Spatial Relations for Perceptual Anchoring

Jonas Melchert and **Silvia Coradeschi** and **Amy Loutfi**[1]

Abstract. In this work we show how a mobile robot can use spatial information of objects to improve communication with humans and other devices located in an intelligent environment. In particular, this work focuses on using spatial relations to facilitate the creation of a connection between symbolic and perceptual representation that refer to the same physical object (anchoring). We extend an anchoring framework to include a set of binary spatial relations which can then be used to exchange information about objects with a human user. To illustrate the performance of the framework, a number of scenarios are presented using a mobile robot. These scenarios are a first step towards the goal of having mobile robots integrated in an intelligent environment and communicating with human users.

1 INTRODUCTION

An emerging trend in the field of robotics is the notion of *symbiotic robotic systems* which consists of a robot, human and (smart) environment cooperating together in performing different tasks [4]. By assisting the robot with information provided by the human or smart objects, some of the current challenges in robotics can be circumvented. For instance, localisation of the robot can be done with a system of surveillance cameras and object recognition tasks can be assisted by passive technologies like RFID. Human assistance and cooperation can also be used to provide instructions to the robot and to assist the robot in case of failure or ambiguous situations when several choices are possible. The motivation behind the symbiotic system is the integration of robotics into everyday life. Therefore, it is essential to allow a range of different users to be able to communicate to the system, this range should include both expert users and even bystanders.

A natural form of communication between humans and the robots is natural language dialogue. In a system where a human provides assistance to the robot it is most convenient for the human to communicate to the robot using dialogue, particularly in the case of a non-expert interacting with the robot. Among the many challenges that this task presents, in this paper, we concentrate on the correspondence that must necessarily exist between the linguistic symbols used by a human and the sensor data perceived by the robot. We call *anchoring* the process of creating and maintaining over time the connection between the symbols and the corresponding perceptual representation that refer to the same physical objects. Already in the field of robotics, anchoring has been explored in systems that use planning and a variety of sensing modalities (e.g. vision and olfaction) [2, 10]. In this paper we examine the possibility to integrate the anchoring framework in a symbiotic robotic system. In particular, we focus on the inclusion of spatial relations in the anchoring framework for the purpose of human-robot communication via language.

[1] AASS Mobile Robotics Lab, Örebro University, Sweden
e-mail: jonas.melchert@aass.oru.se, web: www.aass.oru.se

To accomplish this task, we extend our existing framework [3] to include a set of binary spatial relations; “at”, “near”, “left”, “right”, “in front”, and “behind” for 2D space. As spatial prepositions are inherently rather vague, a technique using fuzzy sets is applied to define graded spatial relations. The proposed method computes a spatial relations-network for anchored symbols and stores that in the anchors. The relations are then used to assist the robot in resolving ambiguities, identifying objects and improving general task performance of the anchoring framework.

This paper is organised as follows: Section 2 summarises related work on spatial relations and perceptual anchoring. In Sections 3 and 4 we detail the perceptual anchoring and the designed spatial relations used in this work. Section 5 describes some initial experimental scenarios and future work. Section 6 gives a conclusion.

2 RELATED WORK

Most of the work on spatial relations is concerned with connecting the visual domain with the verbal domain of humans. Gapp [6] describes a computational model to compute and evaluate graded spatial relations in 3D space for a visual scene description generator. Objects are approximated by their centre of gravity and bounding rectangle, since only the object’s location is required for the applicability of the spatial relation. The semantics of the relations are defined by evaluation functions depending on the proximal distance and orientation angle between a reference object and the object to be located. Abella and Kender [1] present a system that qualitatively describes the spatial layout of objects with binary relations, from a birds eye view. To account for the vagueness of spatial prepositions, they apply a fuzzyfication technique and use a threshold to decide if two objects are no longer describable by a given preposition. Our computational model for the evaluation of spatial relations is mainly based on the one presented by Gapp, and we apply a thresholding function to select relevant relations.

Work that deals with the abstraction of spatial information from sensory data on robotic platforms is e.g. the one by Skubic et al. [14]. They use a more complex computational model based on the “histogram of forces”. Their system generates linguistic expressions that describe spatial relations between a mobile robot and its environment, based on range readings from a ring of SONAR sensors. Luke at al. [11] present a stereo vision system that can generate linguistic spatial relations for 3D scenes, adopting a fuzzy-set approach and the above mentioned histogram of forces. Hois et al. [9] describe an object recognition system based on 3D LASER scans. The recognition process is supported by interaction with the user and ontological deduction. Unidentified objects can be labelled by the user, using a speech interface, or are classified through the designed domain ontology. In a subsequent phase, the user can query the system for scene descriptions, involving spatial relations to specify object locations.

We would like to investigate a similar approach, the exploitation of semantic knowledge, in our future work.

In the above examples the human-robot interaction (HRI) is limited to a (conventional) "master-slave" mode of communication, but our interest is to enable the robot to make use of humans in order to compensate for perceptual or cognitive deficiencies. A good example in this line of thought is the "Peer-to-Peer Human-Robot Interaction" project [5], that aims to develop a range of HRI techniques so that robots and humans can work together in teams and engage in task-oriented dialogue. One of the key components are computational cognitive models for human space perception and spatial reasoning. Of more practical relevance is the work by Moratz and Tenbrink, e.g. [12], that deals with the use of spatial language in human-robot communication. They describe a computational model for a mobile robot platform with a visual object recognition system. The model is evaluated in a number of experiments with uninformed users instructing the robot in spatial identification tasks. Their results provide hints for possible communication scenarios and the employed communication strategies and spatial reference systems, that we will consider in our dialogue system.

So far, the use of spatial relations for anchoring has not been studied in detail. Earlier work [2], investigating the use of planning techniques to recover from perceptual failures and ambiguous cases in perceptual anchoring, incorporated a simple means to refer to an object by specifying its relations to other anchored objects, but only the relations "at" and "near" where supported, and computed on-the-fly using a simple and crisp computational model.

3 PERCEPTUAL ANCHORING

As described in the introduction, the task of anchoring is to create and maintain in time the correspondence between symbols and percepts that refer to the same physical object. This correspondence is reified in a data structure $\alpha(t)$, called an *anchor*. It is indexed by time as the perceptual system continuously generates new percepts; and the created links are dynamic, since the same symbol may be connected to new percepts every time a new observation of the corresponding object is acquired. So at each time instance t, $\alpha(t)$ contains a symbol identifying that object, a percept generated by the latest observation of the object, and a perceptual signature meant to provide the (best) estimate of the values of the observable properties of the object. See figure 1 for a graphical illustration. Following [3] the main parts of anchoring are:

- A *symbol system*, including a set $\mathcal{X} = \{x_1, x_2, \ldots\}$ of individual symbols (variables and constants), a set $\mathcal{P} = \{p_1, p_2, \ldots\}$ of predicate symbols, and an inference mechanism whose details are not relevant here.
- A *perceptual system*, including a set $\Pi = \{\pi_1, \pi_2, \ldots\}$ of possible percepts, a set $\Phi = \{\phi_1, \phi_2, \ldots\}$ of attributes, and perceptual routines whose details are not relevant here. A percept is a structured collection of measurements assumed to originate from the same physical object; an attribute ϕ_i is a measurable property of percepts with values in the domain $D(\phi_i)$. Let $D(\Phi) = \bigcup_{\phi \in \Phi} D(\phi)$.
- A *predicate grounding relation*, $g \subseteq \mathcal{P} \times \Phi \times D(\Phi)$, which embodies the correspondence between (unary) predicates and values of measurable attributes. The relation g maps a certain predicate to compatible attribute values.

The following definitions allow to characterise objects in terms of their (symbolic and perceptual) properties:

- A *symbolic description* σ is a set of unary predicates from $\mathcal{P}$.
- A *perceptual signature* $\gamma : \Phi \mapsto D(\Phi)$ is a partial mapping from attributes to attribute values.

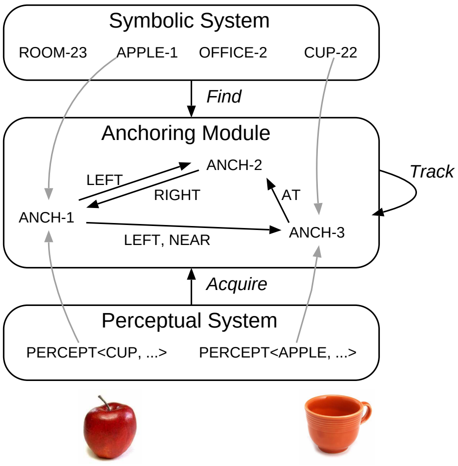

Figure 1. Graphical illustration of the anchoring framework: the anchoring module connects the perceptual and the symbolic systems in a physically embedded intelligent system. Spatial relations between anchored objects are maintained within the anchoring module.

The extension of the framework [3] presented in [10] allows the creation of anchors in both a top-down and a bottom-up fashion: bottom-up acquisition is triggered by recognition events from the sensory system when percepts can not be associated with existing anchors; top-down acquisition occurs when a symbol needs to be anchored to a perceptual description (such a request may come from a top-level planner). These functionalities are realised through:

- *Acquire*: creates a new anchor whenever a percept is received which currently does not match any existing anchor, and inserts symbolic information about the object and its properties into the planner's world model.
- *Find*: takes a symbol x and a symbolic description and returns an anchor α defined at time t (and possibly undefined elsewhere). If an existing anchor, created by *Acquire*, satisfies the symbolic description it selects one; otherwise it searches for matching percepts and, if one is found, creates an anchor for it. Matching of anchor or percept can be either partial or complete: it is partial if all the observed properties in the percept or anchor match the symbolic description, but there are some properties in the description that have not been observed.

At each update cycle of the perceptual system, when new perceptual information is received, it is important to determine if the new information should be associated to an existing anchor (data association problem). The following functionality addresses the problem of tracking objects over time:

- *Track*: takes an anchor α defined at $t - k$ and extends its definition to t. The track assures that the anchor's percept is the most recent and adequate perceptual representation of the object. This facilitates the maintenance of a stable representation of the world on a symbolic level.

By having an anchor structure maintained over time, the challenge is to determine if the association of new percepts is justified or whether certain anchors should be removed. According to [10], this is a difficult problem, because conceptually it is not clear when it is appropriate to remove anchors from the system. The current system adopts a simple solution in which objects that are not perceived when expected decrease in a "life" value. When the anchor has no remaining life, it is removed.

4 SPATIAL RELATIONS

For the computation and evaluation of basic spatial relations' meanings we follow the approach presented in [6] and apply it to 2D space. Two classes of binary spatial relations between a reference object $REFO$ and the object to be located LO (located object) are considered: the topological relations "at" and "near", and the projective relations "front of", "behind", "right", and "left". To model the vagueness of spatial prepositions, the evaluation of a spatial relation results in a degree of applicability in the interval $[0..1]$, representing the range between "not" and "fully" applicable, respectively.

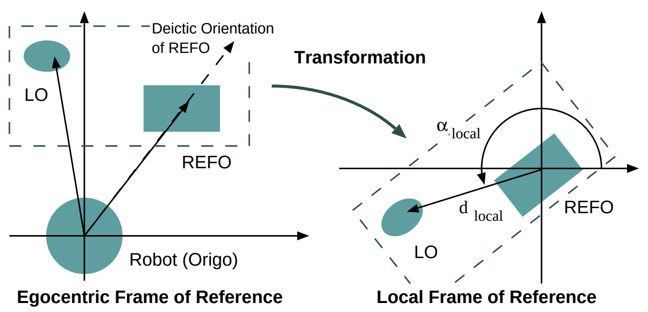

Figure 2. Frame of reference, and computation of distance and orientation angle. Objects are represented by their idealised point location.

4.1 Idealised Object Representation and Frame of Reference

In order to establish spatial relationships between anchored objects we need a geometrical representation of the objects. For the purpose of this work, we assume that the perceptual system provides the relative 2D position of objects with respect to the robot, stored in the perceptual signature. Objects are represented by an idealised point location, derived by projecting the object's centre of gravity (in the video image) onto the floor-plane. See Figure 2 for illustration.

An important aspect is the selection of an appropriate frame of reference [8] for the evaluation of spatial relations. No global frame of reference is used for the robot and therefor also not for the spatial relations. Instead we choose an egocentric frame of reference, as we consider this a more intuitive approach, especially with respect to an intended human-robot interaction (see future work, and [12]).

4.2 Topological Relations

The topological relations "at" and "near" both refer to a region proximal to an object. Following [6] their semantics is defined as: "at" localises an object in the proximal exterior of a $REFO$, and contact is not necessary; for the relation "near" contact between objects is explicitly prohibited.

A local coordinate system at the $REFO$, aligned to its deictic orientation, as shown in Figure 2, is defined, and the local coordinates of the LO w.r.t. the $REFO$ are computed through a transformation $\mathcal{T}_{REFO}$ (rotation and translation). From this the Euclidean distance $d_{\text{local}}(LO) := ||\mathcal{T}_{REFO}(LO)||$ is computed. We use simple trapezoidal membership functions μ_{topo} for the evaluation (others are possible, e.g. spline functions [6]), mapping object distances to the degree of applicability a_{topo}:

$$a_{topo} : (LO, REFO) \mapsto \mu_{topo}(d_{\text{local}}(LO))$$

with $topo \in \{at, near\}$. Figure 3 (top) shows a possible definition for membership functions for the relations "at" and "near".

4.3 Projective Relations

The relations "front of", "behind", "right", and "left" mainly depend on the orientation of the LO w.r.t. the $REFO$, and partition the space in qualitative acceptance areas (as suggested in [8]). But also the distance has to be taken into account: if the distance from the $REFO$ to the LO increases, the degree of applicability a_{proj} decreases. The evaluation function is defined as:

$$a_{proj} : (LO, REFO) \mapsto \mu_{\text{dist}}(d_{\text{local}}(LO)) \cdot \mu_{proj}(\alpha_{\text{local}}(LO))$$

with $proj \in \{front, behind, left, right\}$, mapping the orientation onto the linguistic variables, weighed by the distance. Figure 3 shows a possible definition of the functions μ_{proj} (bottom) and μ_{dist} (top). Although Gapp [7] dropped the distance factor μ_{dist} in an empirically validated revision of the model from [6], we retain it to account for the uncertainty in the visual object localisation, which in our approach weighs heavier than the concern for a cognitively valid model.

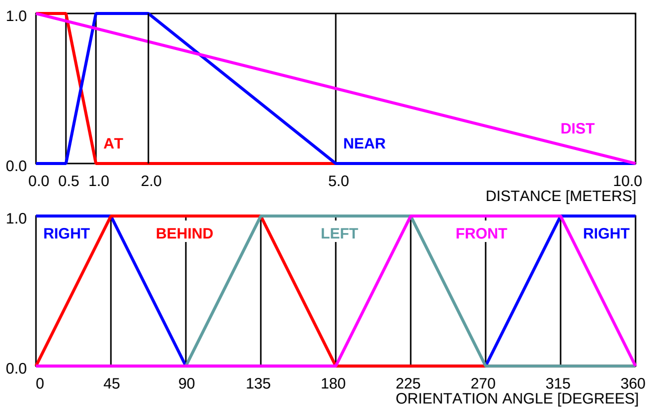

Figure 3. Used membership functions for the evaluation of the spatial relations: μ_{topo} (top), μ_{proj} (bottom), and μ_{dist} (top).

5 ANCHORING WITH SPATIAL RELATIONS

In order to integrate the spatial relations into the existing anchoring framework (see Figure 1), we proceed as follows: At every perceptual update cycle a decision is made for which anchors spatial relations have to be computed. In the current implementation, this is done for all anchored objects. For each anchor, as the located object, all defined spatial relations are computed with respect to all other selected anchors (as reference objects). Only those relations with a degree of applicability greater than a predefined threshold are considered, as in [1], and the others are discarded.

The computed spatial relations, tuples of the form $\langle LO, RO, relation, degree\rangle$, are stored within the anchor of the located object in an additional slot, as we do not consider this information to be part of the anchor's symbolic description. The *find* functionality was extended to include this information for the matching, to allow spatial relations in the symbolic description of a query.

Figure 4. The experimental test-bed, the PEIS-room: view of the kitchen, and robot inspecting the fridge with video camera and electronic nose.

5.1 Example Scenarios

The experimental test-bed for our system is a mobile robot platform that is part of an ambient intelligent environment, called the PEIS Ecology [13]. The robot "shares" a small furnished apartment (the PEIS-room, see Figure 4) with humans and other ambient intelligent devices, and is able to exchange information with these devices.

In the **first example** the robot surveys a static scene with three objects (two green garbage cans and a red ball, see Figure 5) and the anchoring module creates anchors for these objects as soon as they are recognised by the vision system. Then the computation of the spatial relations for these anchors is triggered, resulting in a relation-graph. The list of anchors (in LISP):

```
(ANCHOR ANCH-1 GAR-4
  (SYMBOLIC-DESCRIPTION
    ((SHAPE = GARBAGE) (COLOR = GREEN)))
  (PERCEPTUAL-DESCRIPTION ... )
  (SPATIAL-RELATIONS
    ((GAR-5 ((AT 1.0) (LEFT 0.94)))
     (BALL-2 ((AT 1.0) (BEHIND 0.94)
              (LEFT 0.62)))))
   ... )
(ANCHOR ANCH-2 BALL-2
  (SYMBOLIC-DESCRIPTION
    ((SHAPE = BALL) (COLOR = RED)))
  (PERCEPTUAL-DESCRIPTION ... )
  (SPATIAL-RELATIONS
    ((GAR-5 ((AT 1.0) (FRONT 0.96)
             (LEFT 0.43)))
     (GAR-4 ((AT 1.0) (FRONT 0.96)
             (RIGHT 0.2)))))
   ... )
(ANCHOR ANCH-3 GAR-5
  (SYMBOLIC-DESCRIPTION
    ((SHAPE = GARBAGE) (COLOR = GREEN)))
  (PERCEPTUAL-DESCRIPTION ... )
  (SPATIAL-RELATIONS
```

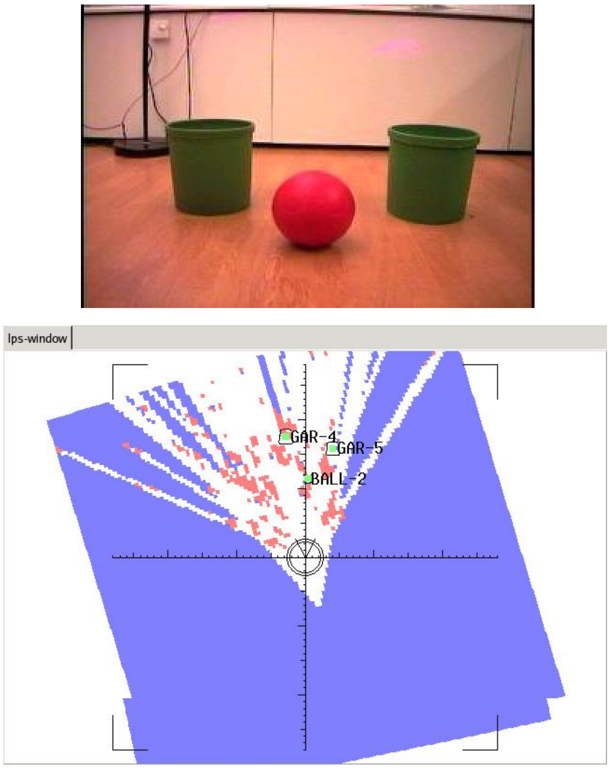

Figure 5. Example scenario: scene from the robot's viewpoint (top) and snapshot of the robot's perceptual space with the created anchors (bottom).

```
    ((GAR-4 ((AT 1.0) (RIGHT 0.94)))
     (BALL-2 ((AT 1.0) (BEHIND 0.96)
              (RIGHT 0.85)))))
   ... )
```

It is now possible to use spatial relations in the *find* functionality (implemented by `(FIND-ANCHOR (NAME SYMBOLIC-DESCRIPTION)))` to search for anchors, for example:

```
(FIND-ANCHOR 'MY-GARBAGE
  '((SHAPE = GARBAGE) (LEFT-TO = BALL-2)))
```

returns `((ANCHOR ANCH-1 MY-GARBAGE ... ))` as result.

In a **second experiment**, a human user is asked to resolve an ambiguity in a *find* request: in the scene from the previous example, the query is "Find the green garbage can". (This experiment is scripted and uses a simple pre-formulated scheme to guide the interaction with the user by text prompts.) As the find request returns more than one anchor (namely ANCH-1 and ANCH-3), the script determines an anchored object that is spatially related to these anchors as reference object and presents the user with a choice, enumerating the returned anchors and their spatial relation(s) to the reference object. Then the query is reformulated using additionally the selected relation(s). For example:

```
? (FIND-ANCHOR 'ANCH
    '((SHAPE = GARBAGE) (COLOR = GREEN)))
- FOUND 2 CANDIDATES: PLEASE CHOOSE
-  1.  GREEN GARBAGE LEFT BEHIND OF RED BALL
-  2.  GREEN GARBAGE RIGHT BEHIND OF RED BALL
? 1
- REFORMULATING:
-  (FIND-ANCHOR 'ANCH '((SHAPE = GARBAGE)
     (COLOR = GREEN) (LEFT-OF = BALL-2)
```

```
    (BEHIND-OF = BALL-2)))
- FOUND: ((ANCHOR ANCH-1 ANCH ...))
```

As outlined in the introduction, the robot should also be able to interact with intelligent devices in the environment, in addition to humans as illustrated in the previous example. Therefore in a possible **third scenario**, the robot could use an external video surveillance system to find an object of interest. In this case, a stationary video surveillance system consisting of several cameras, where each single camera incorporates a private instance of the anchoring module, keeps track of objects in the environment. If the robot is not able to recognise and locate a certain object of interest, but can describe the object in terms of a symbolic description, a request with this description can be sent to the surveillance system. Provided that one of the cameras is able to identify the object, the system, knowing the location of the robot, can qualitatively describe the object's location from the robot's point of view and send a reply.

5.2 Future Work

The current system still lacks a lot of desired functionality and has a number of major shortcomings, e.g.: For now we have not considered ego-motion of the robot; as we use an egocentric frame of reference, spatial relations have to be continuously updated while the robot is moving. The difficulty is to decide when to update the relations (e.g., change of view point), and which anchors are concerned. This demands a more convenient and detailed storage of the relations including view points and reference frames. Furthermore some inference (or reasoning) capability is desirable, to accomplish for example view point-taking (as outlined in [5]).

The linguistic HRI part is still unimplemented and will be one of the next steps. To exploit the anchoring framework in human-robot communication we intend to connect the anchoring module to a symbolic knowledge representation system and a (simple) speech dialogue system, as in [9]. Possible scenarios are semi-autonomous teleoperation of the robot by verbal instructions, like (incremental) navigation instructions, or object identification or localisation tasks involving spatial relations, similar to those described in [12].

6 CONCLUSION

In this work we have extended the anchoring framework to include how objects are spatially related to another in the environment. This is particularly useful for robotic systems working in real environments using real sensor information, as cases of ambiguity may arise where visually identical objects may be present. Furthermore, spatial-relation information facilitates human-robot communication where a human user may find it more intuitive to instruct a robot using spatial communication via language.

The implementation of the anchoring module and the spatial relations-part is still in an early stage and lacking many desired features, so that not all intended scenarios could be tested. This is left for future work.

ACKNOWLEDGEMENTS

This work has been supported through grants from Vetenskapsrådet (the Swedish Research Council) and KK-stiftelsen (the Knowledge Foundation, Sweden). For the first author: This research work was funded in part by CUGS (the National Graduate School in Computer Science, Sweden).

REFERENCES

[1] A. Abella and J. R. Kender, 'Qualitatively describing objects using spatial prepositions', in *Proc. of the 11th Nat. Conf. on Artificial Intelligence (AAAI-93)*, Washington DC, USA, (1993).

[2] M. Broxvall, S. Coradeschi, L. Karlsson, and A. Saffiotti, 'Recovery planning for ambiguous cases in perceptual anchoring', in *Proc. of the 20th Nat. Conf. on Artificial Intelligence (AAAI 2005)*, Pittsburgh, USA, (2005).

[3] S. Coradeschi and A. Saffiotti, 'Anchoring symbols to sensor data: Preliminary report', in *Proc. of the 17th Nat. Conf. on Artificial Intelligence (AAAI-00)*, Austin, USA, (2000).

[4] S. Coradeschi and A. Saffiotti, 'Symbiotic robotic systems: Humans, robots, and smart environments', *IEEE Intelligent Systems*, **21**(3), 82–84, (2006).

[5] T. Fong, I. Nourbakhsh, R. Ambrose, R. Simmons, A. Schultz, and J. Scholtz, 'The peer-to-peer human-robot interaction project', in *Proc. of the AIAA Space 2005 Conf.*, Long Beach, CA, USA, (2005).

[6] K.-P. Gapp, 'Basic meanings of spatial relations: Computation and evaluation in 3D space', in *Proc. of the 12th Nat. Conf. on Artificial Intelligence (AAAI-94)*, Seattle, USA, (1994).

[7] K.-P. Gapp, 'An empirically validated model for computing spatial relations', in *Proc. of the 19th Annual German Conf. on Artificial Intelligence (KI-95)*, Bielefeld, Germany, (1995).

[8] D. Hernández, *Qualitative Representation of Spatial Knowledge*, volume 804 of *LNCS*, Springer-Verlag, 1994.

[9] J. Hois, M. Wünstel, J.A. Bateman, and T. Röfer, 'Dialog-based 3D-image recognition using a domain ontology', in *Proc. of the Int. Conf. Spatial Cognition 2006*, Bremen, Germany, (2006).

[10] A. Loutfi, S. Coradeschi, and A. Saffiotti, 'Maintaining coherent perceptual information using anchoring', in *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI-05)*, Edinburgh, Scotland, (2005).

[11] R. Luke, S. Blisard, J. Keller, and M. Skubic, 'Linguistic spatial relations of three dimensional scenes using SIFT keypoints', in *Proc. of the 14th IEEE Int. Workshop on Robot and Human Interactive Communication (RO-MAN 2005)*, Nashville, TN, USA, (2005).

[12] R. Moratz and T. Tenbrink, 'Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations', *Spatial Cognition and Computation*, **6**, 63–106, (2006).

[13] A. Safiotti and M. Broxvall, 'PEIS Ecologies: Ambient intelligence meets autonomous robotics', in *Proc. of the Joint Conf. on Smart Objects and Ambient Intelligence (sOc-EUSAI '05)*, Grenoble, France, (2005).

[14] M. Skubic, P. Matsakis, G. Chronis, and J. Keller, 'Generating multi-level linguistic spatial descriptions from range sensor readings using the histogram of forces', *Autonomous Robots*, **14**, 51–69, (2003).