We also started work on the implementation of the information booth and the palm pilot scenarios recently (see the system snapshots in this paper) and will be able to test the plausibility of the system's output soon.

Acknowledgments

Three of the authors are members of the Collaborative Research Center 378 "Resource adaptive cognitive processes" partially funded by the DFG (Deutsche Forschungsgemeinschaft).

References

André, E., and Rist, T. 1996. Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the 13th National Confrence on Artificial Intelligence*, 142–147.

Baus, J.; Butz, A.; and Krüger, A. 1999. One way interaction: interactivity over unidirectional links. In *Proceedings of I3 workshop on Adaptive Design of Interactive Multimedia Presentations for Mobile Users, Sunday, March 7, 1999.* www.i3net.org.

Doyle, P., and Hayes-Roth, B. 1997. Agents in Annotated Worlds. Technical Report KSL 97-09, Kowledge Systems Laboratory, Department of Computer Science Stanford University.

Freksa, C.; Habel, C.; and Wender, K., eds. 1998. *Spatial cognition – An interdisciplinary approach to representation and processing of spatial knowledge.* Springer.

Gapp, K.-P. 1994. Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space. In *Proceedings of AAAI-94.*

Herrmann, T.; Schweizer, K.; Janzen, G.; and Katz, S. 1997. Routen– und Überblickswissen. Konzeptuelle Überlegungen. Bericht Nr.1 des Mannheimer Teilprojekts 'Determinanten des Richtungseffekts' im

Schwerpunktprogramm 'Raumkognition', Universität Mannheim, Lehrstuhl Pschychologie III.

Klatzky, R. 1998. Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections. In Freksa, C.; Habel, C.; and Wender, K., eds., *Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge.* Springer. 1–17.

Krüger, A. 1999. *Automatische Abstraktion in 3D-Graphiken.* Ph.D. Dissertation, Universität des Saarlandes.

Maass, W., and Schmauks, D. 1998. MOSES: Ein Beispiel für die Modellierung räumlicher Leistungen durch ein Wegbeschreibungssystem. *Zeitschrift für Semiotik 20* 91–103.

Maass, W.; Baus, J.; and Paul, J. 1995. Visual grounding of route descriptions in dynamic environments. In *Proceedings of the AAAI Fall Symposium on Computational Models for Integrating Language and Vision.* MIT, Cambridge.

Maass, W. 1996. *Von visuellen Daten zu inkrementellen Wegbeschreibungen in dreidimensionalen Umgebungen: Das Modell eines kognitiven Agenten.* Ph.D. Dissertation, Universität des Saarlandes.

Schmauks, D. 1998. Kognitive und semiotische Ressourcen für die Wegfindung. *Kognitionswissenschaft, Sonderheft zum Sonderforschungsbereich (SFB) 378.*

Tversky, B., and Lee, P. U. 1998. How Space Structures Language. In Freksa, C.; Habel, C.; and Wender, K., eds., *Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge.* Springer. 157–177.

Tversky, B. 1993. Cognitive Maps, Cognitive Collages, and Spatial Mental Models. In *Spatial Information Theory. A Theroretical Basis for GIS, COSIT'93*, 14–24.