Sonderforschungsbereich 378
Ressourcenadaptive kognitive Prozesse

KI-Labor am Lehrstuhl für Informatik IV

Leitung: Prof. Dr. W. Wahlster

REAL

Universität des Saarlandes
FB 14 Informatik IV
Postfach 151150
D-66041 Saarbrücken
Germany

Tel. 0681 / 302-2363

Bericht Nr. 135

Processing Spatial Relations in Object Localization Tasks

Klaus-Peter Gapp

November 1996

Sonderforschungsbereich 378
Ressourcenadaptive
Kognitive Prozesse

ISSN 0944-7814

135

Processing Spatial Relations in Object Localization Tasks*

Klaus-Peter Gapp
Cognitive Science Program, Dept. of Computer Science
Universität des Saarlandes
D-66041 Saarbrücken, Germany

Email: gapp@cs.uni-sb.de

Abstract

The localization of an object within a certain environment requires the selection of an appropriate reference object and the establishment of adequate spatial relations between the objects. The decision, which reference object to choose, demands consideration of all applicable spatial relations, which therefore have to be computed before the final decision can be made. However, the calculation of all applicable spatial relations for all potential reference objects is a computationally intensive task. In order to increase performance, it is proposed that the set of spatial relations to be tested can be reduced in two respects without any loss of quality in the generated answer. The two ways in which reduction can be achieved are by taking into account the type of the localization — approximate or precise — and the type of the reference object.

1 Introduction

Object localization is a task in which the integration of language and vision is a required necessity per se. The localization of an object can be processed in two directions: from vision to language (cf. [André *et al.*, 1988; Gapp, 1994c]) and from language to vision (cf. [Olivier *et al.*, 1994; Schirra and Stopp, 1993]). In the former, the position of objects in a given spatial scene is described propositionally or in natural language (NL), in the latter a spatial layout is generated from a NL description. In this paper we focus on the generation of descriptions for answering queries about the spatial position of objects in general domains. A typical query is therefore *"Where is the position of object A?"*, with *object A* being a 2- or 3-dimensional object in the domain.

The task of object localization can be divided up into three major parts: first, the object of interest, the LO, is identified, then an object suitable for use as a referent, the RO, is selected, and finally a linguistic description of the relationship between the objects from some specific point of view is specified (cf. [Gapp, 1995c]). In this paper we will place emphasis on examining the processing of spatial relationships between objects as a whole, which is more than merely the computation of essential relations.

First, a short review about an universal model for the computation of topological and projective relations is given. The model is then extended to interior projective and geographical relations. The extended model provides the basis for our further considerations of the processing of spatial relations in general object localization tasks.

Possible applicable relations in a particular spatial configuration can be manifold and therefore computationally intensive. It is shown that in object localization tasks the set of spatial relations to be considered is situational dependent and can therefore be restricted to specific subsets in advance.

2 A Computational Model of Spatial Relations

Increasingly sophisticated approaches to the computation of spatial relations have been developed in the last couple of decades (e.g. [André *et al.*, 1988; Egenhofer and Herring, 1990; Mukerjee and Joe, 1990; Abella and Kender, 1993; Rajagopalan, 1994; Schirra and Stopp, 1993; Olivier *et al.*, 1994]). In [Gapp, 1994a] the CSR-3D[1] system, a model for the computation of topological and projective relations (cf. [Talmy, 1983; Herskovits, 1986]) in 2D and 3D

*In: Proc. of the AAAI Fall Symposium on "Computational Models for Integrating Language and Vision", MIT, Cambridge, MA, USA, 1995.

[1]The CSR-3D system (Computation of Spatial Relations in 3D-Space) is part of the VITRA (VIsual TRAnslator) project which deals with the relationship between vision and natural language (cf. [André *et al.*, 1989]). The system starts from a geometric scene description, given by a vision system (cf. [Koller, 1992]), and proceeds to describe a recorded video sequence or to answer questions about the location of the objects involved in a cognitively plausible manner.

Figure 1: Distance and angle are measured in a space which is scaled by the RO's extension.

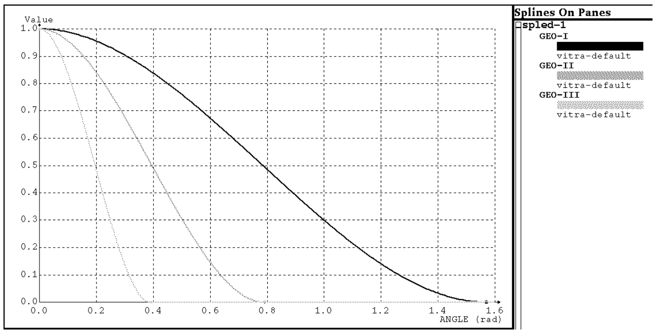

Figure 2: Definition of the evaluation functions for the three groups of cardinal relations.

space, was presented. The model was enhanced to include composite spatial relations, such as *to the left of and behind* or *to the left of and near* which are very common used in German (cf. [Gapp, 1994b]), and proved for cognitive plausibility (cf. [Gapp, 1995a; 1995b]).

The essential idea of the model is to measure the distance (topological relation) and the angle of deviation (projective relation) from the LO's center of gravity to the RO, in a space which is scaled by the extension of the reference object, and to account for the vagueness of spatial relations by mapping these values to specific evaluation functions (cf. Figure 1). The compositional relations are obtained by applying a weighted minimum function to the elementary relations.

To account for object localizations in large scale space, and for the interior of objects it is necessary to extend the model to cardinal, e.g. *north* or *south*, and internal projective relations, such as *top* or *top left*.

2.1 Cardinal Relations

Cardinal relations, e.g. *south, south-west*, or *west-south-west*, are used to ensure appropriate descriptions of the orientation of objects in large-scale geographic space. Cardinal relations are independent from a reference object's extension. The angle of deviation of the located object from the relation's canonical direction can therefore be measured absolutely. Three kinds of evaluation functions are used to model their accuracy (cf. Figure 2). The more specific the relation, the more restricted the evaluation function. The following ranges are defined:

- 90^o for the relations *north, east, south*, and *west* (GEO-I).
- 45^o for the relations *north-east, south-east, south-west*, and *north-west* (GEO-II).
- 22.5^o for the relations *north-north-east, east-north-east, east-south-east, south-south-east, south-south-west, west-south-west, west-north-west*, and *north-north-west* (GEO-III).

Figure 3 shows an example of how the different grades of evaluation functions affect the applicability structure of the relations in 2D space (3D space analogous). The darker the shading in the relation's region of applicability, the higher the degree of applicability.

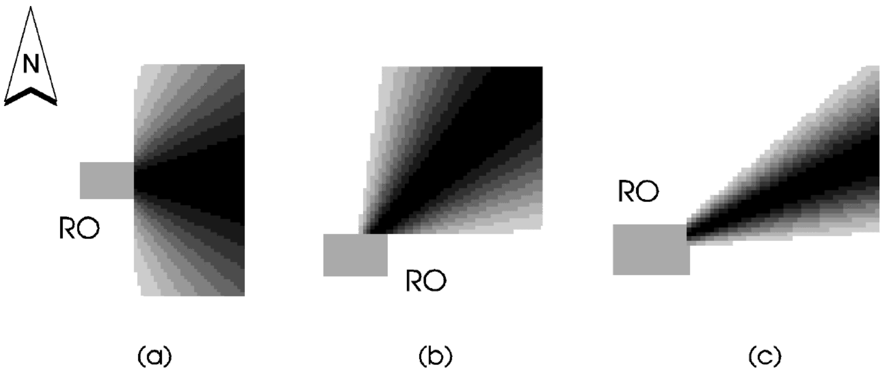

Figure 3: 2D applicability fields for the relations (a) *east* (b) *north-east* (c) *east-north-east*.

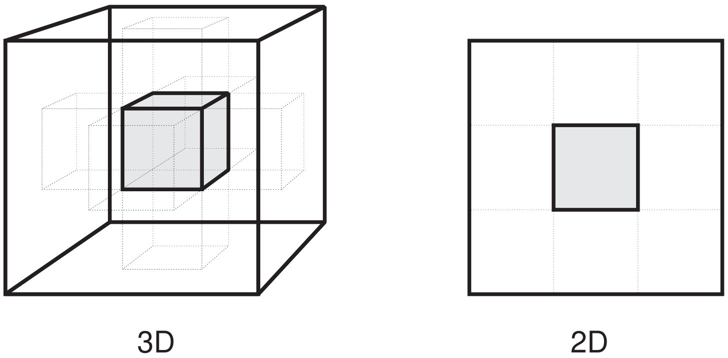

Figure 4: Partitioning of the reference object to compute interior projective relations.

2.2 Internal Projective Relations

In [Wazinski, 1992] a partitioning of the reference object into three parts for each dimension[2] was proposed in order to compute internal projective relations. The regions covered by an object were used to compute its relationship to the reference object.

By dividing up the reference object into three regions for each dimension and using the center region as the reference, CSR-3D now accounts for relations such as *front, back, top, center, top left,* etc. (cf. Figure 4). In contrast to Wazinski's approach, our model can cope with internal projective relations in 3D as well as in 2D space and there are no sharp boundaries between neighboring relations. Also the applicability regions between internal *right/left* and external *to the right/left of* merge into one another as shown in Figure 5.

Combinations of external 3D and internal 2D computations can be used to obtain more detailed information about the location of an object. For example, in Figure 6 an appropriate answer to the query *"Where is the book located?"* would be *"The book is located on the right side of the desk"* which is a combination of the 3D relation *on* and the 2D relation *right*. The computation of the latter requires an abstraction from the original 3D shape of the desk and to use the derived 2D projection.

[2]Only 2-dimensional space was considered.

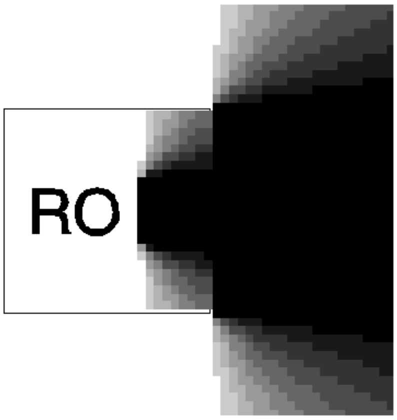

Figure 5: Applicability regions of internal *right* and external *to the right of* merge into one another.

In total the CSR-3D system can now account for more than sixty spatial relations which can be directly linked to linguistic entities, such as spatial prepositions or adverbs.

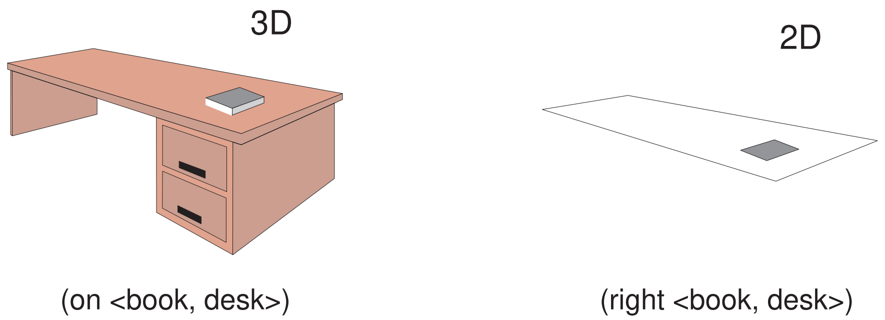

Figure 6: Combined computation in 2D and 3D space achieves more detailed information.

3 Preprocessing

The process of object localization requires the focussing on potential reference objects proximate to the object to be localized. The size of the visual focus depends mainly on the extension of the reference object. Potential reference object candidates are required to be of a certain minimal size, which is again determined by the extension of the reference object. The number of such candidates is situation dependent, and can therefore vary from one to a considerable number of objects. The selection of the best available reference object in the focussed range requires the consideration of all applicable spatial relations to finally decide which object to choose. Figure 7 gives an example of why this might be necessary. In the figure the two reference objects are equal in size and visual salience, and both are the same distance from the located object. However, one would choose RO_1 as the appropriate reference object to describe the LO's position. This is because the spatial relation *below* is simpler and therefore easier to establish than the composite relation *to the right of and below.*[3]

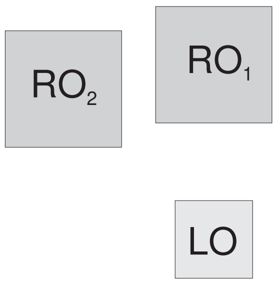

Figure 7: The spatial relationship between the LO and potential ROs can be a decisive criterion for the selection of the best reference object.

Establishing all the applicable spatial relations of potential reference objects in advance is computationally intensive if one considers the large amount of possible relations to be tested. However, the set of relations which are of relevance within a certain context can be restricted in two respects: the type of the reference object and the type of the localization.

[3] Note, that it is necessary in the example for both relations to be of the same degree of applicability.

1.	2.	3.
topological		
projective	cardinal	cardinal projective

Table 1: Three categories of applicable relations depending on the type of the reference object

3.1 Type of the Reference Object

Dependent on the type of the potential reference object are the types of applicable spatial relations. The location of objects in large scale spaces for example are mainly described by using cardinal relations like *"Europe is north of Africa"*, whereas in small scale spaces terms of projective relations are more appropriate: *"The car stands in front of the house"*. For some objects, e.g. buildings, both projective and cardinal relations are permissible in certain situations. If the answer is generated using a deictic or intrinsic frame of reference, projective relations are used. However, it is nevertheless possible to use a cardinal relation, as in: *"The church is north-west of the City Hall"*. The categorization consists, therefore, of the three classes shown in Table 1: the first and the second class deny the use of cardinal respectively projective relations and the third class permits all subclasses.

Using this knowledge the number of spatial relations to be tested can be reduced before their applicability is computed.

3.2 Type of the Localization

The type of the localization depends on the context in which the question is raised. Two types can be observed: approximate and precise localizations. For example, in the question: *"Where is the post office?"* people at first search for approximate localizations using topological relations like: *"The post office is near the City Hall"*. If this is not sufficient exact for the addressee, projective or cardinal relations are used for a more precise localization. There are two reasons for this behavior: First, topological relations are easier to apply, e.g. no frame of reference has to be established, and second, the approximate localization of a salient reference object is used to determine if this object is known to the addressee. If the object is known then a more detailed description can be added, if necessary.

The class of relations used for approximate localizations are topological, like *at, near, in*, etc. Relations which contain orientational information, like projective or cardinal relations, are more precise. However, in an answer to a spatial query about the location of a certain object using the latter relations, the information of proximity is nevertheless included, because a chosen reference object is required to be close to the located object.

In summary, the type of a particular object localization can be classified in an approximate or a precise localization in advance, which yields a further restriction of the possible set of applicable spatial relations.

4 Examples

In this section we will give some examples of how preprocessing affects the performance of the computation of spatial relations in object localization tasks. The number of spatial relations for each group in the CSR-3D system is as follows:

- 8 topological relations
- 12 projective relations (internal and external)
- 24 composite relations
- 16 cardinal relations

We now examine four different but general cases in which the two preprocessing factors help to decrease the number of relations to be tested.

```
USER> Where is the phone book?
SYSTEM> It is on the desk.

USER> Where exactly is the phone book?
SYSTEM> It is on the left side of the desk.

USER> Where is Bavaria located?
SYSTEM> Bavaria is located in Germany.

USER> Where exactly is Bavaria located?
SYSTEM> Bavaria is located in the south-east
        of Germany.
```

In Table 2 the number of spatial relations to be tested are listed with respect to no preprocessing (pp), to the type of RO preprocessing, to the type of localization preprocessing, and to total preprocessing. The example shows a decrease to 60% of the original set of relations in the worst case, and an average improvement down to 28%, which is less than a third of the original number of relations.

case	total no pp	type of RO	type of local.	total with pp	%
1	60	44	8	8	13%
2	60	44	52	36	60%
3	60	24	8	8	13%
4	60	24	52	16	27%

Table 2: Relations to test in the examples with and without preprocessing.

5 Conclusion

After identifying the object of interest, the task of object localization can be divided up into two major parts: the selection of an adequate reference object and the processing of spatial relations between the objects. The number of potential reference object candidates can be large, as can the number of spatial relations to be tested. After extending the computational model for spatial relations, presented in [Gapp, 1994a; 1994b], to internal projective and cardinal relations, it was shown that the set of relations to be tested can be reduced in two respects: the type of the reference object and the type of the localization. The reduction yields a minimal set of relations to be tested and therefore increases the overall performance of object localization processing substantially.

References

[Abella and Kender, 1993] A. Abella and J. R. Kender. Qualitatively describing objects using spatial prepositions. In *Proc. of AAAI-93*, pages 536–540, Washington, DC, 1993.

[André *et al.*, 1988] E. André, G. Herzog, and T. Rist. On the simultaneous interpretation of real world image sequences and their natural language description: The system soccer. In *Proc. of the 8th ECAI*, pages 449–454, Munich, 1988.

[André *et al.*, 1989] E. André, G. Herzog, and T. Rist. Natural language access to visual data: Dealing with space and movement. In F. Nef and M. Borillo, editors, *Proc. of the 1st Workshop of Logical Semantics of Time, Space and Movement in Natural Language.* Hermès, 1989.

[Egenhofer and Herring, 1990] M. J. Egenhofer and J. R. Herring. A mathematical framework for the definition of topological relationships. In K. Brassel and H. Kishimoto, editors, *Proc. of the 4th International Symposium on Spatial Data Handling*, pages 803–813. Zürich, 1990.

[Gapp, 1994a] K.-P. Gapp. Basic meanings of spatial relations: Computation and evaluation in 3d space. In *Proc. of AAAI-94*, pages 1393–1398, Seattle, WA, 1994.

[Gapp, 1994b] K.-P. Gapp. A computational model of the basic meanings of graded composite spatial relations in 3d space. In M. Molenaar and S. De

Hoop, editors, *Proc. of the AGDM'94 Workshop*, pages 66–79, Delft, The Netherlands, 1994. Publications on Geodesy, Netherlands Geodetic Commission.

[Gapp, 1994c] K.-P. Gapp. From vision to language: A cognitive approach to the computation of spatial relations in 3d space. In *Proc. of the 1st European Conf. on Cognitive Science in Industry*, pages 339–358, Luxembourg, 1994.

[Gapp, 1995a] K.-P. Gapp. Angle, distance, shape, and their relationship to projective relations. In *Proc. of the 17th Annual Conference of the Cognitive Science Society*, Pittsburgh, PA, 1995. To appear.

[Gapp, 1995b] K.-P. Gapp. An empirically validated model for computing spatial relations. In *Proc. of KI-95*, Berlin, Heidelberg, 1995. Springer. To appear.

[Gapp, 1995c] K.-P. Gapp. Object localization: Selection of optimal reference objects. In *Proc. of the 2nd int. Conference on Spatial Information Theory*, Berlin, Heidelberg, 1995. Springer. To appear.

[Herskovits, 1986] A. Herskovits. *Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English.* Cambridge University Press, Cambridge, London, 1986.

[Koller, 1992] D. Koller. *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen.* Infix, St. Augustin, 1992.

[Mukerjee and Joe, 1990] A. Mukerjee and G. Joe. A qualitative model for space. In *Proc. of AAAI-90*, pages 721–727, Boston, MA, 1990.

[Olivier *et al.*, 1994] P. Olivier, T. Maeda, and J. Tsuji. Automatic depiction of spatial descriptions. In *Proc. of AAAI-94*, pages 1405–1410, Seattle, WA, 1994.

[Rajagopalan, 1994] R. Rajagopalan. A model for integrated qualitative spatial and dynamic reasoning about physical systems. In *Proc. of AAAI-94*, pages 1411–1417, Seattle, WA, 1994.

[Schirra and Stopp, 1993] J. R. J. Schirra and Eva Stopp. Antlima – a listener model with mental images. In *Proc. of the 13th IJCAI*, pages 175–180, Chambery, France, 1993.

[Talmy, 1983] L. Talmy. How language structures space. In H. Pick and L. Acredolo, editors, *Spatial Orientation: Theory, Research and Application*, pages 225–282. Plenum, New York, London, 1983.

[Wazinski, 1992] P. Wazinski. Generating spatial descriptions for cross-modal references. In *3rd Conference on Applied Natural Language Processing, Trento, Italy*, pages 56–63. Association for Computer Linguistics, 1992.