VITRA GUIDE: Multimodal Route Descriptions for Computer Assisted Vehicle Navigation

Wolfgang Maaß, Peter Wazinski, Gerd Herzog
SFB 314, Project VITRA
Universität des Saarlandes
D-66041 Saarbrücken, Germany

Abstract

We present a system that generates multimodal route descriptions for computer assisted vehicle navigation.[1] The presentation modes available in *Vitra Guide* are natural language, maps and perspective views. Depending on situation-specific constraints, e.g., fully specifed route descriptions for pre-trip planning or incremental route descriptions during driving, *Vitra Guide* allows the coordinated distribution of the route information over the available presentation modes. The coordinated use of the presentation modes is achieved by a plan-based approach. The output for all presentation modes relies on one common 3-dimensional geometric model of the domain.

This paper appeared in: Proc. of the Sixth Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE-93, pp. 144–147, Edinburgh, Scotland, 1993.

[1]The work described here has been supported by the German Science Foundation (DFG) in its Special Collaborative Research Programme on Artificial Intelligence and Knowledge Based Systems (SFB 314), project N2: *Vitra*.

1 Introduction

Due to the continuously increasing amount of vehicular traffic in our urban and extra-urban environment, much attention is paid to the development of traffic-control systems. In order to influence specific traffic patterns, e.g., the "morning peak", certain traffic data (traffic volume, occupancy rates, speed) are collected via induction loops or other detectors (c.f. Jansen et al. [1992]).

This paper addresses the individual counterpart of collective traffic guidance: Computer Assisted Vehicle Navigation (CAVN). Nowadays CAVN has become tractable for real-world situations by the confluence of various techniques. CD-ROM provides enough memory capacity to hold digitised road and city maps. Position fixing of vehicles is possible by map-matching (c.f. Neukirchner and Zechnall [1986]) or even via satellite (Davis [1989]). New or improved output devices raise the possibility of generating of multimodal route descriptions (Maybury [1991]) tailored to situation-dependent needs: For example in the context of pre-trip planning the combined use of maps and perspective views may be the most appropriate presentation, but during driving natural language instructions should be more suitable.

In the *Vitra* project we are concerned with the generation of multimodal route descriptions that combine natural language, maps and perspective views. Only the coordinated use of these different presentation modalities is able to fully exploit the possibilities offered by the new communication technologies. The next two sections discuss the presentation modalities separately, while the last one focuses on their coordinated combination.

2 Natural Language Route Descriptions and Spatial Reference

The mediation of spatial information is the essential task for the generation of natural language route descriptions. The identification of path segments and landmarks as well as the indication of direction changes are necessary parts of natural-language-based navigation assistence:

(1) Turn *right* at the next crossing.

(2) Go *under* the bridge and keep going *along* the field for *about 3 minutes*.

(3) On your *left side* you will see a building that has a car park just in *front of* it.

In (1) a direction change is indicated. In (2) a path segment is characterised by relating it to a landmark via a directional preposition ("under"), a path preposition ("along") and a durative constraint ("about 3 minutes"). In (3) the goal is identified by the use of two features: first, the building is located with respect to the driver's orientation by a directional preposition. Then another directional preposition is used to characterise the building by means of a second landmark (the car park).

In *Vitra Guide*, spatial knowledge is only implicitly present by virtue of a 3-dimensional geometric model. A qualitative representation of the spatial configuration, that serves as a starting point for the NL generation, therefore requires the analysis of this geometric model.

We define a *spatial relation* by specifying certain spatial conditions for a possible spatial configuration. Such spatial relations are considered to be independent of a specific language. Nevertheless they form the basis for the analysis of spatial prepositions within the framework of reference semantics (c.f. Schirra [1990]). Following Lakoff (Lakoff [1973]) we argue that a qualitative representation of spatial relations needs a quantitative counterpart in order to reflect the vagueness of natural language, i.e. *linguistic hedges* such as in "*just* in front of" or "*fairly* behind". As distinguished from purely qualitative approaches like Herskovits [1985] or Aurnague and Borillo [1990], in *Vitra Guide* this quantitative counterpart is realised by the computation of so called applicability degrees. Applicability degrees are real numbers from the interval [0..1], that are related to the various linguistic hedges (c.f. André et al. [1989]). In addition to forming the counterpart of qualitative representations, applicability degrees supply a criterion for selecting between different spatial predicates that hold for the same spatial configuration.

We distinguish three classes of spatial relations (c.f. Retz-Schmidt [1988]):

1. Topological relations ("in", "near", etc.) are based on topological terms, e.g., *neighbourhood* or *interior*. We favor Egenhofer's approach (Egenhofer [1991]) for 2-dimensional binary topological relations. Our work aims at extending this method in order to deal with 3-dimensional space and applicability degrees (Wazinski and Herzog [1992]).

2. Directional relations ("left", "right", etc.) depend on the orientation of the surrounding space and therefore require the construction of a reference frame. We distinguish between the induction of a reference frame by the intrinsic orientation of the reference object ("in front of the house"), by the deictic perspective of the speaker/hearer ("as seen from me/you") or by an extrinsic perspective, given by other external contextually induced factors ("as seen from the church").

3. Path relations ("along", "past") characterise trajectories of objects. As the movement of an object takes place in time, path relations involve spatial and temporal aspects. Our previous work on 2-dimensional path relations (André et al. [1986]) is currently being extended for the 3-dimensional geometric model.

3 Graphical Presentation Modes for Spatial Information

For everyday route descriptions various kinds of maps play a central role. The usage of street maps, region maps, world maps or anatomic maps is quite similar. If you want to get from point A to point B then you start with an orientation. In the next phase you follow the intended route. Both phases, orientation and route finding, are fundamentally guided by visual perception. Therefore, features like landmarks or direction are main guidelines in finding your way.

Within the research field of cognitive maps there is a distinction between *route knowledge* that has more procedural aspects and *survey* or *map knowledge* that has more abstract aspects (Thorndyke and Hayes-Roth [1982]; Gluck [1991]). Maps are commonly used in route description systems (Kuipers [1983]; Neukirchner [1991]). Survey knowledge about the region concerned is given by a 2-dimensional projection onto the horizontal plane. This presentation mode alone is suboptimal because of the cognitive effort the driver has to spend for matching the projected landmarks or crossroads with their appearance in reality. Here perspective views that simulate the driver's perspective can help.

Both presentation modes can be supplemented by several display techniques such as pointing gestures, highlighting, flashing, etc. We distinguish the following modes:

1. maps, i.e., 2-dimensional projections onto the horizontal plane with
   (a) static display attributes (highlighting, flashing, arrows)
   (b) dynamic pointing gestures (the position of the pointer changes in time, c.f. Jung et al. [1989])
   (c) annotated landmarks
2. perspective views with
   (a) static display attributes
   (b) dynamic pointing gestures
   (c) graphical animation.

The left part of figure 1 shows a city map with a pointing gesture and can be used, e.g., for giving survey knowledge in the context of pre-trip planning. The right part of figure 1 gives a perspective view. Such perspective views may be useful for displaying critical choice points of the particular route: a landmark might be only identifiable by a pseudo-realistic image from the driver's perspective.

4 Multimodal Route Descriptions

In *Vitra Guide* we want to distribute the route information to be presented according to the individual strength of each presentation mode. In addition, we have to take into account the specific constraints of the possible application contexts. For pre-trip planning it is desirable to have fully specified route descriptions. In this case all presentation modes can be used. During driving it is necessary to give the driver an incremental route description, i.e. supplying the driver continuously with information. Graphical output should be reduced, e.g. no perspective views, only maps mirrored into the lower part of the windscreen.

A well-tried technique within the field of multimodal user interaces (for an overview see Sullivan and Tyler [1991], Maybury [1993]) is the use of a plan-based approach (Hovy [1988], André and Rist [1993]). An overall planner takes into account the various presentation objectives such as

- the mediation of a certain amount of information,
- the optimal distribution of the information over the presentation modes,
- the multimodal coherence of the presentation to be generated (Bandyopadhyay [1991]) and
- the consideration of resource limitations according to the specific user and application context.

The use of a plan-based approach for route descriptions poses the problem of integrating temporal aspects into the planning process. Temporal relations, e.g., "Turn left *immediately/after 5 minutes*", are an important part of route descriptions. Spatial relations involving movement ("Keep going along the field.") only hold for a certain amount of time. Thus, the planner for *Vitra Guide* has to integrate both spatial and temporal aspects.

5 Conclusion

Computer assisted vehicle navigation has become feasible through the development of new communication technologies. The generation of multimodal route descriptions requires the coordinated use of the available presentation modes. The presentation modes we use in *Vitra Guide* are: natural language, maps and perspective views. For all presentation modes the output is generated from one common 3-dimensional geometric model of our domain. Coordinated use of the presentation modes and consideration of the other presentation objectives can be achieved by a plan-based approach.

Although this paper focuses on the use of *Vitra Guide* for CAVN, we stress that the underlying methods are general enough to be applied in other areas that are concerned with multimodal interfaces for the presentation of spatial information. For example, a stationary route description system in a tourist information office may rely on the same presentation techniques.

Vitra Guide is the extension of a former, fully implemented route description system (Müller [1989]), that works for a 2-dimensional domain model. We have finished the 3-dimensional domain model and the realisation of the new display techniques (2-dimensional projections, perspective views, animation). Our current work is concerned with the seamless coordination of the presentation modes.

References

E. André, G. Bosch, G. Herzog, T. Rist. Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions. In: *Proc. of the 7th ECAI*, vol. 2, pp. 1–8, Brighton, UK, 1986.

E. André, G. Herzog, T. Rist. Natural Language Access to Visual Data: Dealing with Space and Movement. Report 63, Universität des Saarlandes, SFB 314 (VITRA), Saarbrücken, 1989, Presented at the 1st Workshop on Logical Semantics of Time, Space and Movement in Natural Language, Toulouse, France.

E. André, T. Rist. The Design of Illustrated Documents as a Planning Task. In: M. T. Maybury, ed., *Intelligent Multimedia Interfaces*, pp. 94–116, AAAI Press, Menlo Park, CA, 1993.

M. Aurnague, M. Borillo. A Formal Semantics for Internal Localization: An Essay on Spatial Commonsense Knowledge. In: P. Jorrand, V. Sgurev, eds., *Artificial Intelligence IV: Methodology, Systems, Applications*, pp. 305–317, North-Holland, Amsterdam, 1990.

S. Bandyopadhyay. Towards an Understanding of Coherence in Multimodal Discourse. Technical Memo TM-90-01, DFKI GmbH, Kaiserslautern und Saarbrücken, 1991.

J. R. Davis. *Back Seat Driver: Voice Assisted Automobile Navigation*. Ph.D. thesis, Media Arts and Science Section, MIT, Cambridge, MA, 1989.

M. J. Egenhofer. Reasoning about Binary Topological Relations. In: O. Günther, H.-J. Schek, eds., *Advances in Spatial Databases*, pp. 144–160, Springer, Berlin, Heidelberg, 1991.

M. Gluck. Making Sense of Human Wayfinding: Review of Cognitive and Linguistic Knowledge for Personal Navigation with a New Research Direction. In: D. M. Mark, A. U. Frank, eds., *Cognitive and Linguistic Aspects of Geographic Space*, pp. 117–135, Kluwer, Dordrecht, 1991.

A. Herskovits. Semantics and Pragmatics of Locative Expressions. *Cognitive Science*, **9**(3), 341–378, 1985.

E. H. Hovy. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum, Hillsdale, NJ, 1988.

B. Jansen, H. Kirschfink, M. Richter. Traffic data Support Systems via AI-based Methods. In: *Proc. of the First Int. Conf. on Artificial Intelligence Applications in transportation Engeneering*, Santa Buenaventura, CA, 1992.

J. Jung, A. Kresse, N. Reithinger, R. Schäfer. Das System ZORA: Wissensbasierte Generierung von Zeigegesten. In: D. Metzing, ed., *GWAI-89: 13th German Workshop on Artificial Intelligence*, pp. 190–194, Springer, Berlin, Heidelberg, 1989.

B. J. Kuipers. The Cognitive Map: Could It Have Been Any Other Way? In: H. L. Pick, L. P. Acredolo, eds., *Spatial Orientation: Theory, Research, and Application*, pp. 345–359, Plenum, New York, London, 1983.

G. Lakoff. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic*, **2**, 458–508, 1973.

M. T. Maybury. Planning Mulitmedia Explanations Using Communicative Acts. In: *Proc. of AAAI-91*, pp. 61–66, Anaheim, CA, 1991.

M. T. Maybury, ed. *Intelligent Multimedia Interfaces*. AAAI Press, Menlo Park, CA, 1993.

S. Müller. CITYGUIDE: Wegauskünfte vom Computer. Memo 37, Universität des Saarlandes, SFB 314 (VITRA), Saarbrücken, 1989.

E. P. Neukirchner. Fahrerinformations- und Navigationssystem. *Informatik Spektrum*, **14**(2), 65–68, 1991.

E. P. Neukirchner, W. Zechnall. Digital Map Data Bases for Autonomous Map Matched Vehicle Navigation Systems. In: *Proc. of PLANS'86*, Las Vegas, NV, 1986.

G. Retz-Schmidt. Various Views on Spatial Prepositions. *AI Magazine*, **9**(2), 95–105, 1988.

J. R. J. Schirra. Einige Überlegungen zu Bildvorstellungen in kognitiven Systemen. In: C. Freksa, C. Habel, eds., *Repräsentation und Verarbeitung räumlichen Wissens*, pp. 68–82, Springer, Heidelberg, 1990.

J. W. Sullivan, S. W. Tyler, eds. *Intelligent User Interfaces*. ACM press, New York, 1991.

P. W. Thorndyke, B. Hayes-Roth. Differences in Spatial Knowledge acquired from Maps and Navigation. *Cognitive Psychology*, **14**, 560–589, 1982.

P. Wazinski, G. Herzog. Coping with Topological and Directional Relations Based on 3-Dimensional Geometric Representations. In: *Spatial Concepts: Connecting Cognitive Theories with Formal Representations, Workshop Notes, 10th European Conf. on Artificial Intelligence, Vienna, Austria*, 1992.

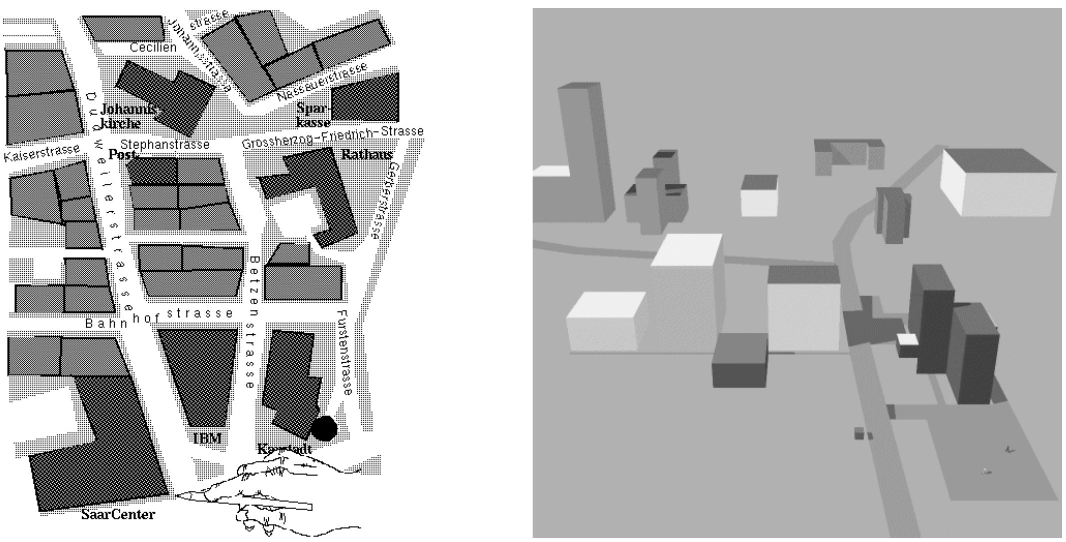

Figure 1: Graphical presentation modes: maps and perspective views