Spatial reasoning in a multi-modal user guide for a complex machine

Nadejda Soudzilovskaia, Rafael Bidarra, Frederik W. Jansen

Delft University of Technology,
Faculty of Information Technology and Systems[1]
E-mail: (N.Soudzilovskaia / R.Bidarra / F.W.Jansen)@its.tudelft.nl

ABSTRACT

This article discusses the application of spatial reasoning techniques in the field of augmented environments. It focuses on resolving spatial references within the flow of a user-computer dialogue. A dialogue-based user guide for a complex machine is presented, utilising a 3D model of the machine and knowledge-based reasoning. The reasoning functionality includes reasoning on the usage of the machine as well as on its construction and location of sub-components. Two issues are discussed. The first issue is the architecture for the user guide supporting both 3D graphical presentations and graphics-based reasoning within the dialogue process. The second issue concerns algorithms for spatial reasoning used in the system. The following algorithms are discussed: the use of functional hierarchy, reduction of dimension, and the use of zones of applicability for spatial references.

1. INTRODUCTION

Nowadays a wide range of tutorials and user guides are commercially available. Companies use computer-based techniques for electronic manuals of their products. Most of these user guides however do not include a dialogue-based expert system but instead propose a fixed flow chart-based navigation through the application. Except for some research oriented systems, as described in [1, 10, 12], few manual-like applications use reasoning algorithms. The main reason for it is that the techniques for reasoning and dialogue management are still not powerful enough to fulfil the user needs. It is therefore still not beneficial to develop dialogue-based user guides for commercial purposes.

If we consider reasoning algorithms to be applied in user guides for complex machines, it is not enough to reason only about the functionality of a machine, but it is also necessary to include its construction, appearance and the location of (sub)-components into the reasoning process.

In the field of spatial reasoning, a remarkable amount of research is already available. One of the most important problems is that people hardly ever use absolute coordinate-like references to describe positions of objects. Instead, people prefer to describe the position of objects relative to other objects (‘to the left of’, ‘above', etc.). In automatic generation of a textual description for an object location, one of the most difficult aspects is to select an adequate reference object. Some experimental investigations about how people describe objects in speech are however available [4, 5, 9]. The most influencing factors in this process are size, colour (brightness), mobility of an object, frame of reference, location of an object relatively to others, distances, functional dependencies, prior knowledge and previously mentioned objects [8]. Several attempts have been made to model the process of choosing an optimal reference object. The use of geometric properties of object [6] and functional dependencies between objects [7] has been suggested.

When going through the process of analysing a textual description of an object location, another question appears: how far and precise can the geometric properties and locations of an object be derived from its linguistic description? The linguistic meaning of spatial prepositions is far much larger than only geometry [3]. Therefore it is essential to apply additional semantics within the geometric based reasoning on the description of an object location.

It appears that most research in this area has been carried out in the domains of navigation (interactive mapping applications and GIS), robot control and image analysis [2, 7, 11, 13]. Few research addresses the spatial reasoning in the domain of a user guide for complex machines and devices. When a complex machine is considered, its components create a hierarchical structure. This article discusses usage of the hierarchical dependencies and knowledge of the machine (sub)-components’ functionality in the process of spatial reasoning.

As a study case of a complex machine, a multi-modal user guide for a fax-machine was chosen. The system provides a 3D model of the fax and supports dialogues on the use of the machine (how

[1] Mekelweg 4, NL-2628 CD Delft, The Netherlands

to install and operate it) and on the appearance and location of the fax sub-components. The hierarchical relations between the fax components are used both for the graphical presentation (building the 3D images of the fax machine) and for the spatial reasoning. This approach leads to a reduction in ambiguity of spatial reasoning results. It is shown that knowledge of the machine component functionality can be used to reduce the three dimensional space of reasoning on three-dimensional objects to two dimensions.

2. SPATIAL REASONING

The components of a complex machine can be considered from two points of view: as a functional block of a machine, with its own place in a hierarchy, and as a geometric object with purely geometric characteristics, such as features and geometry. See Figures 1 and 2 (a, b).

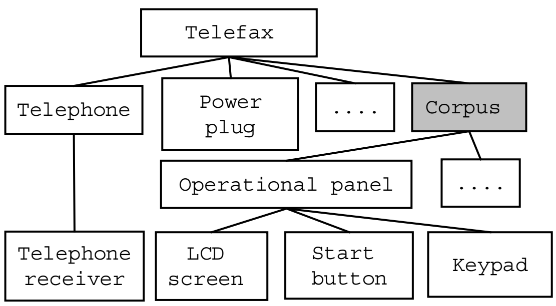

Figure 1: Fragment of a functional graph of the fax

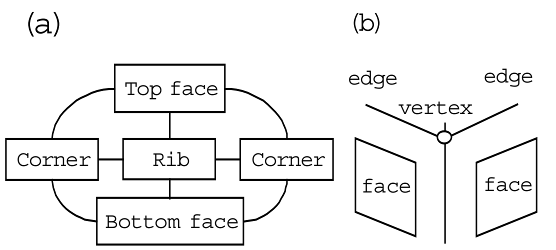

Figure 2: Semantic graph of the object (a), Geometry graph of the object (b).

This section presents algorithms where both functional and geometric approaches complement each other enhancing the reasoning process.

2.1. Use of functional hierarchy

When the location of an object is described by speech, two objects are usually given for reference: the first often a step higher in the hierarchy, the second often on the same level, e.g. “the *start button* is on the *operational panel* below the *LCD screen*”. This fact can be used in the reasoning process. First a hierarchy analysis can be done, reducing the number of possible reference objects; after that, reasoning based on the geometric properties of the objects can be applied. (Algorithms using the geometrical properties (size, centre of gravity and bounding box) are described in [6].)

2.2. Reduction in reasoning dimension

Although in reality objects we deal with are three dimensional, for reasoning, an approximation to a lower dimension is possible. For example, if we reason about the position of a certain key on the keypad, we consider the keypad to be a two-dimensional object, which simplifies the required reasoning (the same applies for slots on a slot panel, etc.).

2.3. Using zone of applicability in reasoning on the location of the surface of an object

When reasoning on the position of an object located on the surface of another object, one can define zones of applicability for references such as 'on the corner', 'on the top side of', 'on the right side of', etc. See Figures 3 and 4 (a,b).

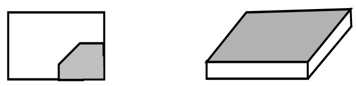

Figure 3: Use of applicability zones for surfaces

Those zones are different for objects that can be considered 2D and for those that are considered 3D. For 3D objects, the 'right side' zone, for instance, will not only cover the left side surface itself, but also adjoining parts of the top and bottom surface. See Figure 4 (a, b).

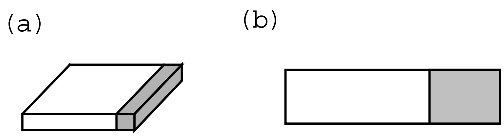

Figure 4: Applicability zones of spatial reasoning rules for 3D object (a) and for 2D object (b)

2.4. Definition of a reasoning dimension

As shown in sections 2.2 and 2.3 the dimension of an object plays an important role in reasoning when the object has children in a hierarchy graph. The reasoning dimension of an object therefore depends not only on a physical dimension of this object but also on the way of location of children on its surface. The proposed solution is to define different categories of objects depending on the location of children on their surfaces, and apply different reasoning rules for these different categoies of objects. The following object types can be defined:

- 3D objects - objects that have a volumetric geometry and can have child nodes on any side face (e.g. fax corpus, telephone)
- 2D objects - objects that have geometry of a flat and can have child nodes only on one side (e.g. operational panel, keypad)
- 1D objects - objects that can have a wire like geometry and can have child nodes situated along that wire or on the ends (e.g. cables)
- SET objects - abstract notions, that represent some group of objects. They have no geometry, but do have child nodes (e.g. group of slots, group of switches)
- LEAF objects - objects that have no child nodes (e.g. buttons, slots, LCD screen, paper tray)

3. MULTI-MODAL FAX USER GUIDE

The fax user guide has the form of a "fax encyclopaedia". The user can ask questions about the fax and get answers supported by graphic demonstrations. The system architecture of the "fax encyclopaedia" is shown in Figure 5.

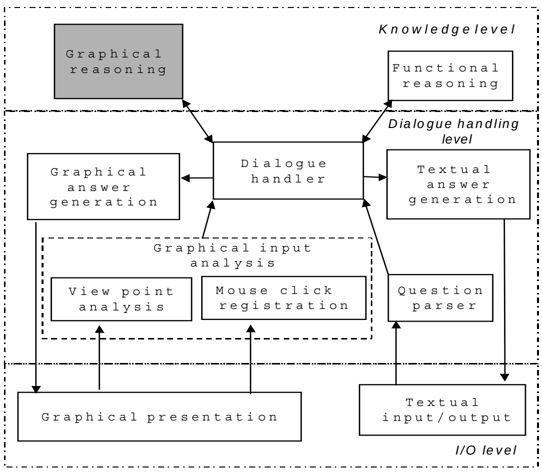

Figure 5: Fax user guide system architecture

The system consists of three levels: the I/O level, the Dialogue handling level and the Knowledge level. The I/O level handles the functionality of the navigation through the virtual world (therefore user can make his own investigations about the fax machine), input of user questions in speech form and presentation of answers, given in the form of speech and graphical illustration. The Dialogue handling level performs control over the flow of the dialogue. At this level the following processes are handled: parsing of questions by analysis of the question itself and analysis of status of the virtual world (i.e. analysis of the view point and mouse clicks); querying the data at the Knowledge level; processing of the query results and generation of the appropriate user answer.

The Knowledge level provides the functionality for data storage and reasoning. Two modules are situated at this level: the Graphical reasoning module and the Functional reasoning module. The Functional reasoning module performs semantics-based reasoning on the fax machine and its components. The Graphical reasoning module provides functionality for graphics-based reasoning (i.e. reasoning about the position and appearance of fax components).

4. GRAPHICAL REASONING

The graphical reasoning module has a dual function in the system. It is used in both the processes of reasoning and presentation of the fax machine construction and functionality. The system applies the graphics-based reasoning in the following three cases:

- to clarify a textual answer by a graphical demonstration
- to derive a textural description of a position of a given component
- to resolve a spatial reference in a text of a user question (for example, when user use a description like "the button in the left corner of the operational panel")

In the first case, the Dialogue handler module informs the Graphical reasoning module which component is to be demonstrated. The Graphical reasoning module analyses the model of the fax in the virtual world, chooses an optimal view point for the requested fax component and highlights the component. Figure 6 demonstrates the operation of the program.

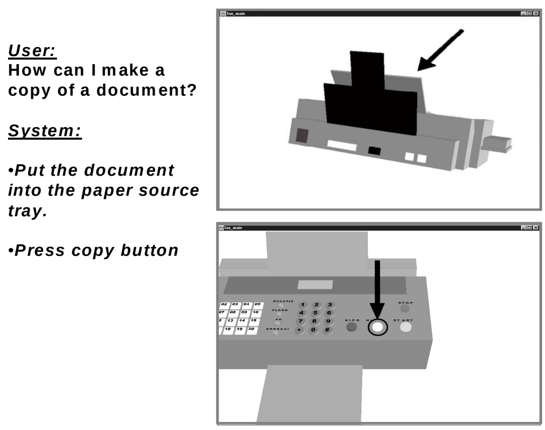

Figure 6: Multi-modal fax user guide: supporting an answer with a graphical demonstration

The choice of an optimal viewpoint is based on the analysis of the size and position of the bounding sphere for the component. The distance from the

viewpoint to the component is calculated as a function of the radius of the bounding sphere. This enables to perform the scaling automatically.
In the second and third cases, the algorithms presented in Section 2 are used. The following rule is applied for analysis of the location of an object:

- Step1: Search for parent of the component in the hierarchy
- Step 2: Develop a description of the component location on the surface of the parent
- Step 3: Develop a description of the component location relatively to other components

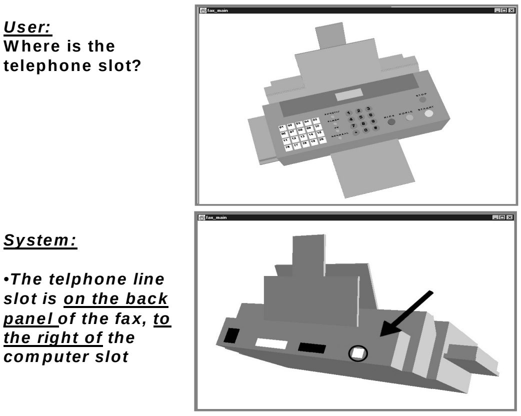

Figure 7: Multi-modal fax user guide: development of a component location description

5. CONCLUSIONS AND FUTURE WORK

This article presents an investigation on spatial reasoning techniques in the application of multi-modal user guides for complex machines. A fax machine user guide is developed, utilising advanced graphical support. The hierarchical structure of the machine is represented in a scene graph and used both for spatial reasoning and graphical presentations for dialogue support. Several algorithms for the spatial reasoning on a complex machine are developed. The hierarchical structure of the machine components is used to reduce the number of possible reference objects. Therefore, a simplification in the process of the optimal reference object selection is achieved. Knowledge of the component functionality allows a reduction in the reasoning dimensions for the spatial reasoning algorithms. The zones of applicability for spatial reference differ for 2D and 3D objects. This can be used to reduce the inherent ambiguity in man-machine communication. All the proposed algorithms lead to reduction of the reasoning complexity and therefore can be applied in a real-time augmented reality system [14]. Future work will be aimed at the improvement of the described a alrithms, using mathematical techniques such as fuzzy logic.

6. ACKNOWLEDGEMENT

This research was carried out as part of the MATIS project (Multimodal Access to Information Services) paid by the Innovation Oriented Programme Man Machine Interaction (IOP MMI) of the Dutch Ministry of Economic Affairs.

7. REFERENCES

[1] H. Bunt, R. Ahn, et al., "Multi-modal co-operation with the DenK system", *Niels Bernen et al. (eds.) "Designing interactive speech systems : from first ideas to user testing"*, pp. 37-69, London : Springer, 1998.
[2] W. Burgard, et al., “Experiences with an interactive museum tour-guide robot”, *Artificial Intelligence*, 114 (1-2): 3-55, 1999.
[3] K. R. Coventry, "Spatial Prepositions, Functional Relations and Lexical Specification", *P. Olivier et al. (eds), "Representation and processing of spatial expressions"*, pp. 247-263, Mahwah: Erlbaum, 1998.
[4] A. Cremers, "Object reference in a task-oriented keyboard dialogues", *Harry Bunt (ed), "Multimodal human-computer communication: systems, techniques, and experiments"*, pp. 279-293, Berlin: Springer, 1998.
[5] K.-P. Gapp, "An empirically Validated Model for Computing Spatial Relations», *KI-95*, pp. 245-256, 1995
[6] K.-P. Gapp, "Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space", *AAAI-94*, pp. 1393-1398, 1994.
[7] K.-P. Gapp, W. Maaβ, "Spatial Layout Identification and Incremental Descriptions", *AAAI-94*, pp. 145-152, 1994.
[8] R. Jackendoff, B. Landau, "Spatial language and spatial cognition", *R. Jackendoff, “Languages of mind. Essays on Mental Representation”*, pp. 99-124, MIT Press, 1993.
[9] T. Kato, Y. I. Nakano, “Referent identification requests in multi-modal dialogs”, *H. Bunt (ed) "Multimodal human-computer communication: systems, techniques and experiments"*, pp. 295-311, Berlin: Springer, 1998.
[10] R.W. Smith, R. Hipp, “Spoken natural language dialogue systems: a practical approach”, Oxford Univ.Press, 1994 .
[11] S. Strippgen, K.Peters “The other way round!: collaborative communication with agents”, *Agents '99*, pp 108-115, 1999.
[12] O. Sundblad, Y. Sundblad, “OLGA - a multimodal interactive information assistant”, *CHI 98*, pp 183-184, 1998.
[13] G.B. Varile, Z. Antonio "Survey of the state of the art in human language technology", Cambridge Univ. Press, 1997.
[14] Ubicom Project: www.ubicom.tudelft.nl