Augmentation of Explicit Spatial Configurations by Knowledge-Based Inference on Geometric Fields

Dan TAPPAN

College of Engineering, Idaho State University
921 S. 8th Ave., Stop 8060
Pocatello, ID 83209-8060, USA

ABSTRACT

A spatial configuration of a rudimentary, static, real-world scene with known objects (animals) and properties (positions and orientations) contains a wealth of syntactic and semantic spatial information that can contribute to a computational understanding far beyond what its quantitative details alone convey. This work presents an approach that (1) quantitatively represents what a configuration explicitly states, (2) integrates this information with implicit, commonsense background knowledge of its objects and properties, (3) infers additional, contextually appropriate, commonsense spatial information from and about their interrelationships, and (4) augments the original representation with this combined information. A semantic network represents explicit, quantitative information in a configuration. An inheritance-based knowledge base of relevant concepts supplies implicit, qualitative background knowledge to support semantic interpretation. Together, these structures provide a simple, nondeductive, constraint-based, geometric logical formalism to infer substantial implicit knowledge for intrinsic and deictic frames of spatial reference.

Keywords: Spatial Knowledge Representation, Reasoning, and Generation.

INTRODUCTION

In most graphics applications, object positions and orientations are represented quantitatively as coordinates and angles, respectively. While this form is very effective at explicitly describing where they are for computational purposes, it neglects the implicit, qualitative relations that people can easily infer and understand; e.g., *the dog, which is near the tree, is facing the cat*. The goal of this work is to take minimal quantitative spatial information and infer unstated relationships to generate additional, qualitative knowledge about a scene. The approach provides a flexible, configurable, knowledge-based framework for geometric constraint satisfaction that can be useful in many applications, especially in artificial intelligence and natural-language processing.

BACKGROUND

Properly interpreting three frames of spatial reference plays a critical role [15, 4, 22, 11, 12, 2]. The intrinsic (or object-centered) frame generally applies to objects that have a canonical front; e.g., *in front of the dog* means some position extending along its orientation axis. The extrinsic (or environment-centered) and deictic (or viewer-centered) frames are generally the opposite case for objects without a canonical front; e.g., *in front of the tree* means extending from it to another position in the world that establishes a virtual front. In the extrinsic frame, this reference position is arbitrary; e.g., *in front of the tree as seen from the horse*. In the deictic frame, which is a specialized case of the extrinsic, it is the (usually implicit) position of the viewer; e.g., *in front of the tree (as seen by the viewer in the south looking north)*. For space reasons, this paper considers only the intrinsic and deictic frames.

This work directly extends and complements previous work by Tappan [31, 32] as the counterpart to generating spatial layouts of objects based on natural-language descriptions. Its approach to inferring spatial knowledge loosely draws upon other work by Neumann [17], Walter et al. [36], Koller et al. [19], and Tsotsos [33] for scene interpretation. Tversky [34] covers in comprehensive detail many of the spatial issues that complicate the problem. Herskovits [15], Claus et al. [4], and Olivier and Tsujii [22], in particular, form the basis for defining and interpreting spatial frames of reference. Most early approaches to spatial analysis adopted purely geometric solutions and did not take advantage of spatial knowledge relevant to the objects [38, 40]. More recent work, especially in Geographic Information Systems, attempts to account for such contextual information [24, 6, 8, 9, 10, 14, 25]. This work follows the latter approach.

EXPLICIT SPATIAL REPRESENTATION

Stage 1 involves quantitatively representing what is explicitly known about a spatial configuration. A configuration is a collection of objects that individually have the numeric properties of a two-dimensional center position (*x*,*y* in meters, increasing from west to east and south to north, respectively) and an orientation (in degrees, with 0 as north). Configurations can come from many sources depending on the application. This work is linguistically motivated, so it generally derives them from handcrafted English sentences. For example, either sentence (2a) or (2b) can entail the same configuration of `shelby.position=(1,3)`, `shelby.orientation=0`, and `tree.position=(1,5)`. Sentence (1) does not define any spatial information, but it does provide important contextual information for understanding what *Shelby* is.

1. *Shelby is a retriever.*

2a. *The tree is north of Shelby, and she is facing north.*
2b. *The tree is north of Shelby, and she is facing the tree.*

The world in which objects reside is a static, two-dimensional, tabletop zoo environment. It currently supports 108 unique, non-articulated kinds of objects, mostly animals and plants, that were selected because they exhibit great variety in their spatial characteristics and interpretations [31]. It is straightforward to add others. The static aspect eliminates the effects of movement, time dependencies, the frame problem, etc., which are well beyond the scope of this work [1, 27, 29, 5].

The underlying representation of a configuration is a simple semantic network, which is particularly suited to this task for three reasons [27]. First, its primary components (nodes and directional arcs) map directly to the objects and properties in a configuration, respectively, and to the relations that will later augment it. For example, Figure 1 is a semantic network that derives from either (2a) or (2b). The object *world-center* (*wc*) is automatically generated at the origin (0,0) to facilitate global position references like *in the north*, etc.

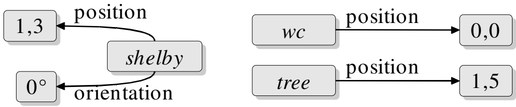

Figure 1: Semantic Network

Second, as a straightforward computational data structure, all standard graph algorithms can operate on it natively. Third, as a well-studied and commonly used formalism for artificial intelligence, it facilitates transferring knowledge representations to and from other related applications [26, 28].

IMPLICIT STATIC INFERENCE

Stage 2 involves deriving the unstated attributes and rules that implicitly describe the spatial relations that could apply to each object in the configuration; e.g., *Shelby is south of the tree*. This form of inference is static in the sense that it considers each object in isolation, not in context with the other objects [6].

Knowledge Representation

Despite its name, the semantic network explicitly represents only the syntax (or structure) of the configuration without any consideration of its real-world semantics (or meaning). To understand the semantics even superficially requires deeper analysis into what the objects are and how their rules apply to them in context [7].

The source of the implicit, commonsense background knowledge for this analysis is a simple knowledge base that is similar to an inheritance hierarchy in object-oriented programming [16]. It currently contains 108 prototypical concepts (or classes), each of which either inherits its attributes and rules for spatial interpretation from its ancestors, or it defines/overrides them itself. Only single inheritance is supported; in principle, multiple inheritance could provide a richer, more compact representation, but the additional complexity, especially for conflict resolution, is currently not justifiable. A simplified example appears in Figure 2.

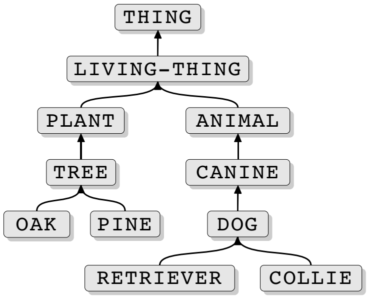

Figure 2: Knowledge Base

An attribute defines whether a concept exhibits a particular spatial behavior. The only one found to be necessary in this paper specifies whether a concept has a canonical front, which generally corresponds to its having a face or eyes. As objects and concepts are not articulated, any head is always fixed in line with the orientation of the body. This simplification eliminates the need to determine the configuration of body parts; e.g., the body of the dog is oriented north, but it is looking east.

A rule specifies when a particular relation, like `near`, applies from one object to another. It uses a formalism of geometric fields that describe a collection of cells in a two-dimensional, top-view, polar projection centered around the object [39, 40, 12, 22, 11]. Experimentation suggests that 32 sectors and 100 rings of the form in Figure 3 are sufficient for the current domain of concepts and relations. Each cell defines a small subregion of the projection that can be conditionally inspected for the presence of other objects. The implementation in this paper does not account for the dimensions of an object, so this check is based only on its center position as a point source.

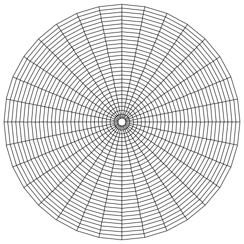

Figure 3: Available Cells

Although any combination of selected cells among the 3,200 available is valid, in practice, only minor variations of two types define all spatial relations in this work: *wedges* apply to position and orientation relations, and *rings* to distance relations. Figures 4a and 4b show respective examples of the relations `front-of` and `far-from` for object c_1, which is facing the direction of the arrow.

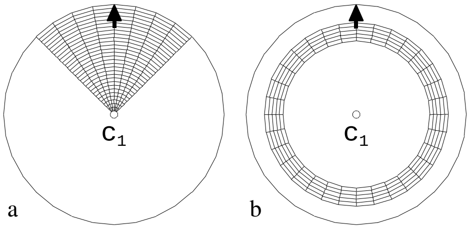

Figure 4: Sample Wedge and Ring Fields

Each concept in the knowledge base has access to the attributes and rules that are appropriate for its spatial interpretation. In particular, these rules define the 77 distance, orientation, and position relations in Tables 1 through 3, respectively. For space reasons, Tables 2 and 3 omit for each relation a prefixed variant `direct`, which specifies a narrower interpretation with the same general meaning; e.g., `direct-front-of` would fan out less to the sides. The interpretation of appropriateness depends on certain *ad hoc* generalities of the concept, which depend on the application of this work. For example, the relation `near` is closer (in absolute terms) for a mouse than it is for an elephant due to their differences in magnitude [13, 22, 30]. Many relations in Table 3 have both `local` and `global` forms, which respectively apply in the intrinsic and deictic frames of reference.

#	Relation	#	Relation
26	`inside`	30	`midrange-from`
27	`outside`	31	`far-from`
28	`adjacent-to`	32	`at-fringe-of`
29	`near`		

Table 1: Distance Relations

#	Relation	#	Relation
33	`facing`	38	`facing-west`
34	`facing-away-from`	39	`facing-northeast`
35	`facing-north`	40	`facing-northwest`
36	`facing-south`	41	`facing-southeast`
37	`facing-east`	42	`facing-southwest`

Table 2: Orientation Relations

#	Relation	#	Relation
1	`local-front-of`	14	`global-front-right-of`
2	`local-back-of`	15	`global-back-left-of`
3	`local-left-of`	16	`global-back-right-of`
4	`local-right-of`	17	`between`
5	`local-front-left-of`	18	`north-of`
6	`local-front-right-of`	19	`south-of`
7	`local-back-left-of`	20	`east-of`
8	`local-back-right-of`	21	`west-of`
9	`global-front-of`	22	`northeast-of`
10	`global-back-of`	23	`northwest-of`
11	`global-left-of`	24	`southeast-of`
12	`global-right-of`	25	`southwest-of`
13	`global-front-left-of`		

Table 3: Position Relations

The final element of this stage combines the explicitly stated information from the semantic network with the implicitly inferred background knowledge from the knowledge base. Figure 5 depicts a simplified example of this process: Objects *Shelby* and *tree* link to concepts `RETRIEVER` and `TREE`, respectively. Through inheritance, *Shelby* derives the rules about her ancestor concepts `DOG`, `CANINE`, `ANIMAL`, `LIVING-THING`, and `THING`. The same process holds for *tree*. It is important to note the distinction between an *object*, which is a unique instance in the configuration, and a *concept*, which is a shared set of attributes and rules that all instances of it must have in common. For clarity, this distinction is rendered typographically through italics and capitalized typewriter font, respectively.

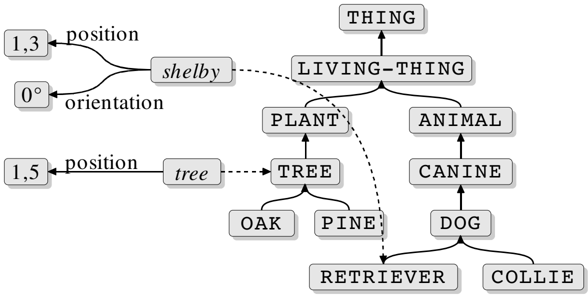

Figure 5: Semantic Network Linked to Knowledge Base

IMPLICIT DYNAMIC INFERENCE

Stage 3 involves determining which rules apply among the candidates derived in Stage 2. This form of inference is dynamic in the sense that it considers each object pairwise in context with all the other objects [6]. It is a nondeductive reasoning mechanism because it uses only straightforward, geometric constraint satisfaction as the logical foundation [23, 35, 21].

For any object o_1 with a canonical-front attribute, the default frame of reference is intrinsic; i.e., any o_2 in front of o_1 is roughly in line with the direction o_1 is facing. A field reflects this spatial behavior by rotating itself so that its arrow aligns with the orientation of o_1. Thus, its front field aligns with this direction, and its back, left, and right fields respectively align 180˚, –90˚, and +90˚ from it. In contrast, for any object o_1 without a canonical front, the only possible frame of reference is deictic; i.e., any o_2 in front of o_1 is in line between o_1 and the implicit viewer, who by default resides in the south-center of the world and looks north. In this case, the arrow aligns to the position of the viewer. Finally, all concepts support absolute compass orientations for relations like `north-of`, `south-of`, etc. In this case, the arrow always aligns north.

Inference Generation

Exhaustive inspection of every pairwise combination of objects produces a list of relations that the semantic network logically entails [4, 37]. The following pseudocode outlines this process:

```
for each object o in semantic network S
  position and orient o from its properties
for each object $o_1$ in S
  for each object $o_2$ in S where $o_2 \neq o_1$
    for each relation r in Tables 1,2,3 …
    contextually applicable for pairing $o_1 r o_2$
      if $o_2$.position is in r.field, …
      add r from $o_1$ to $o_2$ in S
```

This approach augments the purely quantitative representation in Figure 1 with qualitative spatial inferences to produce the augmented semantic network in Figure 6.

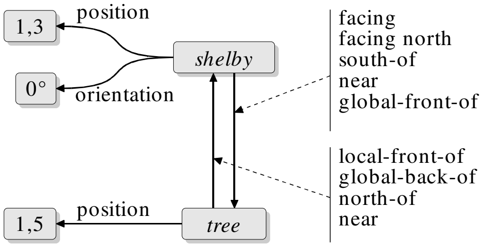

Figure 6: Augmented Semantic Network[1]

Figure 7 depicts an extended example with a configuration containing a tree *T*, a zebra *Z*, and a giraffe *G*, plus world-center *W*. The arrows on *Z* and *G* indicate their orientation.

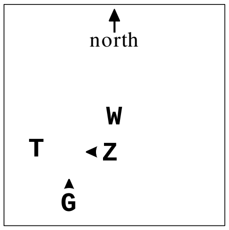

Figure 7: Sample Configuration

These four objects produce the 46 (non-unique) inferences in Table 4, where the numbers correspond to the relations in Tables 1 through 3. The corresponding augmented semantic network for even such a simple configuration is too rich and intertwined to depict here.

		o_2			
	r	G	T	Z	W
	G		14, 24 30, 35	13, 25 30, 35	13,25 31,35
o_1	T	5,15 23,30		1,11 21,30	13,21 31
	Z	6,16 22,30,38	12,20 30,33,38		9,19 29,38
	W	16,22 31	12,20 31	10,18 29	

Table 4: Relationship Inference Matrix

RESULTS AND DISCUSSION

The results were evaluated based on the numerical relationship between the number of objects stated in the input and the number of relations inferred in the output. This relationship is very sensitive to the position and orientation of each object, so a stochastic experiment was performed to demonstrate average results. Independent tests were conducted on 3 through 10 objects (including *world-center*), which referenced the same concepts. Each test was divided into 10,000 independent runs, in which the objects were randomly configured. For each run, the number of unique generated inferences was recorded. Figure 8 shows the average for each test, which varied from 27 for 3 objects to 602 for 10 objects.

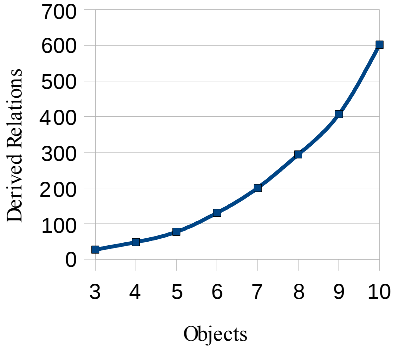

Figure 8: Derived Relations

This relatively simple, straightforward, flexible approach to augmenting explicit spatial configurations is clearly effective for mainstream, accepted interpretations of basic objects in two dimensions. It generates a substantial number of potentially useful spatial inferences in acceptable time complexity: $\Theta(ro^2)$, where *r* is the number of candidate relations to evaluate, and *o* is the number of objects in a configuration. Human cognitive limits usually dictate $o < 6$ in linguistic configurations, so, in practice, the scalability of this approach is not actually a factor for such applications [18].

[1] For space reasons, *world-center* is omitted.

Frame of spatial reference was found to have the following effects on relations in this work:

- Position relations distinguish between intrinsic and deictic interpretations with the respective prefixes `local` and `global`. Both may apply simultaneously because a deictic interpretation always accompanies an intrinsic one; the converse, however, does not hold. For example, *the tree is in front of the dog*, from the intrinsic perspective of the dog, means in line with the orientation of the dog, whereas from the deictic perspective of the viewer, it means between the dog and the viewer, irrespective of the orientation of the dog.
- Orientation relations apply only in an intrinsic interpretation. For example, *the dog is facing the cat* is valid, but *the tree is facing the cat* is not, because dog has a canonical front, and tree does not.
- Distance relations apply in any interpretation. For example, in both *the dog is near the tree* and *the dog is near the cat*, the orientations of the tree and cat are irrelevant to the relation `near`.

FUTURE WORK

Two categories of related work are under consideration. The first extends the current domain of relations to compare the dimensions of objects; i.e., contextually larger, equally sized, or smaller with respect to height, width, and depth. The inference stages are identical, but the rule formalism of geometric fields does not translate easily. The second category has two aspects that involve post-processing the augmented semantic network. The first involves identifying latent pragmatics of scenarios to aid in scene recognition and gisting [20]. For example, a spatial configuration that describes a tiger facing a zebra from behind a nearby tree may imply an impending attack. The second aspect involves generating natural-language descriptions and summaries from an augmented spatial configuration. In crude, verbose form, this capability already exists, as Table 4 shows. However, it lacks a mechanism for determining salience (i.e., what to say and what to omit), and for planning and realizing acceptable text (i.e., how to say it) [3].

REFERENCES

[1] G. Adorni, M. Di Manzo, and F. Giunchiglia, “Natural Language Driven Image Generation”, in **Proceedings of COLING-84**, 1984, pp. 495-500. Stanford, CA.

[2] E. André, G. Bosch, G. Herzog, and T. Rist, “Coping with the Intrinsic and the Deictic Uses of Spatial Prepositions”, in J. Jorrand and L. Sgurev, eds., **Artificial Intelligence II: Methodology, Systems, Applications**, 1987, pp. 375-382, North-Holland, Amsterdam.

[3] J. Bateman, “Upper Modeling: A general organization of knowledge for natural language processing”, in **Proceedings of Standards for Knowledge Representation Systems**, 1990, Santa Barbara, CA.

[4] B. Claus, K. Eyferth, C. Gips, R. Hörnig, U. Schmid, S. Wiebrock, F. Wysotzki, “Reference Frames for Spatial Inference in Text Understanding”, in C. Freksa, C. Habel, and K. Wender, eds., **Spatial Cognition—An interdisciplinary approach to representing and processing spatial knowledge**, 1988, pp. 214-226.

[5] B. Coyne and R. Sproat, “WordsEye: An Automatic Text-to-Scene Conversion System”, in **Proceedings of SIGGRAPH-01**, 2001, pp. 487-496, Los Angeles, CA.

[6] E. Davis, **Representations of Commonsense Knowledge**. Morgan Kaufmann, San Mateo, CA, 1990.

[7] R. Davis, H. Shrobe, and P. Szolovits, “What is Knowledge Representation?” **AI Magazine**, Vol. 14, 1993, pp. 17-33.

[8] M. Egenhofer and R. Franzosa, “Point-Set Topological Spatial Relations”, **International Journal of Geographical Information Systems**, Vol. 5, No. 2, 1991, pp. 161-174.

[9] A. Frank, “Qualitative Reasoning about Distances and Directions in Geographic Space”, **Journal of Visual Languages and Computing**, Vol. 3, No. 4, 1992, pp. 343-371.

[10] A. Frank, “Qualitative Spatial Reasoning: Cardinal Directions as an Example”, **International Journal of Geographical Information Systems**, Vol. 10, No. 3, 1996, pp. 269-290.

[11] C. Freska, “Using Orientation Information for Qualitative Spatial Reasoning”, in A. Frank, I. Campari, and U. Formentini, eds., **Theories and Methods of Spatio-Temporal Reasoning in Geographic Space**, LNCS 639, Springer-Verlag, Berlin, 1992.

[12] K. Gapp, “Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space”, in **Proceedings of AAAI-94**, 1994, pp. 1393-1398, Seattle, WA.

[13] D. Hernández, **Qualitative Representation of Spatial Knowledge**, Springer-Verlag, Berlin, 1994.

[14] D. Hernández, E. Clementini, and P. Di Felice, “Qualitative Distances”, in A. Frank and W. Kuhn, eds., **Third European Conference on Spatial Information Theory**, 1995, pp. 45-58, Semmering, Austria.

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

[16] K. Mahesh, **Ontology Development for Machine Translation: Ideology and Methodology**, Technical Report MCCS-96-292, Computing Research Laboratory, New Mexico State University, 1996.

[17] B. Neumann, “Natural Language Description of Time-Varying Scenes”, in D. L. Waltz, ed., **Semantic Structures: Advances in Natural Language Processing**, 1989, pp. 167-207, Lawrence Erlbaum, Hillsdale, NJ.

[18] P. Johnson-Laird, **Mental Models**, Cambridge, Harvard University Press, 1983.

[19] H. Koller, N. Heinze, and H.-H. Nagel, “Algorithmic Characterization of Vehicle Trajectories from Image Sequences by Motion Verbs”, in **Proceedings of IEEE Conference on Computer Vision and Pattern Recognition**, 1992, pp. 90-95, Maui, Hawaii.

[20] H. Liu and P. Singh, “ConceptNet—a practical commonsense reasoning tool-kit”, **BT Technology Journal**, Vol. 22, No. 4, 2004, pp. 211-226.

[21] A. Mukerjee, “Neat vs Scruffy: A Survey of Computational Models for Spatial Expressions”, in P. Olivier and K. Gapp, eds., **Computational Representation and Processing of Spatial Expressions**, 1998.

[22] P. Olivier, and J. Tsujii, “A computational view of the cognitive semantics of spatial prepositions”, in **Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics**, 1994, Las Cruces, NM.

[23] D. Papadias and M. Kavouras, “Acquiring, Representing and Processing Spatial Relations”, in **Proceedings of 6th International Symposium on Spatial Data Handling**, 1994, Edinburgh, U.K.

[24] S. Peters and H. Shrobe, “Using Semantic Networks for Knowledge Representation in an Intelligent Environment”, in **Proceedings of PerCom '03: 1st Annual IEEE International Conference on Pervasive Computing and Communications**, 2003, Ft. Worth, TX.

[25] D. Randell, Z. Cui, and A. Cohn, “A Spatial Logic based on Regions and Connection”, in **Proceedings of 3rd International Conference on Knowledge Representation and Reasoning**, 1992, pp. 165-176, San Mateo, CA.

[26] S. Russell and P. Norvig, **Artificial Intelligence: A Modern Approach**. Prentice Hall, Upper Saddle River, NJ, 1995.

[27] J. Sowa, ed., **Principles of Semantic Networks: Explorations in the Representation of Knowledge by Computers**. New York, Academic Press, 1991.

[28] J. Sowa, **Knowledge Representation: Logical, Philosophical, and Computational Foundations**, Brooks/Cole, Pacific Grove, CA, 2000.

[29] R. Srihari, “Computational Models for Integrating Linguistic and Visual Information: A Survey”, **Artificial Intelligence Review**, Vol. 8, 1994, pp. 349-369.

[30] A. Stevens and P. Coupe, “Distortions in Judged Spatial Relations”, **Cognitive Psychology**, Vol. 13, 1978, pp. 422-437.

[31] D. Tappan, “Knowledge-Based Spatial Reasoning for Automated Scene Generation from Text Descriptions”, Ph.D. dissertation, New Mexico State University, 2004.

[32] D. Tappan, “Knowledge-Based Spatial Constraint Satisfaction”, in **Proceedings of Florida Artificial Intelligence Research Society International Conference**, 2004, Miami Beach, FL.

[33] J. Tsotsos, “Knowledge Organization and its Role in Representation and Interpretation for Time-Varying Data: the ALVEN System”, **Computational Intelligence**, Vol. 1, 1985, pp. 16-32.

[34] B. Tversky, “Levels and structure of spatial knowledge”, in **Cognitive Mapping: Past, present and future**, R. Kitchin and S. Freundshuh, eds. London and New York, Routledge, 2000.

[35] J. Ullman, **Principles of Data Bases and Knowledge Base Systems, Vol. 1**, Computer Science Press, 1988.

[36] I. Walter, P. Lockemann, and H-H. Nagel, “Database Support for Knowledge-Based Image Evaluation”, in P. M. Stocker, W. Kent, R. Hammersley, eds., **Proceedings of the 13th Conference on Very Large Databases**, 1988, pp. 3-11, Brighton, UK.

[37] S. Wiebrock, L. Wittenburg, U. Schmid, and F. Wysotzki, “Inference and Visualization of Spatial Relations”, in **Lecture Notes on Computer Science**, No. 1849, 2000, pp. 212.

[38] K. Xu, J. Stewart, and E. Fiume, “Constraint-Based Automatic Placement for Scene Composition”, in **Proceedings of the Conference on Human-Computer Interaction and Computer Graphics**, 2002, pp. 25-34, Calgary, Canada.

[39] A. Yamada, T. Yamamoto, H. Ikeda, T. Nishida, and S. Doshita, “Reconstructing Spatial Image from Natural Language Texts”, in **Proceedings of COLING-92**, 1992, pp. 1279-1283, Grenoble, France.

[40] A. Yamada, “Studies on Spatial Description Understanding Based on Geometric Constraints Satisfaction”, Ph.D. dissertation., University of Kyoto, 1993.