# Functional Mapping: Spatial Inferencing to Aid Human-Robot Rescue Efforts in Unstructured Disaster Environments

**Shanker Keshavdas** and **Hendrik Zender** and **Geert-Jan M. Kruijff**
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken (Germany)


**Ming Liu** and **Francis Colas**
Swiss Federal Institute of Technology (ETH)
Zürich (Switzerland)

## Abstract

In this paper we examine the case of a mobile robot that is part of a human-robot urban search and rescue (USAR) team. During USAR scenarios, we would like the robot to have a geometrical-functional understanding of space, using which it can infer where to perform planned tasks in a manner that mimics human behavior. We assess the situation awareness of rescue workers during a simulated USAR scenario and use this as an empirical basis to build our robot's spatial model. Based upon this spatial model, we present "functional mapping" as an approach to identify regions in the USAR environment where planned tasks are likely to be optimally achievable. The system is deployed and evaluated in a simulated rescue scenario.

## Introduction

Rescue workers exploring a disaster environment possess task-specific spatio-cognitive knowledge of such scenes. This task-specific knowledge – for example the likely places to find road accident victims – is then compared to the real disaster scenario. Similarly, we would like our robot to perform *top-down inferencing* – e.g., best places to look for victims – and project this spatial information onto *bottom-up* knowledge gathered – e.g., objects detected in an incomplete map – to plan the best path to complete a particular task. This implies that the robot needs detailed semantic knowledge of (a) the objects to expect in a disaster scenario, (b) corresponding functions performable, and (c) its own perception capabilities. This exploration may lead to further oppurtunities of *knowledge gathering* tasks, requiring still further spatial inferencing computed from the a priori knowledge of the robot. We call this continuous spatial inferencing and planning process *functional mapping*.

In a human-robot interaction context, intelligent behavior implies that the robots knowledge and the actions performed by it should be *human-understandable*, *human-compatible* (having a similar spatial model as possessed by rescue workers), and *team-oriented* (adhering to mission objectives). With individual agents working together in a team, it is important to be in the right place at the right time as demonstrated for robots and remote operators in [Murphy et al., 2008]. These implications are especially important in our case, when the human agents are not familiar with the logic employed by the robot.

Below we first discuss the current approaches in the field. Then we briefly describe the "tunnel accident" use case and observe how rescue workers explore such an accident. The field data thus gathered, forms the empirical basis for our approach. We then present our approach, which is based on inferencing from a similar cognitive spatial model as that of rescue workers, and spatial projection based on it. We conclude with details of future plans with respect to functional mapping.

## Background and Related Work

In the field of semantic mapping for outdoor, unstructured environments as addressed in this paper, the state of the art is at an early stage. Most approaches either use a complex spatial intelligence in structured environments or inversely a low-level spatial intelligence in unstructured environments. Traversibility analysis in unstructured environments either using Hidden Markov models and the Viterbi algorithm [Wolf and Sukhatme, 2005] or using a multiresolution grid [Montemerlo and Thrun, 2006], can be considered to be a computationally interesting form of semantic mapping, however they do not yield more complex spatio-cognitive structures.

There has been a great deal of research in semantic classification and labelling in structured environments such as sparse indoor rooms or corridors. However even in these approaches, the research does not use detailed human-compatible a priori knowledge as present in our approach. The a priori knowledge in those approaches are much more low-level such as classifying rooms based on laser scan char-

acteristics [Goerke and Braun, 2009], abstracting a 3D laser scan into ceiling, floor and wall planes based upon basic ontological relationships [Nüchter and Hertzberg, 2008] or segmenting large spaces into rooms based upon objects typically observed in them [Vasudevan et al., 2007]. In contrast, our approach draws inferences from human-readable ontologies, which are modelled on the task-specific spatial knowledge of human beings. An example of such inferencing is [Zender, Jensfelt, and Kruijff, 2007]. The authors demonstrate a situation-aware indoor robot that can recognize doors and the space required for interacting with them (i.e. opening, closing and passing through) and appropriately act on this information to make space for the user. [Tenorth and Beetz, 2008] also presents a similar approach where a robot that is trained in a simulated kitchen environment, can then use knowledge on common objects and actions performable on them to assist a human being in a real world scenario. Common objects and actions are stored in an ontology built upon the Cyc ontology. In our work, we build on the approach of conceptual mapping of [Zender et al., 2008] and [Sjöö et al., 2010] along with function-specific spatial inferencing, and extend it to outdoor semi-structured environments.

## Empirical Basis

Our scenario is one that involves a human-robot team jointly exploring a traffic incident in a tunnel. Vision is impaired by smoke filling the tunnel. We have performed high-fidelity simulations of the disaster scenario, shown in Fig. 1, with robots and firefighters at the training site of the Italian National Fire Watch Corps (SFO at Montelibretti, Italy) and at the one of the Fire Department of Dortmund, Germany (FDDO). In the setup at SFO, we wanted to observe the vi-
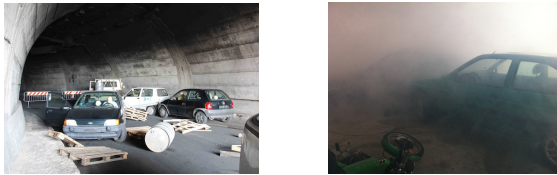


Figure 1: The tunnel accident scenario at SFO, Italy and FDDO, Germany

sual points of attention that firefighters maintained during a rescue operation and match these with their spoken communication. For this reason, they were equipped with eye-gaze machines that track their visual attention [Khambhaita et al., 2011], and their communication during several mock rescue operations was also recorded. A sample audio recording of a firefighter read as follows:

(1) A car, a civil car, with some people inside.

(2) A family. People. A woman drives. A person in *the front seat*. A child. Another child in *the rear seat*. Another child, a baby.

One thing that can be observed here is the felicitous use of hearer-new definite descriptions (marked in italics) [Prince, 1992; Poesio and Vieira, 1998]. Definite descriptions are supposed to refer to mutually known entities in the domain of discourse. The information of the structure of the car (eg: rear seat) is from the mental representation of the firefighter, where the representation of a car has been evoked by the indefinite description "a car" (the so-called *trigger entity*). And through his *prior knowledge* about cars he can be assumed to know that cars in general have (front and rear) seats. Such uses of a definite description to refer to an implicitly evoked entity that can be inferred based on background knowledge are called "inferrables" [Prince, 1992] or "bridging anaphora" [Clark, 1975]. The group of bridging anaphora that come into play in our recordings are the so-called "indirect references by association", which Clark explains with their predictability as being an associated part of the trigger entity. From the transcriptions, we observe that the firefighter's task is tightly correlated with the hierarchical composition of the spatial structure: the tunnel contains cars, which in turn contain victims; a truck, which typically contains goods; and barrels which usually contain (potentially hazardous) substances. It is generally assumed that humans adopt such a (partially) hierarchical representation of spatial organization [Stevens and Coupe, 1978; Hirtle and Jonides, 1985; McNamara, 1986]. This demonstrates the kind of inferences on background knowledge that the robot must perform, not only to autonomously determine a plan for locating victims but to produce and comprehend natural language scene descriptions.

At another simulation scenario at FDDO, firefighters were given tele-operational control of the robot. The scenario was of an unknown smoke-filled environment and where they had to record the positions of vehicles, victims and hazardous material that they observed. Our interest in the experiment was to notice the vantage points the firefighters assumed when observing the inside of a car to look for victims, when looking at a motorbike, and explosive barrels. Once the trials were completed, we marked a boundary of 1 meter around the regions of interest (the car windows, the motorbike, and the barrel). We assumed that this was a sufficient visual range for affording the function of observing these regions of interest. We call the areas marked off by the boundaries as 'functional areas' – since these areas enable the function of observing these regions. In Fig. 2, we show the runs of three of the firefighters who participated in our experiment. Table 1 shows the percentage of 'observation time', or time spent inspecting the regions of interest. We further mention the percentage of the observation time spent in functional areas of objects. From the data, we notice that Participant 1 and 3 spent over half, and Participant 2 spent nearly all observation time in the functional areas, divided into time spent observing vehicles and threats. This confirms our belief that rescue workers do employ strategic vantage points to observe regions of interest. We would like our robot to draw similar *human-compatible* spatial inferences to search for victims.

## System Architecture

The NIFTi project is a multi-disciplinary effort, with a system architecture as described in Fig. 3. The functional mapping layer covers functionality that ranges from the conceptual and ontological understanding of the environ-
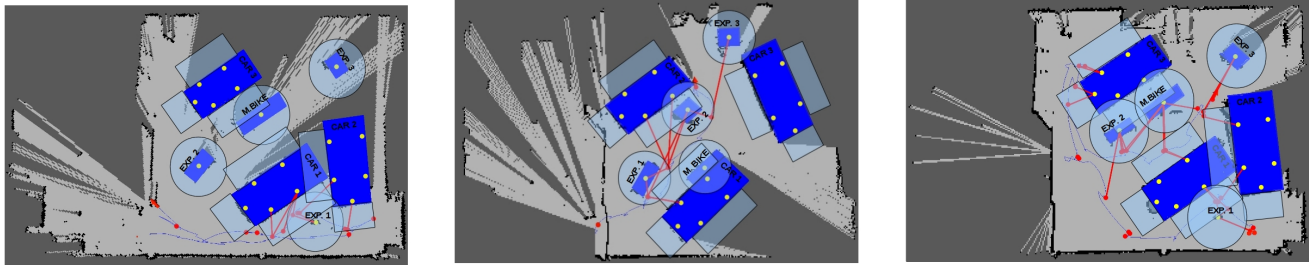
Figure 2: Maps acquired by tele-operation in FDDO, Germany, showing points from where observations/transcriptions were made(red), points of attention which they were observing(yellow), functional areas(light blue) and the path of the robot (blue trajectory).

| Participant | Percentage of observation time | Percentage of observation time in functional areas | Percentage of observation time in functional areas of different objects of interest | |
|---|---|---|---|---|
| | | | Vehicles | Expl. Barrels |
| 1 | 38.17 | 66.7 | 86.67 | 13.33 |
| 2 | 53 | 97.6 | 0 | 100 |
| 3 | 48 | 65.3 | 41.96 | 58.04 |

Table 1: An analysis of the time spent for the tele-operated runs shown in Fig. 2

ment down to the low-level metrical maps created from the laser mounted on the robot, and stored together in multi-hierarchical maps. This layer communicates the understanding it derives from its concepts about the environment including regions, objects and actions to be performed to other higher level processes such as user interfaces, human-robot communication through natural dialogue and planning. The system uses the ROS architecture for lower-level control, e.g., motor and cameras, and the CAST middleware [Hawes and Wyatt, 2010] for higher-level processing. Several components simultaneously communicate messages on both ROS and CAST servers.

**Mapping and Navigation**

To perform Simultaneous Localization and Mapping (SLAM), we use the ROS GMapping toolkit[1], a state-of-the-art implementation of a SLAM algorithm. Using Rao-Blackwellized particle filtering, it successively estimates the pose and the map based on sensor readings [Grisetti, Stachniss, and Burgard, 2007] and generates a 2D occupancy grid map. The robotic platform we used was equipped with a 2D laser range finder with an angle of $180°$.

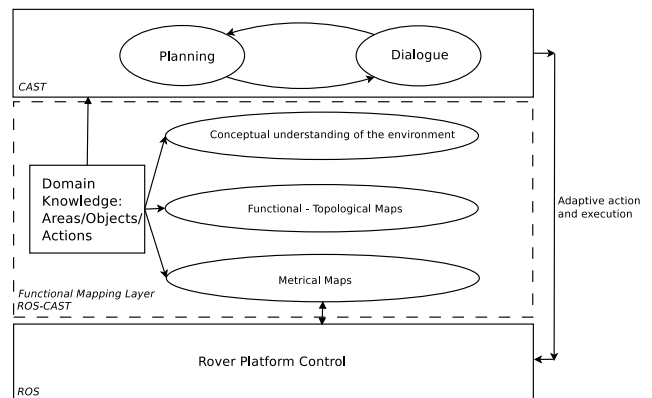Navigation and motion control of the robot are handled



Figure 3: Architecture schema showing integration of various components of the NIFTi Project

using the ROS navigation stack[2] toolkit. It enables a mobile robot to fulfill navigation tasks using the occupancy map. Users send global goal coordinates and the navigation stack will plan a trajectory, taking into account reachability and traversability to the target.

**Vision**

Cars are detected and localized by an online learnable object detector, which is based on the efficient combination of [Lepetit, Lagger, and Fua, 2005; Kalal, Matas, and Mikolajczyk, 2010; Hurych, Zimmermann, and Svoboda, 2011]. We use several rapidly updated detectors instead of a single tracker. These detectors, each using the same set of features but updated with different parameters, yield similar boosting ability as the tracker in [Kalal, Matas, and Mikolajczyk, 2010], while preserving a real-time performance. Car detections from individual camera frames are collected over time. Once a sufficient number of detections from sufficiently distant poses are available, the 3D position of a detected car is estimated via localization method described in [Hartley and Zisserman, 2000].

## Approach

We have observed that rescue workers perform their tasks based upon their possession of a rich, well-defined spatial structure of entities that they expect to observe in the environment. In a similar fashion, functional mapping is equipped with a hand-written OWL/RDF-based ontology [Baader, Horrocks, and Sattler, 2005] of the two domains that we require 3D spatial information about consumer cars and optical properties of cameras used for robotics. Both domains are filled in from manufacturer datasheet information.

Also, the German Automotive Industry Union (VDA) provides datasheets to rescue workers [3], with exploded-view schematics that contain vital measurements, e.g., the positions of embedded gas cylinders that may explode during a rescue attempt. These datasheets do not contain the dimensions we require and are not presently machine-readable, but we hope that such information will be available in the future.

Fig. 5 shows an excerpt of our ontology, relating to a specific car and camera model. In order to preserve standardization, the classes in our domain are based upon the hierarchy of the Wordnet lexical database [Fellbaum, 1998]. Our system is endowed with the forward chaining engine HFC [Krieger, 2011], which allows us to perform inference over OWL/RDF-based ontologies. We have equipped the HFC reasoner with a standard OWL-DL rule set and a number of custom rules for drawing *default conclusions* [Reiter, 1980; Antoniou, 1997].

Fig. 4 summarizes the working of the system. Perception of an entity in the environment triggers the Ontology-Inferencing module. Based upon the current task and under normal mission status (absence of fire emergencies etc.), this module sends a query to the ontology database. The queries are resolved in the ontology. From the responses generated, the spatial inferencer can determine the optimal geometrical configurations to perform the given task. This procedure is explained in the next paragraph. Achieving these tasks can trigger re-detections in the perception unit, leading to more accurate re-estimations of the functional mapping configurations. This cycle is applied repeatedly for different tasks.
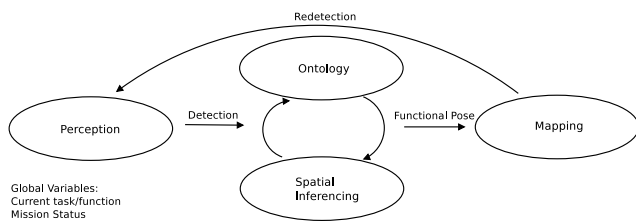


Figure 4: Functional Mapping Workflow

Spatial Inferencing for the functional mapping workflow victim search can be seen in Fig. 6. The trigger is the detection of a car of known 'CarModel' class. Following the approach presented in [Sjöö et al., 2010], the reasoner then derives the default knowledge about the car's dimensions and the location and dimensions of its windows. The next step is to search for the robot positions that will afford looking into these windows. Search spaces of sufficient size for this search are generated in front of the windows. The spatial inferencer then queries the ontology for the visual capabilities of the robot. The ontology returns the camera parameters including horizontal $H$ and vertical $V$ angles of the view cone of the camera and the reliable range $R$. With these parameters, the search space is linearly sampled for the position of the robot and the position of the camera with respect to the robot. For each position of the camera, the view cone is linearly sampled ray by ray. Each ray is projected into the plane of the corresponding window and if the point of intersection is included in the polygon corresponding to the car's window[4], the point of intersection is stored. These points are accumulated as a patch $A$ and if the patch area is greater than the average size of a face used in the car detection algorithm $p$, the corresponding robot position is stored. These robot positions are then accumulated and abstracted into a polygon, composed of points closest to the edges of the search space. All these points are then converted into robot poses, with the robot direction facing the window and sent to the planning and navigation components as planning coordinates.

The above description of a functional mapping cycle for victim search, has been implemented and in Fig. 6, is a screenshot of one of our runs at FDDO. Another function that can be integrated into the functional mapping cycle is car detection. As we have described earlier, the car detector detects the car models with a certain probability. Also, based upon the features used in the car detection, certain positions of the robot and subsequently the camera, would be more benificial for car detection. In our case, the most reliable features for car detection are observed from the back of the car. Thus when a car is detected with a low confidence of detection, functional mapping can generate poses facing the rear of the car. In keeping with our cyclic concept of functional mapping for knowledge gathering triggering re-detections, the position of the car can be redetected until the detector reaches a sufficient level of confidence. Once this level is reached, we can proceed to our function of victim search.

In this discussion, it is necessary to point out that datasheets linking registered license plates of an area to car models are generally available to government authorities. If the visual system of the robot is able to read the license plates through algorithms such as [Anagnostopoulos et al., 2008], it should be possible to find out the model of the car, and then infer its dimensions through ontologies such as ours.

Since our approach makes use of default knowledge, it serves as a top-down process that can raise expectations about functional areas, even when these areas have not yet been explored. A discussion of how a robot system can make use of such default knowledge in order to automatically generate meaningful plans under partial observability and incomplete knowledge about its environment can be found in
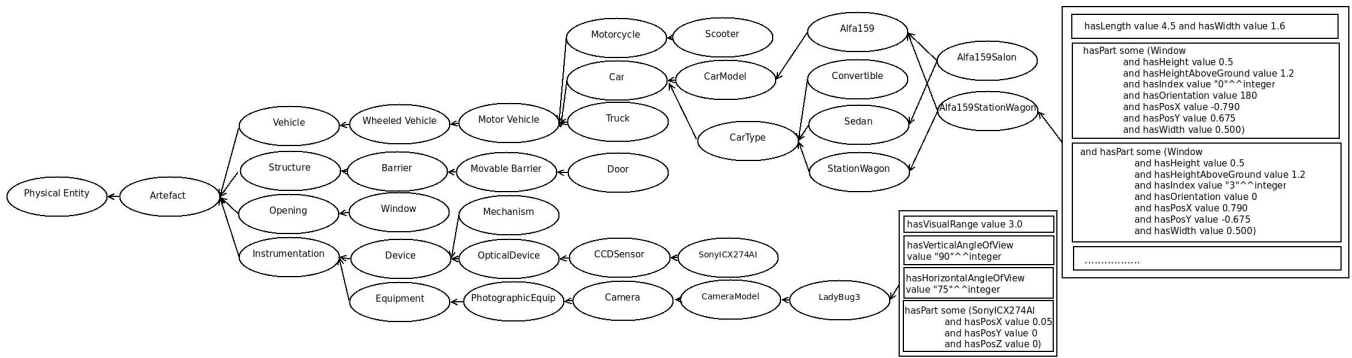
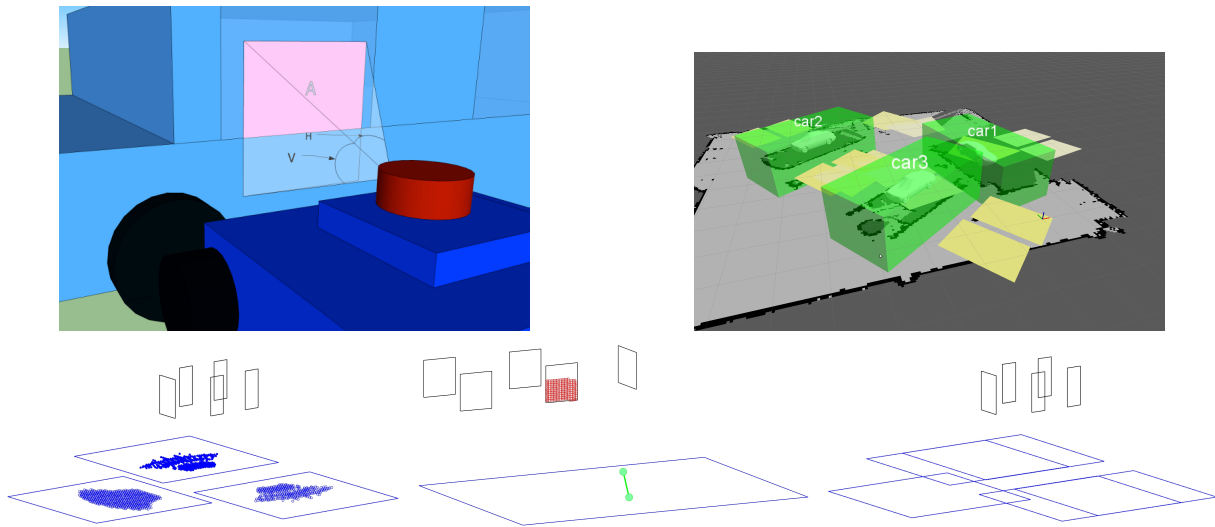Figure 5: An excerpt of the car accident domain ontology. Default properties for two classes are shown.



Figure 6: (Clockwise from top): Sketch showing window visibility patch *A*, with a visibility cone of vertical angle *V* and horizontal angle *H*; Functional areas observed during pilot tests at FDDO; window positions(black) queried from the ontology and search spaces(blue) ; window visibility (red circles) case of a single robot pose, robot and camera(green); accumulated robot positions with visibility patch greater than face size patch (blue circles)

[Hawes et al., 2009].

## Conclusion

We have presented an approach to outdoor conceptual-functional mapping for intelligent robots operating in the USAR domain. The approach makes use of state-of-the-art methods in robot mapping, conceptual mapping, computer vision and ontological reasoning.

We have shown empirically the usage of our concept of functional areas by rescue personel and based our ontologies on similar information and principles to the knowledge possessed by the rescue workers in disaster scenarios. Our system successfully interprets objects detected in a search and rescue scenario and derives ontological inferences based on these detections. It infers functions that these objects could afford, and derives areas based on our reasoner where these functions could be afforded. It then projects these areas back onto the map for display to the user and adds the information

of these areas onto the working memory of the CAST system from where other components may then use this information. Our approach is informed by high-fidelity field experiments with expert rescue workers at SFO Montelibretti. It was also successfully deployed and tested at FDDO Dortmund, where in a simulated rescue operation, firefighters were able to control the robot in an effort to find victims in the scenario.

In the future, we will focus on testing our algorithm in relation to executing plans in a simulated accident environment, possibly using USARSim. Our robot has adaptable flipper tracks for traversing uneven terrain. We will look into including the variable morphology of our robot into the ontology, for performing more complex navigation in order to reach a goal. We will also investigate an approach for acquiring a high-coverage car ontology with car dimensions, number of doors, windows, etc. from available databases.

# References

Anagnostopoulos, C. N. E.; Anagnostopoulos, I. E.; Psoroulas, I. D.; Loumos, V.; and Kayafas, E. 2008. License Plate Recognition From Still Images and Video Sequences: A Survey. *Intelligent Transportation Systems, IEEE Transactions on* 9(3):377–391.

Antoniou, G. 1997. *Nonmonotonic Reasoning*. Cambridge, MA, USA: The MIT Press.

Baader, F.; Horrocks, I.; and Sattler, U. 2005. Description logics as ontology languages for the semantic web. In Hutter, D., and Stephan, W., eds., *Mechanizing Mathematical Reasoning*, volume 2605 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 228–248.

Clark, H. H. 1975. Bridging. In Schank, R., and Nash-Webber, B., eds., *Theoretical Issues in Natural Language Processing*. New York, NY, USA: Association for Computing Machinery.

Fellbaum, C., ed. 1998. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, illustrated edition edition.

Goerke, N., and Braun, S. 2009. Building semantic annotated maps by mobile robots. In *Proc. of TAROS 2009: Towards Autonomous Robotic Systems*, 149–156. Univ. of Ulster, UK.

Grisetti, G.; Stachniss, C.; and Burgard, W. 2007. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Transactions on Robotics* 23:34–46.

Hartley, R. I., and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hawes, N., and Wyatt, J. 2010. Engineering intelligent information-processing systems with CAST. *Advanced Engineering Infomatics* 24:27 – 39.

Hawes, N.; Zender, H.; Sjöö, K.; Brenner, M.; Kruijff, G.-J. M.; and Jensfelt, P. 2009. Planning and acting with an integrated sense of space. In *Proc. of the 1st International Workshop on Hybrid Control of Autonomous Systems – Integrating Learning, Deliberation and Reactive Control*, 25–32.

Hirtle, S. C., and Jonides, J. 1985. Evidence for hierarchies in cognitive maps. *Memory and Cognition* 13:208–217.

Hurych, D.; Zimmermann, K.; and Svoboda, T. 2011. Fast learnable object tracking and detection in high-resolution omnidirectional images. In *Proc. of International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 2011*.

Kalal, Z.; Matas, J.; and Mikolajczyk, K. 2010. P-n learning: Bootstrapping binary classifiers by structural constraints. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition 2010*, 49–56. IEEE.

Khambhaita, H.; Kruijff, G.; Mancas, M.; Gianni, M.; Papadakis, P.; Pirri, F.; and Pizzoli, M. 2011. Help me to help you: How to learn intentions, actions and plans. In *Proc. of the AAAI Spring Symposium 2011*.

Krieger, H. 2011. A temporal extension of the Hayes and ter Horst entailment rules for RDFS and OWL. In *Proc. of AAAI 2011 Spring Symposium "Logical Formalizations of Commonsense Reasoning"*.

Lepetit, V.; Lagger, P.; and Fua, P. 2005. Randomized trees for real-time keypoint recognition. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005*, 775–781.

McNamara, T. P. 1986. Mental representations of spatial relations. *COGNITIVE. PSYCHOL.* 18:87–121.

Montemerlo, M., and Thrun, S. 2006. Large-scale robotic 3-d mapping of urban structures. In Ang, M., and Khatib, O., eds., *Experimental Robotics IX*, volume 21 of *Springer Tracts in Advanced Robotics*. Springer Berlin / Heidelberg. 141–150.

Murphy, R.; Tadokoro, S.; Nardi, D.; Jacoff, A.; Fiorini, P.; Choset, H.; and Erkmen, A. 2008. Search and rescue robotics. In Siciliano, B., and Khatib, O., eds., *Springer Handbook of Robotics*. Springer Verlag. Part F, 1151–1173.

Nüchter, A., and Hertzberg, J. 2008. Towards semantic maps for mobile robots. *ROBOT. AUTON. SYST.* 56(11):915–926.

Poesio, M., and Vieira, R. 1998. A corpus-based investigation of definite description use. *COMPUT. LINGUIST.* 24(2):183–216.

Prince, E. 1992. The ZPG letter: subjects, definiteness, and information status. In Thompson, S., and Mann, W., eds., *Discourse Description: diverse analyses of a fund raising text*. John Benjamins. 295–325.

Reiter, R. 1980. A logic for default reasoning. *ARTIF. INTELL.* 13(1-2):81–132.

Sjöö, K.; Zender, H.; Jensfelt, P.; Kruijff, G.-J. M.; Pronobis, A.; Hawes, N.; and Brenner, M. 2010. The Explorer system. In Christensen, H. I.; Kruijff, G.-J. M.; and Wyatt, J. L., eds., *Cognitive Systems*, volume 8 of *Cognitive Systems Monographs*. Berlin/Heidelberg, Germany: Springer Verlag. chapter 10, 395–421.

Stevens, A., and Coupe, P. 1978. Distortions in judged spatial relations. *COGNITIVE. PSYCHOL.* 10:422–437.

Tenorth, M., and Beetz, M. 2008. Towards practical and grounded knowledge representation systems for autonomous household robots. In *Proc. of the 1st International Workshop on Cognition for Technical Systems*.

Vasudevan, S.; Gachter, S.; Nguyen, V.; and Siegwart, R. 2007. Cognitive Maps For Mobile Robots: An Object Based Approach. *ROBOT. AUTON. SYST.* 55(5):359–371.

Wolf, D. F., and Sukhatme, G. S. 2005. Autonomous terrain mapping and classification using hidden markov models. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA*, 2026–2031.

Zender, H.; Mozos, O. M.; Jensfelt, P.; Kruijff, G.-J. M.; and Burgard, W. 2008. Conceptual spatial representations for indoor mobile robots. *ROBOT. AUTON. SYST.* 56(6):493–502.

Zender, H.; Jensfelt, P.; and Kruijff, G.-J. M. 2007. Human- and situation-aware people following. In *Proc. of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007)*, 1131–1136.