Taylor & Francis
Taylor & Francis Group

Advanced Robotics

ISSN: 0169-1864 (Print) 1568-5535 (Online) Journal homepage: https://www.tandfonline.com/loi/tadr20

A review of spatial reasoning and interaction for real-world robotics

C. Landsiedel, V. Rieser, M. Walter & D. Wollherr

To cite this article: C. Landsiedel, V. Rieser, M. Walter & D. Wollherr (2017) A review of spatial reasoning and interaction for real-world robotics, Advanced Robotics, 31:5, 222-242, DOI: 10.1080/01691864.2016.1277554

To link to this article: https://doi.org/10.1080/01691864.2016.1277554

Published online: 19 Jan 2017.

ADVANCED ROBOTICS, 2017
VOL. 31, NO. 5, 222–242
http://dx.doi.org/10.1080/01691864.2016.1277554

SURVEY PAPER

A review of spatial reasoning and interaction for real-world robotics

C. Landsiedel[a], V. Rieser[b], M. Walter[c] and D. Wollherr[a]

[a]Chair of Automatic Control Engineering, Technical University of Munich, Munich, Germany; [b]School of Mathematical and Computer Sciences (MACS), Heriot-Watt-University, Edinburgh, UK; [c]Toyota Technological Institute at Chicago, Chicago, IL, USA

ABSTRACT

Truly universal helper robots capable of coping with unknown, unstructured environments must be capable of spatial reasoning, i.e. establishing geometric relations between objects and locations, expressing those in terms understandable by humans. It is therefore desirable that spatial and semantic environment representations are tightly interlinked. 3D robotic mapping and the generation of consistent metric representations of space are highly useful for navigation and exploration, but they do not capture symbol-level information about the environment. This is, however, essential for reasoning, and enables interaction via natural language, which is arguably the most common and natural communication channel used and understood by humans. This article presents a review of research in three major fields relevant for this discussion of spatial reasoning and interaction. Firstly, dialogue systems are an integral part of modern approaches to situated human–robot interaction. Secondly, interactive robots must be equipped with environment representations and reasoning methods that are suitable for both navigation and task fulfillment, as well as for interaction with human partners. Thirdly, at the interface between these domains are systems that ground language in systemic environment representation and which allow the integration of information from natural language descriptions into robotic maps. For each of these areas, important approaches are outlined and relations between the fields are highlighted, and challenging applications as well as open problems are discussed.

ARTICLE HISTORY
Received 29 August 2016
Revised 14 December 2016
Accepted 17 December 2016

KEYWORDS
Situated human–robot interaction; natural language grounding; spatial reasoning; environment modelling; semantic mapping

1. Introduction

Robots, as we see them being developed now and envision them for the future are unique among the recent technological innovations in that their features include acting and navigating in spaces shared with humans, perceiving the joint surroundings, and communicating about them. For many imaginable use cases of robotics, the specification and communication of spatial arrangements and layouts are essential. Scenarios such as search and rescue, urban or indoor navigation, and collaborative manipulation all require a understanding of space that is shared between robots and humans. Under the paradigm of natural and accessible user interfaces for these types of robot, spoken natural language is an essential modality for communication between robots and their human users. For example, in a collaborative industrial construction scenario, this might allow a user to instruct their assistive robot with a command like ‘Take the screwdriver from the toolbox in the back of the storage room and use it to tighten the upper bolt of the left car door that I just positioned’. The robot would then parse this information, identify the necessary concepts and objects at the appropriate locations in its commonsense and environment knowledge, find and clarify possibly missing information necessary to carry out the task, and then plan and execute the actions asked for.

The combination of abilities necessary for such scenarios poses a set of unique challenges. The robot must be able to reason about space both in high-level human terms, as in understanding spatial language, and on a lower level that is related to its own sensors and manipulators. For most robotic tasks, in particular those that involve navigation, a metric map of the environment is necessary. This format, however, is not amenable for communicating concepts in a way that is intuitive to humans. Abstractions of quantities perceived through sensors to symbolic, qualitative terms and the inclusion of semantic information are necessary for the robotic knowledge to allow a dialogue that is close to the human way of reasoning about space. This type of interaction, where human and robot share an environment which is the topic of their discourse is known as *situated* interaction. It requires the robot to *ground* spatial language in its internal environment representation, and vice versa reference objects contained therein in interaction. Furthermore, interaction can deliver environment information

CONTACT C. Landsiedel ✉ christian.landsiedel@tum.de

that is not perceivable with the robot's sensor repertoire or outside the sensory horizon.

This article reviews fundamental and recent work on two main aspects of spatial reasoning and interaction in robotics. The first part concerns natural language interaction between robots and their human users about spatial relationships. Dialogue systems for such grounded, situated human–robot interaction (HRI) are discussed in Section 2.1. Section 2.2 reviews the grounding of spatial language in environment representations. Spatial information from both sensors and interaction must be grounded and referenced in the robot's internal knowledge representation. Environment representations, in particular ones that allow to store semantic information, and symbolic spatial reasoning systems in robotics make up the second main part in Section 3. Since environment representations have been traditionally based on studies of how humans handle spatial information cognitively, a short overview over the most important findings in this field is given in Section 3.1. Qualitative approaches to spatial reasoning and mapping are discussed in Section 3.2. Section 3.3 gives an overview over mapping approaches used in robotics, especially ones that incorporate qualitative components and allow the representation of semantic information. Specific attention is devoted to the derivation of semantic map information from natural language descriptions in Section 3.4. The article concludes with a discussion of challenges that remain for robots reasoning and interacting about space in the real world, and ideas about how to address them.

2. Natural language human–robot interaction about space

One of the most direct and natural ways to communicate with robots is natural language. In order for robots to understand what we say and to respond with a coherent, well-formed utterance, they need to be able to process and generate natural language, as well as to reason about the current context. This section first introduces Spoken Dialogue Systems (SDS), which are traditionally used to model these skills. Of particular importance for interaction about spatial relationships is the grounding of the human-robot communication in the environment representation of the robot, methods for which are reviewed in Section 2.2.

2.1. Situated dialogue for human-robot interaction

Broadly speaking, a dialogue system has three modules, one each for input, output and control, as shown in Figure 1 after [1]. The input module commonly comprises Automatic Speech Recognition (ASR) and Spoken Language Understanding (SLU). The control module corresponds to the Dialogue Manager (DM), which executes a dialogue strategy. The output module consists of a Natural Language Generation (NLG) system and a Text-To-Speech (TTS) engine. Usually, these modules are placed in a pipeline model. The ASR converts the user's speech input (1) into text (2), see Figure 1. SLU parses the text into a string of meaningful concepts, intentions, or Speech Acts (SA) (3). The DM maintains an internal state and decides what SA action to take next (4). This is what we call a dialogue strategy. For most applications the DM is also connected to a back-end database. In the output module, NLG renders the communicative acts (4) as text (5), and the TTS engine converts text to audio (6) for the user. Interested readers are referred to introductory texts such as [2,3].

The distinctive characteristic of *situated dialogue* is that participants are placed in a shared spatio-temporal context. When communicating, both participants can refer to objects in the environment, while each participant has a individual perceptual perspective. As such, participants need to make sure they understand each others' utterances and can uniquely resolve references to the world around them. This is also referred to as *grounding*. Linguists have developed advanced theories of how this grounding might work in human–human conversation, e.g. [4], some of which are implemented in SDS, e.g. [5]. When moving to HRI, several new challenges arise, some of which are covered in this special issue as detailed in Section 4.

2.2. Interpreting spatial natural language

Natural language provides an efficient, flexible means by which users can convey information to their robot partners. Natural language utterances may come in the form of commands that instruct robots to carry out manipulation tasks [6,7] or to navigate [8–12] within their environment. The problem of interpreting free-form instructions can be formulated as what Harnad [13] refers to as the symbol grounding problem, in which the objective is to map linguistic elements within the utterance to their corresponding referents in the physical world. Early research in natural language symbol grounding relies upon manually engineered rules that exploit the compositional structure of language to associate words in the utterance to sets of predefined environment features (e.g. a metric map in the form of an occupancy grid) and actions [14–17]. The use of static language-to-symbol mappings limits understanding to a small, fixed set of phrases and consequently does not scale to the diversity of natural language. Later work employs statistical methods to model the symbol grounding problem using a flat

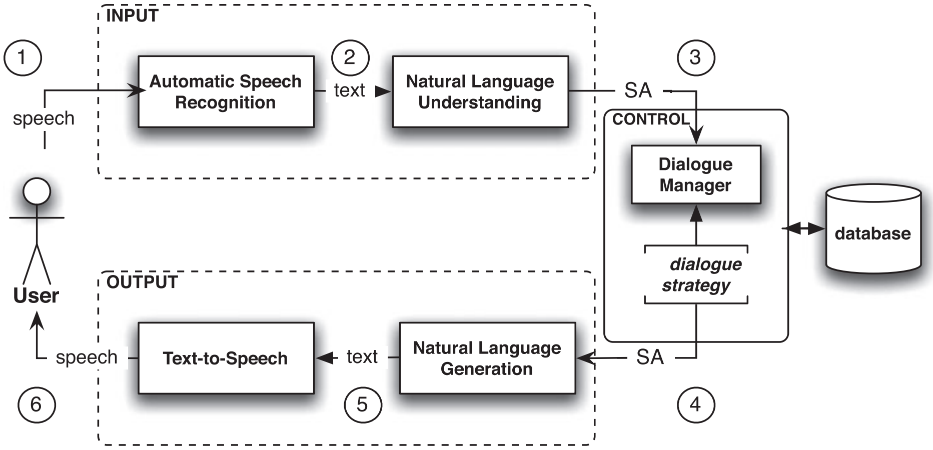


Figure 1. Architecture for SDS, also see [1].

representation of the free-form utterance. These techniques [8,9,18–20] learn to convert free-form utterances into their referent symbols by employing language in a perceptual context [21]. The symbols may take the form of features in a hybrid map of the environment (Section 3.3.3) that express spatial, semantic, and topological properties of different objects and places. These symbol grounding methods model natural language grounding in terms of a fixed set of manually defined linguistic, spatial, and/or semantic features, but are unable to resolve more complex expressions that require modelling the hierarchical structure of language.

One approach to the problem of grounded language acquisition is to treat language understanding as a problem of learning a parser that converts natural language into a formal language equivalent. Importantly, many of these methods do not require a prior representation of the environment and instead rely upon a rules- or constraint-based planner to satisfy the parsed formal language according to sensor data. Matuszek et al. [9] parse free-form language using a general purpose parser trained in a supervised fashion on natural language utterances paired with their formal language groundings. Similarly, Chen and Mooney [22] parse natural language navigation instructions into a formal specification of mobility actions that a downstream robot control process can process and execute. The parser is trained in a weakly supervised fashion from pairs of natural language instructions and their corresponding action sequence, along with a symbolic representation of the robot's environment. Meanwhile, Kim and Mooney [23] formulate grounded language learning as the induction of a probabilistic context free grammar (PCFG). They employ learned lexicons [22] to constrain the set of production rules, and are thereby able to scale PCFGs to the robot's action space. Kim and Mooney [24] extend their model by incorporating re-ranking using a discriminative classifier trained in a weakly supervised manner. Alternatively, Artzi and Zettlemoyer [25] and Artzi et al. [26] model the parsing problem using a combinatory categorical grammar that converts natural language utterances to their corresponding lambda calculus referents. Meanwhile, Mei et al. [12] introduce a neural sequence-to-sequence model that maps natural language instructions to action sequences. The model takes the form of an alignment-based recurrent neural network that encodes the free-form instruction and subsequently decodes the resulting representation into an action sequence based upon the visible environment. The model has the advantage that it does not use any specialized linguistic resources (e.g. parsers) or task-specific annotations (e.g. seed lexicons), and can be trained in an end-to-end fashion.

A second approach to grounded language acquisition is to map natural language utterances to their corresponding locations and objects in the robot's environment model and the actions in its action space. In this case, the environment is often represented as a hybrid map that expresses the spatial, topological and semantic properties associated with specific objects and locations in the environment (Section 3.3.3). These techniques learn a probabilistic model that captures the correspondence between each word in the free-form utterance and its matching referent in the world model (i.e. the symbols contained in the map and the robot's action space). The task of interpreting a new utterance is then one of performing inference in this learned probabilistic model. Kollar et al. [8] take this approach by constructing a generative model over a flat, sequential representation of free-form language consisting of both prespecified and learned models of adverbs, verbs and spatial relations.

Alternatively, Tellex et al. [6] propose a discriminative model that exploits the hierarchical, compositional structure of language. The Generalized Grounding Graph (G^3) builds a factor graph according to the parse structure of language (e.g. using the Cocke–Kasami–Younger algorithm [27] or the Stanford Parser [28]), resulting in a distribution over the space of groundings. The G^3 model then assumes that the groundings for linguistic elements are independent, factoring the distribution across individual phrases. These factored distributions take the form of a log-linear model that expresses the mapping between linguistic elements and their referent groundings in terms of binary correspondence variables. Again, the possible groundings are contained in a map that takes the form of a hybrid model of the environment. The model is trained on a corpus of utterances paired with their corresponding groundings. Consequently, the diversity of language that can be grounded is limited only by the rules of grammar and the richness of the training data. Inference in the G^3 model involves setting the correspondence variables to TRUE and searching over the space of possible groundings. This space can include all possible motions (actions) and can be arbitrarily large for non-trivial robot domains. Consequently, the computational cost of inference is proportional to the power set of the symbols in the world model (i.e. the objects, locations and actions). Approximating this space in a manner that affords efficient inference without sacrificing the diversity of the resulting symbols is challenging in practice.

Howard et al. [7] propose the Distributed Correspondence Graph (DCG) that extends the G^3 model in a manner that maintains the ability to interpret diverse natural language utterances, while improving the efficiency of inference. The DCG model grounds language into a discrete set of constraints (e.g. those suitable for a constraint-based motion planner, though symbols such as those employed by G^3 can also be used in place of the constraints) that can then be converted to continuous actions via a downstream planner, as opposed to approximating the continuum of paths with a set of samples. Rather than search over the space of groundings (constraints), inference in the DCG model involves searching over the correspondence variables associated with each constituent grounding. The number of factors that the DCG associates with each phrase is proportional to the number of conditionally independent components of the grounding. Consequently, the model distributes inference across multiple factors in the graphical model, which reduces the cost of inference from exponential to linear in the number of constraints. While more efficient than G^3, the runtime of DCG can still be prohibitive in the case of complex environments and tasks for which the number of grounding constituents is large. To address this limitation, Chung et al. [29] propose the Hierarchical Distributed Correspondence Graph (HDCG), an extension of the DCG that assumes that the space of candidate groundings can be constrained based upon the structure of the utterance in the context of the environment. The HDCG first employs the DCG to define a distribution over a set of rules that determine which symbols to consider as constituents. This distribution is then used by a second DCG to define the distribution over a reduced set of candidate groundings. Inference in the HDCG then proceeds as with DCG by searching over the set of correspondence variables. For tasks and environments for which the space of groundings is large, the HDCG outperforms both the DCG and G^3 models, without sacrificing accuracy.

3. Spatial reasoning and mapping in robotics

Robots moving and acting in the world need to be endowed with an understanding of their environment. This section gives an overview over the cognitive theories that have been developed about spatial representations that humans use and that have influenced models used in robotics. Furthermore, an introduction to qualitative spatial representations and to different approaches to creating maps for the use of robots is given. Finally, specific attention is given to the problem of deriving environment and semantic information from natural language descriptions.

3.1. Cognitive models

In order to create functional and efficient abstractions of space for intelligent robots, research has often looked to insights on the way humans and animals organize their spatial knowledge. Spatial representations for technical systems that are close to a human understanding of space are often easier to design and interpret, and facilitate information exchange with humans. In the following section, some basic terms and distinctions from the study of human and animal spatial cognition are highlighted that have shown to be helpful spatial models for technical systems. These ideas from cognitive studies have influenced research especially in hierarchical hybrid and semantic maps, which are discussed in Sections 3.3.3 and 3.3.4.

Two basic paradigms that have been used to describe human spatial cognition are those of *route* and *survey knowledge* [30,31]. Route knowledge represents space on a person-to-object basis, with the perspective of the visual system, while survey knowledge represents object-to-object relations at a global, world-centred view [32]. Survey knowledge is often also referred to by the term *cognitive map* [33]. On the lowest level, there is also *location knowledge*, which identifies a single location by a salient configuration of objects, which should be robust against change to reduce the uncertainty of the mental environment model [34,35].

Corresponding to these levels of spatial knowledge, *frames of reference* are defined. The *egocentric* frame takes the person-centred view, and the *allocentric* frame designates the world-centred view. Additional useful designations of frames are the relative, intrinsic and extrinsic frames, which stand for a person-centred, object-centred or global view, respectively [36].

Both route and survey knowledge are acquired when moving through an environment. Once learned, route and survey levels of spatial representation are tied to navigational tasks that they are most useful for. Route knowledge is used when navigating along a known path between identifiable places, where the navigational decisions have to be made at decision points to identify the correct continuation of the paths. On the other hand, survey knowledge is needed for pre-meditated navigational planning, where an unknown route to a target in a known or partially known environment has to be determined before actually executing the plan [37]. Insights on the different forms of spatial representations in cognitive models have influenced the research on hybrid maps for technical systems, in which environments are represented at multiple hierarchically organized levels.

While the ability to perform these tasks shows that both levels of knowledge are accessible, humans generally do not acquire full survey knowledge by exploration. Instead, they store topological relationships along with coarse, imprecise spatial relations between places that enable some Euclidean reasoning, for example about shortcuts through unexplored areas [30]. Experiments have shown that humans do not perform very well on recreating exact Euclidean measurements for known large-scale environments, with recalled distances distorted and affected by properties such as the number of landmarks on a route, and angles between alternative paths generally regressing towards right angles [38].

The non-Euclidean nature of the cognitive spatial model is further illustrated by the observations that recalled spatial relations may depend on an (imagined) vantage point, and that symmetric relations tend to be recalled asymmetrically depending on the properties of the involved objects. Cognitive load expended on retrieving spatial relations is also a factor that allows some insight into the mental spatial representation, which can be seen in some spatial relations being faster to recall than others, and in the fact that recalled spatial arrangements are more accurate when more information is asked for than when only partial information is inquired [38]. Tversky [38] calls the ensuing representation *spatial mental models*, eschewing the term 'cognitive map', since its properties are rather different from a standard Euclidean map. The notion of the representation being not fully Euclidean, but topological with added imprecise general spatial relations is corroborated by experiments in Virtual Reality, where participants have no problems navigating in worlds that are physically impossible [39]. For technical systems, formalisms that do not rely on quantitative Euclidean geometry have been explored with qualitative spatial representations as discussed in Section 3.2 and topological maps, which are introduced in Section 3.3.2.

A further important characteristic in the discussion of mental spatial models is their hierarchical nature. Non-hierarchical models rely on all spatial elements being stored at the same level, while hierarchical theories postulate that different areas or aspects of space are organized in different branches of a hierarchy. Hierarchical models can be strongly or partially hierarchical, where the latter permits additional attributes between elements of different branches. Experiments have shown that human spatial memory is likely to be organized partially hierarchically [40,41].

Another distinction that has proven useful in discussing human spatial cognition is the dichotomy of *propositional* and *imagistic* representations in human cognition [42]. Imagistic representations are common as spatial representation such as maps, sketches and figures. On the other hand, propositional representations are closer to the way spatial arrangements are expressed with language, and can be computed from imagistic representations.

3.2. Qualitative spatial representation and reasoning

Traditionally, formal mathematical reasoning about space primarily used the tools of topology and Euclidean or Cartesian geometry. While this type of reasoning about metric quantities is essential to many aspects of robotics, the disciplines of robotics and Artificial Intelligence have also developed an interest in a qualitative, symbolic system of reasoning about space. Arguably, a quantitative representation of space is closer to the cognitive and, in particular, the linguistic ways of representing space. Thus, it can bridge the gap between physical space where robots operate, and commonsense space, which are commonly addressed by language. Deliberate quantization can also bring robustness against noise and parameter errors. Dealing with metric values can also bring a degree of unwanted precision in the presence of uncertainty or in interaction scenarios. Finally, qualitative reasoning can be beneficial in terms of memory and computational complexity.

3.2.1. Qualitative representations of space

The aspects of space that need to be represented by a specific representation depend on the application, and many different formalisms for different requirements have been developed in the Qualitative Spatial Reasoning community.

A spatial representation consists of a set of basic spatial entities, and the relations that can be defined between them. Basic entities can be points, lines, line segments, rectangles, cubes or arbitrary regions of any dimension. The *dimensionality* of the basic entities and the space that is being modelled depends on the modelling depth and the application as well: As a practical example, a road is one-dimensional for trip planning, two-dimensional when planning overtaking behaviour, and three-dimensional when trying to estimate the curb position.

For brevity, the focus here is on representations used or usable for robotics. The basics of reasoning systems will be mentioned; more in-depth treatments can be found in the review articles by Vieu et al. and Chen et al. [43–45].

3.2.1.1. Mereotopology. An important set of qualitative spatial representations is based on the topology, i.e. relations of connectedness and enclosure, and mereology, i.e. the relations of parthood, of basic entities. These are known as mereotopological representations. In the following, some important instances of these formalisms will be briefly introduced.

As a very basic reasoning system, the *point calculus* for scalar values defines the relations $<, =, >$. The *interval calculus* [46] extends the reasoning in a single dimension to intervals, originally for reasoning about intervals in time. The 13 resulting binary relations are illustrated in Figure 2(a).

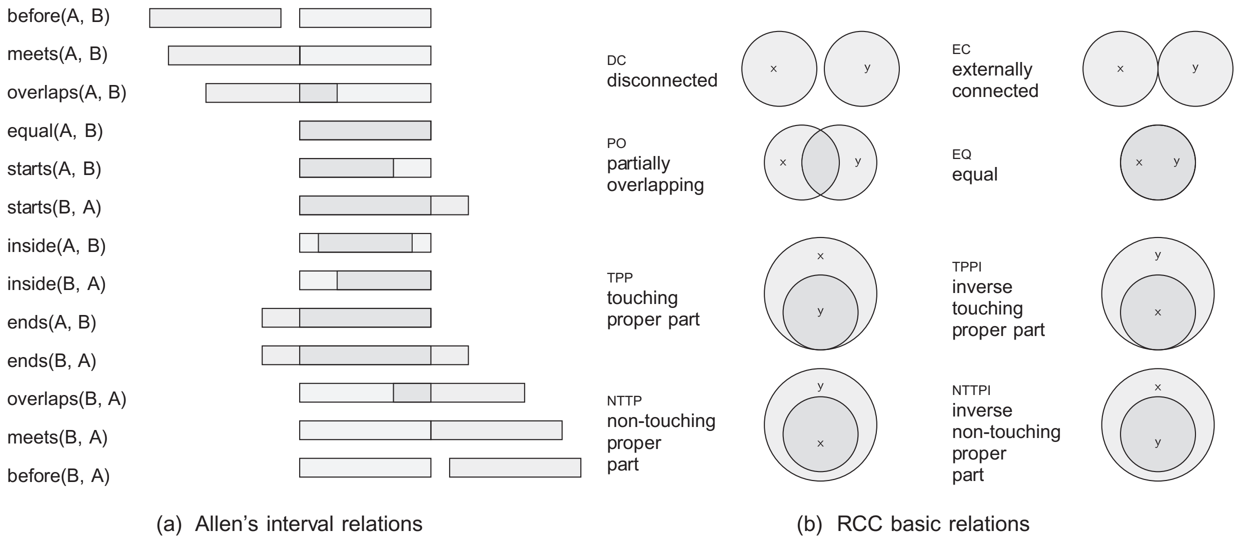
This type of reasoning is extended to two dimensions to form the *Rectangle Algebra* [47,48]. For this type of representation, shapes are projected to the axes of an extrinsically defined coordinate system, and the relations between the resulting intervals are constructed separately for each axis. It can be noted that this representation not only conveys topological information, but also has an orientation component. The spatial representation defined by the interval calculus has been analysed for cognitive adequacy by Knauff [49], who has shown that this representation aligns well with cognitive models.

Another important representation, which builds entirely on the notion of connectedness between regions, is the *Region Connection Calculus* (RCC) [50]. The canonical set of eight relations between two regions, which is known as RCC-8, that can be defined using connectedness is shown in Figure 2(b). Based on this set of base relations, different reasoning systems are possible depending on the intricacies of handling open and closed sets. Easier calculi are possible when border regions are not considered explicitly for reasoning [44]. A reduced set of base relations that does not take the boundary of regions into account comprises the five relations EQ, PO, PP (subsumes TPP and NTPP), PPI (subsumes TPPI and NTPPI), and DR (subsumes DC and EC). The cognitive plausibility of RCC-8 has been evaluated by Renz and Nebel [51] with the result that test subjects cluster pairs of regions according to the topological information it represents; thus showing its cognitive adequacy.

RCC can also serve as a good example for the concept of *conceptual neighborhoods* [52]. These define a system of neighbourhood for relations in a reasoning system, as opposed to a system of neighbourhood of objects. The conceptual neighbourhood of a relation contains all those relations that can be reached directly through transformations of one of the involved objects. For example, in RCC-8, the cognitive neighbourhood of the EC relation consists of DC and PO, but none of the other five relations. The conceptual neighbourhood depends on the transformations that are allowed, but can usually be used to limit the complexity of reasoning in a system, in particular if the reasoning entails the movement of objects.

3.2.1.2. Orientation calculi. Mereotopological relations are important for qualitative modelling of space; however, the information they can represent is limited. In the following, some simple representations that focus on orientation and direction between two objects, a primary object and a reference object, are introduced. For reasoning about orientation, a *frame of reference* is necessary. This can be either extrinsic, such as the cardinal directions given by a compass, or intrinsic to the problem. Frank [53] presents two spatial calculi based on cardinal directions: a cone-based one, where the angular direction towards the reference object is rounded to the nearest cardinal direction, and a projection-based one, which overlays the two pairs of half-planes associated with the two pairs of opposing cardinal directions. Both divide the plane by two lines intersecting in the reference point. They are illustrated in Figure 3(a) and (b), respectively. A generalization of this representation to an arbitrary number of lines is the Star calculus [54].

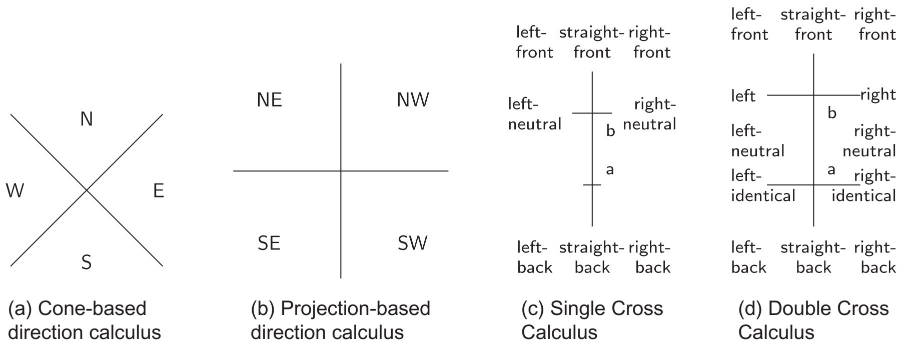
The single cross calculus and the double cross calculus [55,56] are example for relative orientation representation, where orientation is given as a ternary relation between a point on the plane, the *referent*, and the oriented line segment defined by the *origin* a and the *relatum* b. These representations are illustrated in Figure 3(c) and (d).



(a) Allen's interval relations

(b) RCC basic relations

Figure 2. Mereotopological calculi.



(a) Cone-based direction calculus

(b) Projection-based direction calculus

(c) Single Cross Calculus

(d) Double Cross Calculus

Figure 3. Orientation calculi. 3(a) and (b) are binary calculi; 3(c) and (d) are ternary. For the latter two, the origin is denoted by *a*, the referent by *b*, and the relatum can be any point in the plane.

The Cardinal direction calculus (CDC) [57] is a representation for relations between two extended regions in the plane. For the primary region, the minimum bounding rectangle is determined. The continuations of its edges separate the plane into nine sections, and the relation to the reference region is given by the set of sections the reference region intersects with. This is usually written as a 3 × 3 Boolean matrix, where each element indicates the non-emptiness of the corresponding intersection.

3.2.1.3. Other relations: size, distance and shape.

More predicates can be introduced to describe other aspects of objects or tuples of objects like relative or absolute distance, size, or shape. Distance and size properties are often based on quantization into a small number of categories like *far* or *close* or relative to other objects, as in a predicate $Closer((o_1, o_2), (o_3, o_4))$, which compares the distances of two pairs of objects. Reasoning about the shape of objects is a more recent development in qualitative spatial reasoning. The high complexity of most approaches and formalisms has led to only very simple formalisms being adapted into robotics applications, mostly based on representing objects by their centroid as a single point, their convex hull or a minimum bounding rectangle.

More recent work has focused on combining reasoning mechanisms from different calculi to jointly reason about different aspects of a spatial arrangement, e.g. topology and orientation using RCC-8 and RA or CDC simultaneously [58]. Another approach at combining reasoning about orientation and distance is the ternary point configuration (TPCC) calculus [59], which separates the plane into eight radial segments based on orientation with respect to the origin-relatum line segment, and additionally qualifies distance of the relatum to the referent as greater or smaller than the distance between origin and referent.

3.2.2. Qualitative spatial reasoning

Qualitative Spatial Reasoning is tightly connected with methods and results from mathematical logic. Reasoning systems can be formulated as *axiomatic systems*, which are generally first-order [43]. Due to the high complexity of reasoning in axiomatic systems, most spatial reasoning systems are defined as *relational algebras* or *calculi* [60]. These define a finite set of qualitative relations as described for various representation systems above. Usually, this set of base relations is required to be jointly exhaustive and pairwise disjoint (JEPD). If there are multiple possible base relations between a pair or tuple of objects, their relation is described by the disjunction of the individual base relations, which is generally denoted by the union of these relations. The full set of possible relations is the power set of the base relations, but it can also be restricted further, e.g. to ensure tractability. In addition to the relations, two important operations need to be defined to enable reasoning with a spatial algebra. For a binary calculus, the *converse* operation defines the relation S that hold for the pair (x, y) if relation R holds for the pair (y, x). The *composition* operator defines the relation for the pair (x, z) if the relations for pairs (x, y) and (y, z) are known. For many calculi, compositions of pairs of base relations are given in *composition tables*, which allow to determine the composition of arbitrary relations as the union of the compositions of the contained base relations by a simple table lookup. For ternary calculi, corresponding ternary operators have to be defined.

Different spatial reasoning problems can be posed. An important reasoning problem is the question whether there is an arrangement of objects that fulfil a set of given relations, which is known as *consistency checking* or *satisfiability*. From a computational standpoint, the consistency checking problem is a convenient choice, since many other decision or counting problems can be converted to this problem with polynomial complexity, and it has been studied for a long time for general-purpose logical formulations. Among these related tasks are the problem of finding one or all variable assignments that conform with a given constraint network, removing redundant constraints, or deciding whether a constraint network can be realized in a particular dimension, for example on a plane. For a propositional algebra, the consistency checking problem can be posed as a constraint satisfaction problem, and the corresponding methods from literature can be applied. The more restricted structure of spatial problems, as compared to general problems in logical formulations, allows to make simplifications to the reasoning process, which make the reasoning more efficient than general logical inference.

In many cases, the operators defined for the relations can be used in constraint propagation algorithms such as *path-consistency* and *algebraic-closure* [44] to decide the consistency of a constraint network. For decidable calculi, the complexity of inference is an important characteristic. For most calculi, deciding consistency is NP-complete, for example for the interval and rectangle algebras, as well as for RCC-8 and RCC-5 [44]. Reasoning problems in relation systems that are able to distinguish left from right is NP-hard [61,62], since in these cases, local consistency algorithms such as algebraic closure cannot decide consistency of a global scenario.

Research has been directed towards improving the practical applicability of the algorithms based on local consistency by trying to identify (maximal) tractable subsets of existing calculi, which make the backtracking search in algebraic-closure-based constraint processing more efficient. This can, for example, be done by searching for subsets that can be expressed using Horn clauses.

While generic constraint programming and logical inference tools can be used for spatial reasoning with the formalisms mentioned above [63], a number of specialized software toolboxes specifically for QSR have been developed. Among them are SPARq (Spatial Reasoning done Qualitatively) [64], GQR (Generic Qualitative Reasoner) [65], PelletSpatial [66], CHOROS [67] and the Qualitative Algebra Toolkit [68].

Wolter and Wallgrün [64] list some applications other than satisfiability checking via constraint processing that have practical relevance. Among these is *qualification*, the translation of a quantitative description of a scenario to a qualitative one considering rounding errors and noise, and the process of producing a (cognitively valid) rendition of a qualitative scenarion, e.g. for visualization. The qualification problem has also been addressed in the context of machine learning. Wollherr et al. [69] present a system based on Markov Logic Networks that estimates relations between objects in an annotated map of an urban environment. The approach put forward by Sjöö et al. [70] relies on a Graphical Model to determine the relations 'On' and 'In' between everyday objects. Support relations between objects in household scenes are the result of the estimation process performed by Silberman et al. [71].

3.3. Mapping in robotics

For an overwhelming majority of robotic tasks, robots need to develop and keep a representation of their surroundings based on sensor readings and possibly prior knowledge. Independent of the actual properties of this representation, this field of research is known as mapping. This section will give a brief overview of the different types of maps used in robotics, with a focus on representations that is wholly or partially qualitative in nature, and those that have a semantic component.

3.3.1. Metric maps

Learning and maintaining a metric map is central to many robotic tasks that rely on navigation. Based on early work by Smith and Cheeseman [72] and Leonard and Durrant-Whyte [73], the probabilistic formulation of the problem of building a globally consistent map has become known as the Simultaneous Localization and Mapping (SLAM) problem. Data from a very diverse range of sensors such as cameras, sonars, laser sensors, odometry and GPS need to be integrated over the course of potentially long exploration runs of a robot. A basic distinction between SLAM approaches is whether they are filter-based or graph-based. Filter-based SLAM emphasizes the temporal aspect of consecutive sensor measurements, while the graph-based variant emphasizes the spatial aspect by adding spatial constraints between robot poses where landmarks are jointly visible [74]. The underlying representation for the metric map can vary independently of the SLAM formalism, from landmark-based formalisms that store the positions of salient features in the environment to low-level representations like occupancy grids [75], surface maps [76], or raw sensor measurements like point clouds. A central challenge in SLAM is the data association problem of aligning real-world features across multiple sensor measurements. A good solution is important when the robot revisits a location where it has been before *(closing the loop)*, where a wrong association of features leads to an inconsistent map.

An example of a metric map generated using a SLAM algorithm from laser sensor data is shown in Figure 4.

3.3.2. Topological maps

Topological maps represent environments using a graph, the nodes of which represent *places* in free space, and edges denote traversability or connection in free space between pairs of nodes. There are different approaches to define the notion of places. One common approach is to define nodes in the topological map for each distinct part of the environment, separated by gateways such as doors or entryways. Other approaches define nodes every time the robot has travelled a fixed, specified distance, or use the structure of the *Generalized Voronoi Graph* [77]. An example of a topological map overlaid over a metric map of the same environment is given in Figure 4.

Like the problem of loop closure in metric mapping, topological mapping also faces the problem of identifying a place that is being revisited by the robot. This is known as the correspondence problem, which is made difficult in environments where possible matching candidate places look exactly or approximately the same in the available sensor data, which is known as *perceptual aliasing*.

An axiomatic theory and full ontological definition of topological maps were presented by Remolina [78]. Map learning is accomplished in a purely logical fashion using nested abnormality theories, which use causal, topological and metrical properties of the environment to determine the topological map as the minimal map that explains the robot's percepts.

For topological maps, the space of maps is combinatorial, but still much smaller than the space of all possible metric maps. Thus, multi-hypothesis or probabilistic methods that keep a distribution over all possible hypotheses are possible. The probabilistic topological map [79] keeps a distribution over all possible topologies using a Rao-Blackwellized particle filter. Wallgrün [80] presents a topological mapping algorithm that exclusively relies on qualitative spatial reasoning to keep track of multiple hypotheses about the structure of the environment. Two different qualitative reasoning calculi are compared on the task of building a consistent map from sparse qualitative connection information, using various constraints on the spatial structure of the resulting network to reduce the size of the search space. An extensive review of SLAM in topological maps is presented by Boal et al. [81]

A topological map is also a convenient and efficient representation of environments for route-based navigation. The *route graph* [82] is a topological map designed for this purpose. Its nodes are *places* connected by *courses*, which together make up *route segments* and entire *routes*. Elements of the route graph can be labelled with additional information to convey categories such as the medium of transport to be used on a particular route segment.

3.3.3. Hybrid maps

Each type of map has its own strengths, and the term hybrid maps describes approaches that combine different representations to form a stronger overall environment representation. Buschka and Safiotti [83] define a hybrid map as a tuple of maps, where usually one is metric and one is topological. The benefit of the hybrid maps comes from links between the two, which maps objects from one map to objects of the other. Other combinations are possible, however. Some advantages of this combination of different maps are improved loop closure, lower complexity, improved localization, easier planning and high-level reasoning, and the possibility to define a system state on different levels. A particular benefit for hybrid maps can be the possibility to relax the requirement for global consistency of metric representations, and keep a consistent topological representation instead, which can have computational advantages.

An early instance of hybrid maps that has received much attention is the *Spatial Semantic Hierarchy* (SSH)

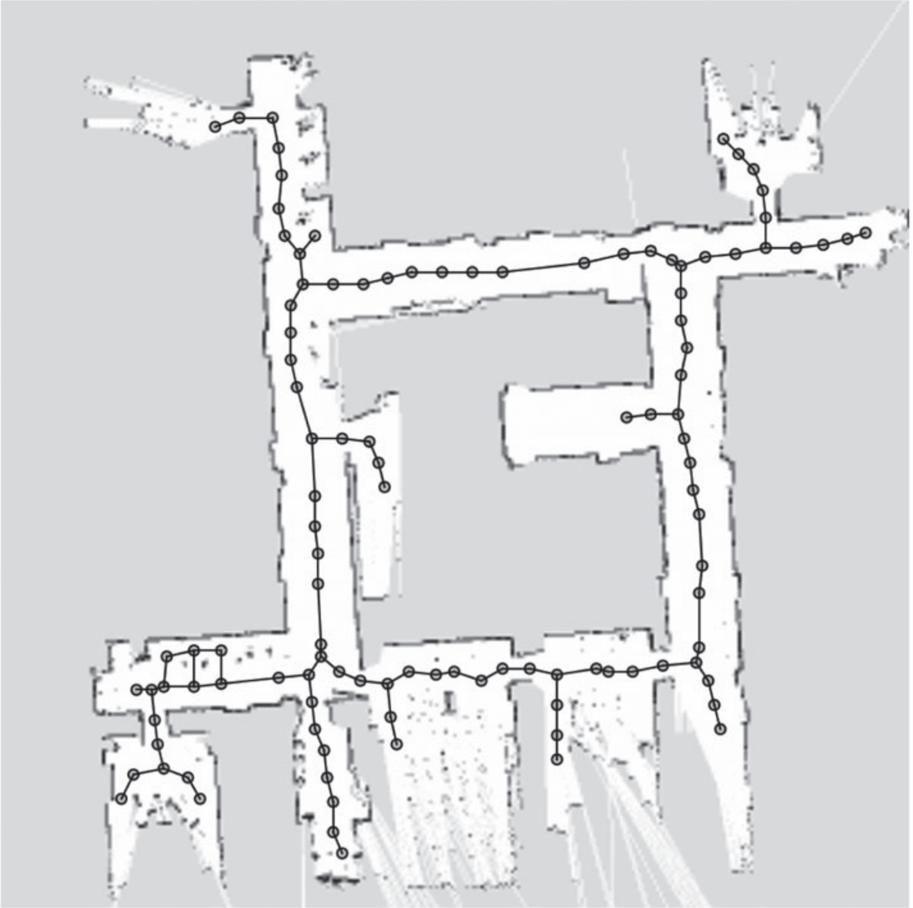


Figure 4. Metric map with overlaid topological structure. The metric map is generated with a SLAM algorithm on laser data. For the topological map, the structure of the environment is extracted from the metric map as the Voronoi graph, and edges are placed at junctions of the Voronoi graph as well as at constant intervals between junctions.

[84,85], which is inspired by cognitive studies about human spatial representations. It models space on four levels, where each level depends on information from the levels below: The lowest level is the *control level*, which defines a dynamical system where *distinctive state*, known poses in the environment, can be reached by hill-climbing, and trajectories between these states or their attractor regions can be followed. Sensor percepts, so-called *views*, allow the unique identification of these states. On the *causal level*, a finite state automaton is defined, in which state transitions correspond to movements between places. The states and edges of this automaton map to places and paths on the *topological level*. Finally, the *metric level* stores a geometric representation, such as occupancy grids, for each place, which can be combined to form a global metric map. Not all levels must be present or available at all times, depending on whether the region has been explored, availability of computation resources, sensor data etc. The hierarchical structure of the SSH is illustrated in Figure 5.

This formalism was extended with ideas from the SLAM community to form the *hybrid SSH* [86], where local maps are used instead of views to identify places locally. This allows more tolerance for noise and dynamics in the environment in small-scale space (within the sensor horizon), but does not require loop closure in large-scale space, where the topological representation can be used. Beeson et al. [87,88] integrate semantic aspects in the hybrid SSH by reasoning about gateways and integrating the approach with a natural language interface.

3.3.4. Semantic maps

While metric and topological maps only describe the spatial arrangement of an environment, additional information is necessary for many robotics tasks. Semantic maps broaden the scope of the elements represented in a map to instances of objects, their categories and possible attributes, and to common-sense knowledge about entities represented in the map [89]. This is particularly

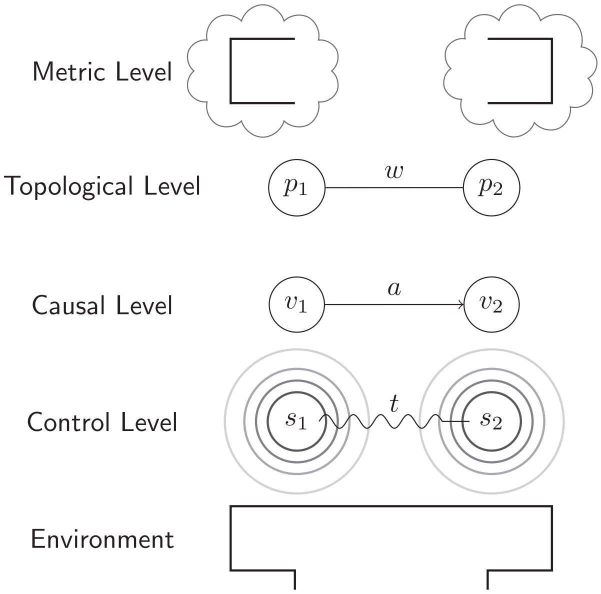


Figure 5. Illustration of the SSH. The environment is represented by two distinctive states s on the control level, which each have a region of attraction and are connected by a trajectory t. On the causal level, the distinctive states can be identified from sensor percepts with the views v, and transitioning from one state to the other is possible by taking action a. On the topological level, the environment has two places p, which are connected by a path w. On the metric level, a local metric representation for each place can be stored.

beneficial in applications where a higher level understanding of scenarios is necessary, and when applications require human–robot interaction.

Semantic mapping requires that information about the objects in an environment is available for reasoning. Like the spatial information represented in metric maps, this information is often inferred from typical sensor data, coming from 2D and 3D sensors including sonars, LIDAR scanners, monocular, stereo and omnidirectional camera set-ups and RGB-D sensors. For building semantic maps, high-level techniques like character recognition [90], interaction with humans [91,92] or databases of common-sense knowledge are used as additional modalities. Techniques for using language to acquire semantic knowledge from human interaction partners for map building and annotation are discussed in detail in Section 3.4, while this section focuses on the use of purely technical sensor streams. Advanced perception algorithms for object detection, segmentation and classification have been adapted from the robotic perception and computer vision literature and developed specifically for semantic mapping. This area of research is out of scope for this survey, which instead focuses on the mapping and representational aspects of semantic mapping. An overview of perception approaches to semantic mapping is given by Kostavelis and Gasteratos [93].

There is a broad range of different types of semantic information in maps, depending on factors like the intended application, the sensor repertoire of the robot, and the type of environment that is being mapped. A broad categorization can be made between maps that add semantic attributes to objects in the map, maps that categorize regions, and those that add semantic categories to sensor percepts on the trajectory of an exploration of the environment.

3.3.4.1. Object-based semantic maps.

The first category of mapping approaches relies on techniques for scene interpretation to label objects in the robot's sensor stream and localize them using a metric environment representation. In this vein, Limketkai et al. [94] label line segments in a metric map as *wall*, *door* or *other* using a relational Markov Network that uses unary and pairwise as well as higher order spatial relations between objects as input. Nüchter and Hertzberg [95] use a constraint network expressing common properties of spatial arrangements of planes in buildings to classify points from a point cloud into different categories (ceiling, wall, floor, etc.). Additionally, other objects like humans and printers are detected and classified, forming a semantically annotated 3D point cloud. A more perception-oriented approach is presented by Meger et al. [96], where objects detected and classified based on camera images are mapped into their locations in a global occupancy grid. A place categorization method based on object co-occurrence statistics and clustering of objects to places based on spatial distance and a Bayesian criterion for the number of clusters is presented by Viswanathan et al. [97].

3.3.4.2. Region-based semantic maps.

Many semantic mapping approaches discretize space to a topological map on some level of their hierarchy of maps into areas of conceptual meaning, which are often called *places*. One distinguishing factor between semantic mapping approaches is the way places (or generally nodes in a corresponding topological formulation) are separated.

Some mapping systems recognize that distinct places are usually separated by gateway structure like doors, and devise ways of identifying these structures. The work by Vasudevan et al. [98] builds on a way to recognize objects and doors. A probabilistic relative object graph tracks object positions relative to the place they are found in, and allows to compute probabilistic spatial relations between them. These graphs are used for place recognition, and classification of places is based on the types of objects present in the scene. An extension of the approach [99] uses spatial information even in the reasoning about place categories, where the category of objects, the number of occurrences and simple spatial relationships between them are taken into account when classifying rooms into different categories. Places can have hierarchical structure, so places that afford particular functions, such as a 'printer area' or a 'couch area' can be contained in a more general place of type 'office'. A more detailed subcategorization of gateways is part of the approach by Rituerto et al. [100], which distinguishes the categories *door*, *stairs*, *elevator* and *jamb*.

Other work on segmenting metric maps of indoor environments into semantically meaningful clusters is based on a semi-supervised scheme employing a Markov process model [101], spectral clustering on a graph that encodes visibility between randomly sampled free space points in its edges [102], clustering based on mutual information [103] and fitting models of basic room shapes in a Markov Chain [104].

Pronobis et al. [105] present an approach for semantic mapping where low-level classifiers are used to determine properties of areas such as room shape, size or the existence of certain objects, which are then used to determine room types in a probabilistic reasoning step through inference in a chain graph. Later work [106] includes this technique in a complete semantic mapping system for indoor environments. It accepts multimodal sensor input, including input from humans via natural language, which is treated as a separate sensor modality with an appropriate sensor model. For mapping an environment, first a global metric and topological map are built. Places are created at constant distance intervals on the trajectory of the robot, which are further clustered into rooms separated by door places.

3.3.4.3. Semantic maps from segmenting the robot trajectory and from user interaction.

Environments can also be segmented into semantically distinct regions in an online process by recognizing significant changes in the surroundings of the robot while it is exploring the environment. Mozos et al. [107] use a boosting classifier in combination with a hidden Markov model to segment the trajectory of the robot into contiguous segments, where the surrounding environment corresponds to a place. The same classifier together with probabilistic smoothing techniques is used to cluster an occupancy grid into areas of semantic meaning.

Sünderhauf et al. [108] create a semantic occupancy grid by classifying camera data with a convolutional neural network and propagating the classification results along laser beams similar to the probability update in a standard occupancy grid. A number of other approaches rely on classifying and segmenting environments based on the stream of images from the robot's sensors. A topic modelling approach is used by Murphy and Sibley [109], while Ranganathan & Dellaert [110] use an information-theoretic approach. A string encoding of appearance features is used for segmentation of places by Tapus and Siegwart [111].

A segmentation of an environment can also be determined through user interaction. Thrun et al. [112] determine distinctive places by having users push a button to communicate that the robot has arrived at a distinctive place. Nieto-Granda et al. [113] define the assignment of places to the environment as a mixture-of-Gaussians distribution, where the centres of the individual components are taught by human interaction partners during a tour of the surroundings.

3.3.4.4. Ontologies and high-level reasoning.

High-level reasoning about the map and its elements require the robot's understanding of task-relevant concepts as they are used in human reasoning and in language in their own right, and their connection to the corresponding sensor impressions, which is one aspect of the *symbol grounding problem* [13]. A common trait to many approaches that combine metric or topological mapping with reasoning on higher level concepts is the introduction of an ontology, where world knowledge is stored in a taxonomy and sensor experience from the map is encoded to domain knowledge, which is then linked based on overlapping semantic attributes. Zender et al. [114] present one instance of such an approach, where ontological reasoning complements a multi-level spatial map to form a conceptual representation of an indoor environment. The ontology is handcrafted to represent different room types and the typical objects present in them. Grounding instances of places and objects found in the environment allow to refine knowledge about the environment, and to generate a linguistic representation of a scene, for example for clarification dialogues. Hawes et al. [115] builds on this mapping approach to build a system that can identify, reason about and autonomously fill gaps in its knowledge about the environment, both its structure and conceptual knowledge as well as semantic knowledge such as room categories.

The multi-hierarchic semantic map for indoor environments presented by Galindo et al. [116] maintains hierarchical representations both for spatial and for semantic knowledge, where the latter takes the form of an ontology. The bottom level of the spatial hierarchy is made up of an occupancy grid, which is segmented into rooms using image-processing techniques to form a topological map. Based on properties of the rooms and objects found in them, regions can be classified and anchored to the corresponding concepts in the ontology, and further reasoning can be performed based on the world knowledge stored there. Tenorth et al. [117], Pangercic et al. [118] and Riazuelo et al. [119] introduce semantic mapping approaches which link objects detected in the environment to a large database of commonsense, probabilistic knowledge including high-level attributes like affordances or object articulations, which allows to execute high-level plans like 'clear the table'. A different type of world knowledge is tapped by works that use the large-scale structure of buildings to determine the function of rooms by their typical topology or by conditioning classifiers on the type of building [120–122].

3.3.4.5. Outdoor semantic mapping. While the research in semantic mapping has primarily been directed towards the application in indoor environments, outdoor environments have been addressed as well, using a similar array of techniques. Lang et al. [89] apply a multilevel spatial representation along with ontological reasoning to urban outdoor environments. Multiple other methods to add semantic labels to metric maps of urban road environments have been presented, e.g. [123–125]. A topological description of environments for off-road driving is defined by Bernuy and Ruiz de Solar [126]. Wolf and Sukhatme [127] create a terrain map of a robot's driving surface that is annotated with semantic labels, and includes traversability information. In addition to common appearance features for static environments, observed dynamics are included as activity measurements to distinguish different environments in that work.

3.4. Learning semantic maps from natural language descriptions

Grounding-based approaches to natural language understanding, such as those described in Section 2.2, require a priori knowledge of the space of symbols that express the objects and locations that comprise the robot's environment and the actions that it is able to execute. Importantly, these symbols should express low-level properties (e.g. the metric pose of each object and location) needed for planning and control, as well as higher level semantic properties (e.g. the colloquial name associated with each object and location) that are integral to language grounding. Semantic maps (Section 3.3.4) provide environment representations that are useful in expressing these symbols. These symbolic representations are typically created either by manually labelling metric maps of the environment or by automatically inferring semantic properties from the robot's sensor stream as part of SLAM framework. As described above, the latter involves populating the semantic map with scene attributes extracted using scene classifiers [107,108,128,129] and object detectors [96,99,130,131].

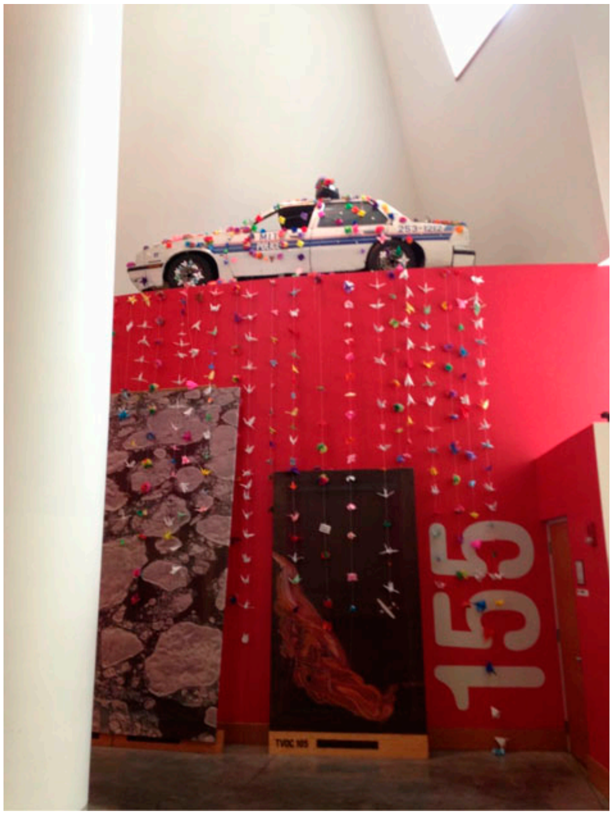


Figure 6. General-purpose visual classifiers would find it difficult to recognize the police car uniquely marking a campus building lobby.

However, scene and object classifiers alone are not sufficient to infer many of the semantic properties that people commonly associate with their environments. This is partially a consequence of the fact that their effectiveness is a function of the richness of the available training data. As such, they perform best when the environments have similar appearance and regular geometry (particularly for LIDAR-based classifiers), and when the objects are drawn from a common set. Even in structured settings, it is not uncommon for the regions to be irregular and for the objects to be difficult to recognize, either because they are out of context or are singletons. For example, Figure 6 shows a police car elevated 15 m off the floor that marks the lobby of a campus building. Few image-based object recognition methods would be able to recognize the car's presence, however students and faculty often use it as a salient landmark. Similarly, scene classification doesn't provide a means to infer the specific labels that humans use to refer to a location, such as 'Carrie's office' or the 'Kiva conference room.' Additionally, the information that can be extracted is inherently limited to the field-of-view of the robot's sensors and the geometry of the environment (e.g. line-of-sight).

An effective means of overcoming these limitations are to use human supervision to learn properties that are difficult to extract from traditional sensor streams, such as the common name and class of different spaces and objects in the environment. Recognizing the efficiency of language as a means of providing supervision, researchers have developed methods that integrate user-spoken cues within the semantic mapping framework [106,114,132–134]. For example, Zender et al. [114] allow people to assign labels to objects nearby the robot using speech. These labels are then combined with a LIDAR-based scene classifier to generate semantic maps of indoor environments that express the relationship between room categories and the types of objects that they contain. Alternatively, Hemachandra et al. [133] propose a framework that enables robots to acquire spatial-semantic environment models by autonomously following humans as part of a narrated tour. During the tour, the robot extracts object and location labels from the user's utterances and uses these labels to augment a hybrid map of the environment that expresses its metric, topological and semantic properties. Meanwhile, Pronobis and Jensfelt [106] introduce a semantic mapping framework that integrates semantic information extracted from multiple modalities, including human speech and object classifiers. They fuse this information with observations of the metric and appearance properties of the scene to arrive at a joint spatial-semantic environment model.

Methods that employ language as a form of weak supervision tend to be limited to templated, domain-specific descriptions that reference the robot's immediate surround. Walter et al. [92,135] propose a probabilistic framework that learns a hybrid metric, topological and semantic map from natural language descriptions. The Semantic Graph takes the form of a region-based semantic map (Section 3.3.4.2) in which nodes in the topology define distinct places and edges express observed spatial relations between places. Poses are associated with each node and the resulting topology gives rise to a pose graph, similar to that employed for metric SLAM. Semantic properties are also associated with each node in the topology. The Semantic Graph algorithm employs a learned G^3 model of free-form descriptions to reason over utterances for which the structure is limited only by the rules of grammar and the diversity of the training data. The method is able to learn properties of the environment including labels and spatial relations not only for the robot's immediate surroundings, but also for distant areas as well as regions of the environment that the robot has not yet visited. The framework uses a Rao-Blackwellized particle filter to maintain a factored distribution over the metric, topological and semantic properties of the environment based upon the stream of natural language descriptions and sensor-based observations. By performing joint inference over the metric, topological and semantic layers in the hybrid map, the framework exploits updates to any one layer to improve the other layers in the hierarchy. These updates can come in the form of spatial-semantic observations extracted from the user's descriptions during a guided tour, as well as information gleaned from the robot's traditional sensor streams [136]. As an example, the addition of semantic information from language can be used to recognize loop closures and thereby improve the metric and topological accuracy of the map. Hemachandra and Walter [137] build upon the Semantic Graph framework incorporating a mechanism that enables the robot to engage the user in dialogue during the narrated tour. They propose an information-theoretic algorithm that decides whether and what question to ask of the user in order to minimize the uncertainty in the learned distribution over the spatial-semantic map.

4. Conclusions

As the previous sections have attempted to show, large advances have been made towards robotic spatial reasoning and interaction in real-world applications. This review article has highlighted both the interaction and the representational aspects of this particular problem. Dialogue Systems enable natural language as an efficient and natural modality for situated user interaction. They specifically require robust parsing and understanding of spatial language to reconcile information gathered in interaction with the robot's world knowledge for task planning or learning of new facts. Different metric and symbolic types of representation that support both navigation and interaction have been discussed along with their motivation from cognitive studies of human spatial representations.

In spite of the large body of work in this area, there remain challenging research questions in this multidisciplinary field of study. First, robot understanding of natural language is incomplete and error-prone. Dialogue input components are cumbersome and ASR is especially difficult in noisy environments, such as outdoors or in crowded rooms. Robotic vision systems are also prone to error. Unambiguous recognition and interpretation of spatial scenes are essential to resolve language in context. In this special issue, Schuette et al. [138] use dialogue to resolve perception errors as e.g. induced by object recognition or mismatches in the understanding of spatial relations. However, not only robotic perception is error-prone. Also humans can make mistakes. Topp [139] addresses the issue of learning from a human teacher in situated communications, where the robot needs to detect inconsistencies in the information provided by the human teacher. In the work put forward by

Rangel et al. [140], this information is not learned from dialogue, but instead learned from manually annotated images. These annotations are used to learn a spatial representation of the environment from the visual processing of raw images. Finally, physically situated interaction brings further challenges that are not covered in this special issue. For example, situated HRI often includes multi-party conversation, where the challenge is to manage engagement [141,142]. In addition, generating situated language is a challenge with respect to, e.g. providing complex way-finding directions or generating referring expressions [143,144]. Finally, open scenarios frequently evoke out-of-domain queries, which are a challenge for Natural Language Understanding [142].

The fact that interactive robots are highly integrated, complex systems raises the problem of standardization for evaluation and comparison of different approaches on a common ground. Different system often varies greatly in their intended application, for example in the type and scope of semantic information that is considered, and in the intended domain, such that direct comparisons are often difficult. Recent efforts to make large data-sets available and to standardize evaluation protocols could be a step towards more standardized evaluation scenarios. The large variety of possible application domains has also lead to some scenarios receiving more scientific attention than others. In particular, indoor office-like environments often serve as a testbed due to their accessibility, their relatively structured nature and the availability of technical solutions for objects and place classification, while less attention has been devoted to outdoor scenarios. These more complex domains could also benefit from extending work on including common-sense and location-based knowledge from a wide range of sources, such as open-source maps and databases as well as natural language online resources. In the present special issue, Landsiedel and Wollherr [145] present a method to augment hybrid maps of urban environments by fusing information from 3D point clouds on the sensor level and human-annotated Open Data from OpenStreetMap.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research received funding from the EPSRC projects GUI [grant number GUI EP/L026775/1], [grant number DILiGENt EP/M005429/1], [grant number MaDrIgAL EP/N017536/1]; was supported in part by the Robotics Consortium of the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement [grant number W911NF-10-2-0016].

Notes on contributors

C. Landsiedel is currently studying towards a Doctor of Engineering degree in robotics at the Chair of Automatic Control Engineering at the Technical University of Munich, working in the field of semantic mapping for autonomous robots in urban environments. He received the Diplom-Ingenieur degree from the Technical University of Munich in 2011. His research interests include semantic mapping and spatial reasoning, machine learning and statistical relational reasoning.

V. Rieser is an associate professor at Heriot-Watt University, Edinburgh. She currently leads her own research group in Conversational Agents. Before joining HWU in 2011, she was a postdoctoral researcher at Edinburgh University. She gained her PhD from Saarland University in 2008 with “summa cum laude” and was awarded the Dr.-Eduard-Martin prize for outstanding research. She has authored over 70 papers in the areas of natural language processing, machine learning and multi-agent systems. She is a steering group member for Natural Language Generation (SIGGEN) and is on the editorial board for Science Scotland. Recently, she and her team were selected to enter the Amazon Alexa Challenge, where they compete to build an open domain social bot.

M. Walter is an assistant professor at the Toyota Technological Institute at Chicago (TTIC), where he directs the Robot Intelligence through Perception Laboratory. Before joining TTIC, he was a research scientist in the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology (MIT). He earned his PhD from MIT in 2008. His interests revolve around the realization of intelligent, perceptually aware robots that are able to act robustly and effectively in unstructured environments, particularly with and alongside people. His research focuses on machine learning-based solutions that allow robots to learn to understand and interact with the people, places and objects in their surroundings.

D. Wollherr received the Diplom-Ingenieur degree and the Doctor of Engineering degree, both in Electrical Engineering, and the Habilitation degree from Technical University Munich, Germany, in 2000, 2005, and 2013, respectively. From 2001 to 2004, he was a research assistant with the Control Systems Group, Technische Universität Berlin, Germany. In 2004, he was with the Yoshihiko Nakamura Laboratory, The University of Tokyo, Japan. Since 2014, he has been a professor with the Chair of Automatic Control Engineering, Department of Electrical and Computer Engineering, Technical University Munich.

His research interests include automatic control, robotics, autonomous mobile robots, human-robot interaction, and humanoid walking.

References

[1] Rieser V, Lemon O. Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation. Theory and applications of natural language processing. Berlin: Springer; 2011.

[2] Jurafsky D, Martin JH. Speech and language processing. Upper Saddle River (NJ): Prentice Hall; 2000.

[3] McTear MF. Towards the conversational user interface. Berlin: Springer; 2004.

[4] Clark HH, Brennan SE. Grounding in communication. Perspect Soc Shared Cognition. 1991;1991(13):127–149.

[5] Paek T, Horvitz E. Uncertainty, utility, and misunderstanding: a decision-theoretic perspective on grounding in conversational systems. AAAI fall symposium on psychological models of communication in collaborative systems. Cape Cod (MA): AAAI; 1999.

[6] Tellex S, Kollar T, Dickerson S, et al. Understanding natural language commands for robotic navigation and mobile manipulation. In: Proceeding of National Conference on Artificial Intelligence (AAAI). San Francisco (CA): AAAI; 2011. p. 1507–1514.

[7] Howard T, Tellex S, Roy N. A natural language planner interface for mobile manipulators. In: Proceeding of IEEE International Conference on Robotics and Automation (ICRA). Hong Kong: IEEE; 2014. p. 6652–6659.

[8] Kollar T, Tellex S, Roy D, et al. Toward understanding natural language directions. In: Proceeding ACM/IEEE International Conference on Human-robot Interaction (HRI). Osaka, Japan: ACM; 2010. p. 259–266.

[9] Matuszek C, Fox D, Koscher K. Following directions using statistical machine translation. In: Proceeding ACM/IEEE International Conference on Human-robot Interaction (HRI). Osaka, Japan: ACM; 2010. p. 251–258.

[10] Duvallet F, Walter MR, Howard TM, et al. Inferring maps and behaviors from natural language instructions. In Hsieh MA, Khatib O, Kumar V, editors. Proceeding International Symposium Experimental Robotics (ISER). Vol. 109 of Springer Tracts in Advanced Robotics. Marrakesh, Morocco: Springer; 2014. p. 373–338.

[11] Hemachandra S, Duvallet F, Howard TM, et al. Learning models for following natural language directions in unknown environments. In: Proceedings of IEEE International Conference on robotics and automation (ICRA). Seattle (WA): IEEE; 2015 May; p. 5608–5615.

[12] Mei H, Bansal M, Walter MR. Listen, attend, and walk: neural mapping of navigational instructions to action sequences. In: Schuurmans D, Wellman MP, editors. Proceedings of National Conference on Artificial Intelligence (AAAI). Phoenix (AZ): AAAI; 2016. p. 2772–2778.

[13] Harnad S. The symbol grounding problem. Physica D. 1990;42(1–3):335–346.

[14] Winograd T. Procedures as a representation for data in a computer program for understanding natural language [Ph.D. thesis]. Cambridge (MA): MIT; 1971.

[15] Skubic M, Perzanowski D, Blisard S, et al. Spatial language for human-robot dialogs. IEEE Trans Syst Man Cybern Part C: Appl Rev. 2004;34(2):154–167.

[16] MacMahon M, Stankiewicz B, Kuipers B. Walk the talk: connecting language, knowledge, and action in route instructions. In: Proceedings of National Conference on Artificial Intelligence (AAAI). Boston (MA): AAAI; 2006. p. 1475–1482.

[17] Hsiao K, Tellex S, Vosoughi S, et al. Object schemas for grounding language in a responsive robot. Connection Sci. 2008;20(4):253–276.

[18] Branavan S, Chen H, Zettlemoyer LS, et al. Reinforcement learning for mapping instructions to actions. In: Proceedings of Annual Meeting Association for Computational Linguistics (ACL). Singapore: Suntec; 2009 August. p. 82–90.

[19] Shimizu N, Haas A. Learning to follow navigational route instructions. In: Proceedings of Interantional Joint Conference Artificial Intelligence (IJCAI). Pasadena (CA): AAAI; 2009. p. 1488–1493.

[20] Vogel A, Jurafsky D. Learning to follow navigational directions. In: Proceedings of Annual Meeting Association for Computational Linguistics (ACL). Uppsala, Sweden: The Association for Computational Linguists; 2010. p. 806–814.

[21] Mooney RJ. Learning to connect language and perception. In: Proceedings of National Conference on Artificial Intelligence (AAAI). Chicago (IL): AAAI; 2008. p. 1598–1601.

[22] Chen DL, Mooney RJ. Learning to interpret natural language navigation instructions from observations. In Burgard W, Roth D, editors. Proceedings of National Conference on Artificial Intelligence (AAAI). San Francisco (CL): AAAI; 2011.

[23] Kim J, Mooney RJ. Unsupervised PCFG induction for grounded language learning with highly ambiguous supervision. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP). Jeju Island, Korea: The Association for Computer Linguistics; 2012. p. 433–444.

[24] Kim J, Mooney RJ. Adapting discriminative reranking to grounded language learning. In: Proceedings of Annual Meeting Association for Computational Linguistics (ACL). Sofia, Bulgaria: The Association for Computer Linguistics; 2013. p. 218–227.

[25] Artzi Y, Zettlemoyer L. Weakly supervised learning of semantic parsers for mapping instructions to actions. Trans Assoc Comput Linguistics. 2013;1:49–62.

[26] Artzi Y, Das D, Petrov S. Learning compact lexicons for CCG semantic parsing. In Moschitti A, Pang B, Daelemans W, editors. Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: ACL; 2014. p. 1273–1283.

[27] Younger D. Recognition and parsing of context-free languages in time n^3. Infor Control. 1967;10(2):189–208.

[28] Klein D, Manning C. Accurate unlexicalized parsing. In: Proceedings of Annual Meeting Association for Computational Linguistics (ACL). Sapporo, Japan: ACL; 2003 Jul. p. 423–430.

[29] Chung I, Propp O, Walter MR, et al. On the performance of hierarchical distributed correspondence graphs for efficient symbol grounding of robot instructions. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany: IEEE; 2015 Oct. p. 5247–5252.

[30] Chrastil ER, Warren WH. From cognitive maps to cognitive graphs. PloS one. 2014;9(11):e112544.

[31] Siegel AW, White SH. The development of spatial representations of large-scale environments. Adv Child Dev Behav. 1975;10:9–55. Available from: http://www.sciencedirect.com/science/article/pii/S0065240708600075

[32] Golledge RG, Jacobson RD, Kitchin R, et al. Cognitive maps, spatial abilities, and human wayfinding. Geog Rev Japan Ser B. 2000;73(2):93–104.

[33] Tolman EC. Cognitive maps in rats and men. Psychological Rev. 1948;55(4):189.

[34] Krieg-Brückner B, Röfer T, Carmesin HO, et al. A taxonomy of spatial knowledge for navigation and its application to the Bremen autonomous wheelchair. In: Freksa C, Habel C, Wender KF, editors. Proceedings of Spatial Cognition. Berlin: Springer; 1998. p. 373–397.

[35] Chown E. Making predictions in an uncertain world: environmental structure and cognitive maps. Adapt Behavior. 1999;7(1):17–33.

[36] Tversky B. Levels and structure of spatial knowledge. In Kitchin R, Freundschuh SM, editors. Cognitive Mapping: Past Present Future. New York (NY): Routledge; 2000. p. 24–34.

[37] Werner S, Krieg-Brückner B, Mallot HA, et al. Spatial cognition: the role of landmark, route, and survey knowledge in human and robot navigation. Informatik '97 - Informatik als Innovationsmotor. Aachen, Germany: Springer; 1997. p. 41–50.

[38] Tversky B. Cognitive maps, cognitive collages, and spatial mental models. In: Frank AU, Campari I, editors. Spatial information theory: a theoretical basis for GIS. Berlin: Springer; 1993. p. 14–24.

[39] Kluss T, Marsh WE, Zetzsche C, et al. Representation of impossible worlds in the cognitive map. Cognitive Process. 2015;16(1):271–276.

[40] McNamara TP. Mental representations of spatial relations. Cognitive Psychology. 1986;18(1):87–121.

[41] Hirtle SC, Jonides J. Evidence of hierarchies in cognitive maps. Memory Cognition. 1985;13(3):208–217.

[42] Hobbs JR, Narayanan S. Spatial representation and reasoning. In Encyclopedia of cognitive science. Hoboken (NJ): Wiley Online Library; 2002.

[43] Vieu L. Spatial representation and reasoning in artificial intelligence. In: Stock O, editor. Spatial and temporal reasoning. Dordrecht: Springer; 1997. p. 5–41. DOI: 10.1007/978-0-585-28322-7_1

[44] Cohn AG, Renz J. Qualitative spatial representation and reasoning. Handbook Knowl Representation. 2008;3:551–596. Available from: http://www.sciencedirect.com/science/article/pii/S1574652607030131

[45] Chen J, Cohn AG, Liu D, et al. A survey of qualitative spatial representations. Knowledge Eng Rev. 2015;30:106–136. Available from: http://journals.cambridge.org/article_S0269888913000350

[46] Allen JF. Maintaining knowledge about temporal intervals. Commun ACM. 1983;26(11):832–843.

[47] Balbiani P, Condotta JF, del Cerro LF. A new tractable subclass of the rectangle algebra. In: Proceedings of International Joint Conference on Artificial Intelligence (IJCAI). Stockholm, Sweden: Morgan Kaufman; 1999. Vol. 99, p. 442–447.

[48] Guesgen HW. Spatial reasoning based on Allen's temporal logic. Berkeley, CA: International Computer Science Institute; 1989. (Tech Report; TR-89-049).

[49] Knauff M. The cognitive adequacy of Allen's interval calculus for qualitative spatial representation and reasoning. Spatial Cognition Comput. 1999;1(3):261–290.

[50] Randell DA, Cui Z, Cohn AG. A spatial logic based on regions and connection. In: Proceedings of International Conference on Knowledge Representation and Reasoning (kr). Cambridge (MA): Morgan Kaufman; 1992. p. 165–176.

[51] Renz J, Nebel B. Spatial reasoning with topological information. In: Freksa C, Habel C, Wender KF, editors. Spatial cognition, an interdisciplinary approach to representing and processing spatial knowledge. Vol. 1404 of Lecture Notes in Computer Science. Berlin: Springer; 1998. p. 351–372.

[52] Freksa C. Qualitative spatial reasoning. Berlin: Springer; 1991.

[53] Frank AU. Qualitative spatial reasoning with cardinal directions. In: Kaindl H, editor. Proceedings of Austrian Conference on Artificial Intelligence. Berlin: Springer; 1991. p. 157–167.

[54] Renz J, Mitra D. Qualitative direction calculi with arbitrary granularity. In: Proceedings of Pacific Rim International Conference on Artificial Intelligence (PRICAI). Auckland, New Zealand: Springer; 2004. Vol. 3157, p. 65–74.

[55] Freksa C. Using orientation information for qualitative spatial reasoning. In: Frank AU, Campari I, Formentini U, editors. International Conference on theories and methods of spatio-temporal reasoning in geographic space. Vol. 1992, Berlin: Springer; p. 162–178. DOI:10.1007/3-540-55966-3_10

[56] Wallgrün JO, Frommberger L, Wolter D, et al. Qualitative spatial representation and reasoning in the SparQ-toolbox. In: Proceedings of Spatial Cognition. Bremen, Germany: Springer; 2006. p. 39–58.

[57] Goyal RK, Egenhofer MJ. Similarity of cardinal directions. In: Jensen CS, Schneider M, Seeger B, Tsotras VJ, editors. Proceedings International Symposium advances in spatial and temporal databases. Redondo Beach, CA: Springer; 2000;36–55.

[58] Cohn AG, Li S, Liu W, et al. Reasoning about topological and cardinal direction relations between 2-dimensional spatial objects. J Artif Intell Res. 2014;493–532.

[59] Moratz R, Ragni M. Qualitative spatial reasoning about relative point position. J Visual Languages Comput. 2008;19(1):75–98. Available from: http://www.sciencedirect.com/science/article/pii/S1045926X06000723

[60] Ligozat G, Renz J. What is a qualitative calculus? A general framework. In: Proceedings of Pacific Rim International Conference on artificial intelligence

(PRICAI). Auckland, New Zealand: Springer; 2004. p. 53–64.

[61] Lücke D, Mossakowski T, Wolter D. Qualitative reasoning about convex relations. In: Freksa C, Newcombe NS, Gärdenfors P, Wölfl S , editors. Proceedings of spatial cognition. Berlin: Springer; 2008. p. 426–440.

[62] Wolter D, Lee J. Qualitative reasoning with directional relations. Artif Intell. 2010;174(18):1498–1507. Available from: http://www.sciencedirect.com/science/article/pii/S0004370210001530

[63] Westphal M, Wölfl S. Confirming the QSR promise. AAAI Spring Symposium: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems. Menlo Park (CA): AAAI; 2009.

[64] Wolter D, Wallgrün JO. Qualitative spatial reasoning for applications: new challenges and the SparQ toolbox. Qualitative Spatio-Temporal Representation Reasoning: Trends Future Directions. 2012;336–362.

[65] Gantner Z, Westphal M, Wölfl S. GQR–a fast reasoner for binary qualitative constraint calculi. AAAI Workshop on Spatial and Temporal Reasoning. Chicago (IL): AAAI; 2008.

[66] Stocker M, Sirin E. PelletSpatial: a hybrid RCC-8 and RDF/OWL reasoning and query engine. In: Proceedings of 6th International Conference on OWL: Experiences and Directions (owled). Aachen, Germany; 2009. p. 39–48. Available from: http://dl.acm.org/citation.cfm?id=2890046.2890051

[67] Christodoulou G, Petrakis EGM, Batsakis S. Qualitative spatial reasoning using topological and directional information in OWL. In: Proceedings of IEEE 24th International Conference on Tools with Artificial Intelligence. Vol. 1, Athens, Greece: IEEE Computer Society; 2012. p. 596–602.

[68] Condotta JF, Saade M, Ligozat G. A generic toolkit for n-ary qualitative temporal and spatial calculi. International Symposium Temporal Representation and Reasoning. Budapest, Hungary: IEEE Computer Society; 2006 Jun. p. 78–86

[69] Wollherr D, Khan S, Landsiedel C, et al. The interactive urban robot IURO: towards robot action in human environments. In: Hsieh AM, Khatib O, Kumar V, editors. Proceedings of International Symposium Experimental Robotics (ISER). Cham, Switzerland: Springer International Publishing; 2016. p. 277–291. DOI:10.1007/978-3-319-23778-7_19

[70] Sjöö K, Pronobis A, Jensfelt P. Functional topological relations for qualitative spatial representation. In: Fitzgibbon AW, Lazebnik S, Perona P, Sato Y, Schmid C, editors. Proceedings of IEEE International Conference on Advanced Robotics. Tallinn, Estonia: IEEE; 2011. p. 130–136.

[71] Silberman N, Hoiem D, Kohli P, et al. Indoor segmentation and support inference from RGBD images. In: Fitzgibbon AW, Lazebnik S, Perona P, Sato Y, Schmid C, editors. Proceedings of European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science. Vol. 7576, Florence, Italy: Springer; 2012. p. 746–760. DOI:10.1007/978-3-642-33715-4_54

[72] Smith RC, Cheeseman P. On the representation and estimation of spatial uncertainty. Int J Rob Res. 1986;5(4):56–68.

[73] Leonard JJ, Durrant-Whyte HF. Simultaneous map building and localization for an autonomous mobile robot. In: Proceedings of IEEE/RSJ International Workshop Intelligence for Mechanical Systems, Intelligent Robots and Systems. Osaka, Japan: IEEE; 1991 Nov. p. 1442–1447.

[74] Grisetti G, Kummerle R, Stachniss C, et al. A tutorial on graph-based SLAM. IEEE Intell Trans Syst Mag. 2010;2(4):31–43.

[75] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters. Trans Rob. 2007;23(1):34–46. DOI:10.1109/TRO.2006.889486.

[76] Triebel R, Pfaff P, Burgard W. Multi-level surface maps for outdoor terrain mapping and loop closing. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Beijing: IEEE; 2006. p. 2276–2282.

[77] Aurenhammer F. Voronoi diagrams - a survey of a fundamental geometric data structure. ACM Comput Surv (CSUR). 1991;23(3):345–405.

[78] Remolina E, Kuipers B. Towards a general theory of topological maps. Artif Intell. 2004;152(1):47–104. Available from: http://www.sciencedirect.com/science/article/pii/S0004370203001140

[79] Ranganathan A, Dellaert F. Online probabilistic topological mapping. Int J Rob Res. 2011;30(6):755–771. Available from: http://ijr.sagepub.com/content/30/6/755.full.pdf+html http://ijr.sagepub.com/content/30/6/755.abstract

[80] Wallgrün JO. Qualitative spatial reasoning for topological map learning. Spatial Cognition Comput. 2010;10(4):207–246.

[81] Boal J, Sánchez-Miralles A, Arranz A. Topological simultaneous localization and mapping: a survey. Robotica. 2014;32:803–821. Available from: http://journals.cambridge.org/article_S0263574713001070

[82] Werner S, Krieg-Brückner B, Herrmann T. Modelling navigational knowledge by route graphs. In: Proceedings of Spatial Cognition. Berlin: Springer; 2000. p. 295–316.

[83] Buschka P, Saffiotti A. Some notes on the use of hybrid maps for mobile robots. In: Proceedings of International Conference on Intelligent Autonomous Systems. Amsterdam: IOS Press; 2004. p. 547–556.

[84] Kuipers B. An intellectual history of the spatial semantic hierarchy. Robotics and cognitive approaches to spatial mapping. Berlin: Springer; 2007. p. 243–264.

[85] Kuipers B, Byun YT. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. Rob Auton Syst. 1991;8(1):47–63.

[86] Kuipers B, Modayil J, Beeson P, et al. Local metrical and global topological maps in the hybrid spatial semantic hierarchy. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Vol. 5, New Orleans (LA): IEEE; 2004. p. 4845–4851.

[87] Beeson P, MacMahon M, Modayil J, et al. Integrating multiple representations of spatial knowledge for mapping, navigation, and communication. Interaction challenges for intelligent assistants. Stanford University, CA: AAAI; 2007. p. 1–9.

[88] Beeson P, Modayil J, Kuipers B. Factoring the mapping problem: mobile robot map-building in the hybrid spatial semantic hierarchy. Int J Rob Res. 2010;29(4):428–459.

[89] Lang D, Friedmann S, H M, et al. Definition of semantic maps for outdoor robotic tasks. In: Proceedings of IEEE International Conference on Robotics and Biomimetics. Bali: IEEE; 2014. p. 2547–2552.

[90] Case C, Suresh B, Coates A, et al. Autonomous sign reading for semantic mapping. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai: IEEE; 2011. p. 3297–3303.

[91] Trevor AJ, Rogers JG, Nieto-Granda C, et al. Feature-based mapping with grounded landmark and place labels. In RSS Workshop on Grounding Human-Robot Dialog for Spatial Tasks. Los Angeles (CA): MIT Press; 2011.

[92] Walter MR, Hemachandra S, Homberg B, et al. A framework for learning semantic maps from grounded natural language descriptions. Int J Rob Res. 2014;33(9):1167–1190.

[93] Kostavelis I, Gasteratos A. Semantic mapping for mobile robotics tasks: a survey. Rob Auton Syst. 2015;66:86–103.

[94] Limketkai B, Liao L, Fox D. Relational object maps for mobile robots. In: Proceedings of International Joint Conference on Artificial Intelligence (IJCAI). Edinburgh: Professional Book Center; 2005. p. 1471–1476.

[95] Nüchter A, Hertzberg J. Towards semantic maps for mobile robots. Rob Auton Syst. 2008;56(11):915–926.

[96] Meger D, Forssén PE, Lai K, et al. Curious george: an attentive semantic robot. Rob Auton Syst. 2008;56(6):503–511.

[97] Viswanathan P, Meger D, Southey T, et al. Automated spatial-semantic modeling with applications to place labeling and informed search. Proceedings of Canadian Conference on Computer and Robot Vision. Kelowna: IEEE Computer Society; 2009. pp. 284–291

[98] Vasudevan S, Gächter S, Nguyen V, et al. Cognitive maps for mobile robots–an object-based approach. Rob Auton Syst. 2007;55(5):359–371.

[99] Vasudevan S, Siegwart R. Bayesian space conceptualization and place classification for semantic maps in mobile robotics. Rob Auton Syst. 2008;56(6):522–537.

[100] Rituerto A, Murillo A, Guerrero J. Semantic labeling for indoor topological mapping using a wearable catadioptric system. Rob Auton Syst. 2014;62(5):685–695. special Issue Semantic Perception, Mapping and Exploration. Available from: http://www.sciencedirect.com/science/article/pii/S0921889012001856

[101] Liu M, Colas F, Pomerleau F, et al. A Markov semi-supervised clustering approach and its application in topological map extraction. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vilamoura, Algarve: IEEE; 2012. p. 4743–4748.

[102] Brunskill E, Kollar T, Roy N. Topological mapping using spectral clustering and classification. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). San Diego (CA): IEEE; 2007. p. 3491–3496

[103] Liu M, Colas F, Siegwart R. Regional topological segmentation based on mutual information graphs. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai: IEEE; 2011. p. 3269–3274.

[104] Liu Z, Chen D, von Wichert G. Online semantic exploration of indoor maps. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). St. Paul (MN): IEEE; 2012. p. 4361–4366. DOI:10.1109/ICRA.2012.6224871

[105] Pronobis A, Jensfelt P. Multi-modal semantic mapping. RSS workshop on grounding human-robot dialog for spatial tasks. Los Angeles (CA): Robotics, Science and Systems; 2011.

[106] Pronobis A, Jensfelt P. Large-scale mapping and reasoning with heterogeneous modalities. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). St. Paul (MN): IEEE; 2012. p. 28.

[107] Mozos OM, Triebel R, Jensfelt P, et al. Supervised semantic labeling of places using information extracted from sensor data. Rob Auton Syst. 2007;55(5):391–402.

[108] Sünderhauf N, Dayoub F, McMahon S, et al. Place categorization and semantic mapping on a mobile robot. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Stockholm: IEEE; 2016 May; p. 5729–5736.

[109] Murphy L, Sibley G. Incremental unsupervised topological place discovery. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Hong Kong: IEEE; 2014 May; p. 1312–1318.

[110] Ranganathan A, Dellaert F. Bayesian surprise and landmark detection. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Kobe: IEEE; 2009 May; p. 2017–2023.

[111] Tapus A, Siegwart R. Incremental robot mapping with fingerprints of places. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Edmonton: IEEE; 2005. p. 2429–2434.

[112] Thrun S, Gutmann JS, Fox D, et al. Integrating topological and metric maps for mobile robot navigation: a statistical approach. In: Proceedings of National Conference on Artificial Intelligence (AAAI). Madison (WI): AAAI; 1998. p. 989–995.

[113] Nieto-Granda C, Rogers JG, Trevor AJ, et al. Semantic map partitioning in indoor environments using regional analysis. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Taipei: IEEE; 2010. p. 1451–1456.

[114] Zender H, Mozos OM, Jensfelt P, et al. Conceptual spatial representations for indoor mobile robots. Rob Auton Syst. 2008;56(6):493–502.

[115] Hawes N, Hanheide M, Hargreaves J, et al. Home alone: autonomous extension and correction of spatial representations. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai: IEEE; 2011. p. 3907–3914.

[116] Galindo C, Saffiotti A, Coradeschi S, et al. Multi-hierarchical semantic maps for mobile robotics. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Edmonton: IEEE; 2005. p. 2278–2283.

[117] Tenorth M, Kunze L, Jain D, et al. Knowrob-map – knowledge-linked semantic object maps. In: Proceedings of IEEE/RAS International Conference on Humanoid Robots. Nashville (TN): IEEE; 2010. p. 430–435.

[118] Pangercic D, Pitzer B, Tenorth M, et al. Semantic object maps for robotic housework–representation, acquisition and use. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vilamoura, Algarve: IEEE; 2012. p. 4644–4651.

[119] Riazuelo L, Tenorth M, Di Marco D, et al. RoboEarth semantic mapping: A cloud enabled knowledge-based approach. IEEE Trans Autom Sci Eng. 2015;12(2):432–443.

[120] Aydemir A, Jensfelt P, Folkesson J. What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vilamoura, Algarve: IEEE; 2012. p. 4675–4682.

[121] Luperto M, D'Emilio L, Amigoni F. A generative spectral model for semantic mapping of buildings. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE; 2015. p. 4451–4458.

[122] Luperto M, Quattrini Li A, Amigoni F. A system for building semantic maps of indoor environments exploiting the concept of building typology. In: Behnke S, Veloso M, Visser A, Xiong R, editors. Proceedings of robocup. Berlin: Springer; 2014. p. 504–515. DOI:10.1007/978-3-662-44468-9_44

[123] Sengupta S, Greveson E, Shahrokni A, et al. Urban 3D semantic modelling using stereo vision. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Karlsruhe: IEEE; 2013. p. 580–585.

[124] Douillard B, Fox D, Ramos F, et al. Classification and semantic mapping of urban environments. Int J Rob Res. 2011;30(1):5–32.

[125] Singh G, Koeck J. Acquiring semantics induced topology in urban environments. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). St. Paul (MN): IEEE; 2012 May; p. 3509–3514.

[126] Bernuy F, Ruiz del Solar J. Semantic mapping of large-scale outdoor scenes for autonomous off-road driving. In: Proceedings of IEEE International Conference on Computer Vision Workshop. Santiago: IEEE; 2015. p. 35–41.

[127] Wolf DF, Sukhatme GS. Semantic mapping using mobile robots. Trans Rob. 2008;24(2):245–258.

[128] Nüchter A, Surmann H, Lingemann K, et al. Semantic scene analysis of scanned 3D indoor environments. In: Proceedings of International Workshop on Vision, Modeling and Visualization (VMV). Munich, Germany: Aka GmbH; 2003 Nov.. p. 215–221.

[129] Pronobis A, Jensfelt P, Sjöö K, et al. Semantic modelling of space. Cognitive systems. Berlin: Springer; 2010. p. 165–221.

[130] Torralba A, Murphy KP, Freeman WT, et al. Context-based vision system for place and object recognition. In: Proceedings of International Conference on Computer Vision (ICCV). Nice, France: IEEE Computer Society; 2003 Oct. p. 273–280.

[131] Kollar T, Roy N. Utilizing object-object and object-scene context when planning to find things. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Kobe, Japan: IEEE; 2009 May. p. 4116–4121.

[132] Spexard T, Li S, Wrede B, et al. BIRON, where are you? Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Beijing, China: IEEE; 2006 Oct. p. 934–940.

[133] Hemachandra S, Kollar T, Roy N, et al. Following and interpreting narrated guided tours. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai, China: IEEE; 2011 May. p. 2574–2579.

[134] Walter MR, Antone M, Chuangsuwanich E, et al. A situationally-aware voice-commandable robotic forklift working alongside people in unstructured outdoor environments. J Field Rob. 2015;32(4):590–628.

[135] Walter MR, Hemachandra S, Homberg B, et al. Learning semantic maps from natural language descriptions. In: Newman P, Fox D, Hsu D, editors. Proceedings of Robotics: Science and systems (RSS). Berlin, Germany: Robotics: Science and Systems; 2013 Jun.

[136] Hemachandra S, Walter MR, Tellex S, et al. Learning spatial-semantic representations from natural language descriptions and scene classifications. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE; 2014 May. p. 2623–2630.

[137] Hemachandra S, Walter MR. Information-theoretic dialog to improve spatial-semantic representations. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany: IEEE; 2015 Oct. p. 5115–5121.

[138] Schuette N, Kelleher J, Namee BM. Robot perception errors and human resolution strategies in situated human-robot dialogue. Adv Rob. 2016;31.

[139] Topp EA. Linking user actions to spatial concepts in human augmented mapping. Adv Rob. 2016;31.

[140] Rangel JC, Martinez J, Garcia-Varea I, et al. Lextomap: lexical-based topological mapping. Adv Rob. 2016;31.

[141] Keizer S, Foster ME, Gaschler A, et al. Handling uncertain input in multi-user human-robot interaction. Proceedings of IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Edinburgh: IEEE; 2014 Aug. p. 312–317.

[142] Bohus D, Saw CW, Horvitz E. Directions robot: in-the-wild experiences and lessons learned. In: Bazzan ALC, Huhns MN, Lomuscio A, Scerri P, editors. Proceedings of International Conference on Autonomous Agents and Multi-agent Systems (AAMAS). Paris: IFAAMAS/ACM; 2014. p. 637–644.

[143] Gkatzia D, Rieser V, Bartie P, et al. From the virtual to the real world: referring to objects in real-world spatial scenes. In: Proceedings of Conference Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics; 2015 Sep. p. 1936–1942. Available from: https://aclweb.org/anthology/D/D15/D15-1224

[144] Cercas Curry A, Gkatzia D, Rieser V. Generating and evaluating landmark-based navigation instructions in virtual environments. In: Proceedings of European Workshop Natural Language Generation (ENLG). Brighton, UK: Association for Computational Linguistics; 2015 Sep. p. 90–94. Available from: http://www.aclweb.org/anthology/W15-4715

[145] Landsiedel C, Wollherr D. Road geometry estimation for urban semantic maps using open data. Adv Rob. 2016;31.