Talbot, Ben, Lam, Obadiah, Schulz, Ruth, Dayoub, Feras, Upcroft, Ben, & Wyeth, Gordon
(2016)
Find my office: Navigating real space from semantic descriptions. In *IEEE International Conference on Robotics and Automation (ICRA 2016)*, 15-21 May 2016, Stockholm, Sweden.

Find my office: Navigating real space from semantic descriptions

Ben Talbot[1], Obadiah Lam[1], Ruth Schulz[1], Feras Dayoub[1], Ben Upcroft[2], and Gordon Wyeth[2]

***Abstract*— This paper shows that by using only symbolic language phrases, a mobile robot can purposefully navigate to specified rooms in previously unexplored environments. The robot intelligently organises a symbolic language description of the unseen environment and "imagines" a representative map, called the abstract map. The abstract map is an internal representation of the topological structure and spatial layout of symbolically defined locations. To perform goal-directed exploration, the abstract map creates a high-level semantic plan to reason about spaces beyond the robot's known world. While completing the plan, the robot uses the metric guidance provided by a spatial layout, and grounded observations of door labels, to efficiently guide its navigation. The system is shown to complete exploration in unexplored spaces by travelling only 13.3% further than the optimal path.**

I. INTRODUCTION

Navigating through the world is a requirement of daily life for humans, and is also vital for mobile robots to be useful within their environments of operation. Furthermore, navigation must be intelligent even in places that a robot has never been before. To achieve this, spatial navigation depends on spatial cues from the environment and self motion, computational mechanisms, and spatial representations [1].

When people navigate built environments, many of the spatial cues that they use to aid in their wayfinding have been left by other people in the form of labels, signs, maps, and planners [2]. One commonly used cue is a set of directions or descriptions that can guide someone through an unfamiliar environment to their goal. People follow directions with ease in most situations, even though language is ambiguous and can be difficult to determine how the directions correspond to the real world.

In contrast to people, robots typically rely on constructing a detailed map of their environment prior to autonomous operation, often requiring manual teleoperation to build a useful geometric map. However, unlike geometric maps, symbolic spatial information is typically devoid of metric meaning. In order to make sense of a set of natural language directions or descriptions, robots need to make sense of the ambiguity in descriptions and changing frames of reference, as well as be able to identify other cues in the environment. Some progress has been made in this area, with a number of robotic systems developed to use symbolic spatial information to guide navigation, including gestures [3], maps [4], and natural language route instructions [5]. These systems are still restricted in the types of symbolic spatial information that they can interpret and use for navigation.

[1] The authors are with the School of Electrical Engineering and Computer Science, Queensland University of Technology (QUT), Brisbane, Australia {b.talbot, o1.lam, ruth.schulz, feras.dayoub}@qut.edu.au

[2] The authors are with the ARC Centre of Excellence for Robotic Vision, Queensland University of Technology (QUT), Brisbane, Australia. http://roboticvision.org/ email: {ben.upcroft, gordon.wyeth}@qut.edu.au

Fig. 1: The GuiaBot navigating towards its goal, guided by a structured language environment description and grounded observations of door labels

This paper extends preliminary work on constructing abstract maps from spatial descriptions that can be used to guide goal-directed exploration [6]. Symbolic language phrases of spatial descriptions are ambiguous, probabilistic, and non-metric. This paper presents two key contributions that allow a robot to efficiently navigate in environments which it has never been to before, guided by symbolic spatial language (see Figure 1):

- a novel method for parametrically defining the topological structure and spatial layout information encoded in spatial language phrases. This allows the robot to decode the spatial meaning communicated through symbolic language phrases.
- a complete symbolic navigation system comprising of large, topologically organised, spring-damper systems. The system leverages both the conciseness of semantic structure, and the metric guidance available from spatial layouts. This allows a robot navigation system to efficiently find symbolic goals in unseen environments.

The remainder of this paper is organised as follows. After reviewing the related literature, section 3 explains the approach behind a symbolic navigation system that represents, interprets, and reasons about symbolic language phrases. In section 4 we present our experiment configuration, with the results of simulated and real world experiments presented in section 5. The paper concludes with a discussion of demonstrated outcomes.

II. Related work

People can use natural language to convey information about an environment either through a set of directions leading from one location to another or through a survey or route description of the environment. Research into robot navigation systems using natural language as a resource has made improvements on algorithms for understanding route directions provided in natural language [7], [8], [9], created semantic maps from natural language descriptions [10], and allowed robots to follow route directions [5], [11], [12], [13], [14]. While most existing models require prior knowledge of the world to be able to follow the directions or a method to follow directions when the world is unknown, some use the language to build hypothesized world models [5] or add unvisited locations to maps [9].

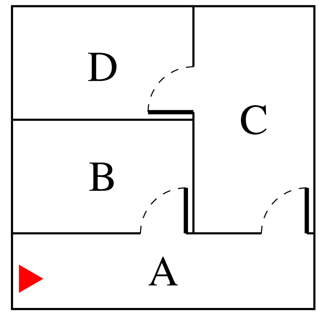
While route directions describe a single path through an environment, a survey or route description, comprised of natural language phrases of spatial information, allows others to form fairly complete mental maps of the environment [15]. Each phrase typically contains or refers to four components: three toponyms (a noun denoting the name of a place) and a spatial relationship. The three toponyms are the figure location, which is the space being described, the reference location, which provides context for understanding the phrase, and the context location, which may also define the frame of reference [16] for interpreting the phrase. The spatial relationship is described with a spatial preposition [17]. A robot with access to many sets of route descriptions or a complete description of the environment would be able to form a more complete mental map of the environment.

III. Approach

This paper describes a symbolic navigation system that purposefully finds a goal in an unfamiliar built environment using only structured language phrases and information available within the environment. The targeted built environments are complex office floors with multiple nested corridors and adjoining rooms (a simple example is shown in Figure 2a).

In this system, the provided semantic knowledge is a structured language description of the spatial configuration of locations in the environment. Each individual structured language phrase (Figure 2b) is interpreted as a combination of both topological and spatial properties. From this, an abstract map is constructed that contains two levels of spatial information: a high-level topological representation, and a series of low-level spatial representations (Figure 3). This internal representation provides a series of symbolic navigation actions (e.g. *"find and enter room A"*) and estimated spatial positions (e.g. an estimated position for the entry to Room A) for efficiently finding semantic goals in unseen built environments.

As the robot attempts to navigate to a semantic goal's predicted location, door label observations provide the abstract map with grounded symbolic information about the state of



(a)

i. *"You are in A"*
ii. *"B is down A"*
iii. *"C is down A"*
iv. *"C is beyond B"*
v. *"D is through C"*

(b)

Fig. 2: Sample Environment: a) floor plan and b) structured language phrases describing the environment

the world. This information is used to adapt and improve the robot's internal semantic and spatial understanding of the surrounding environment. Currently, this work assumes that the toponyms have a unique referent in the real world. For example, in most cases it is safe to assume the room label *GP-S-1101* only refers to one place in the robot's operational space. On the other hand, the toponym *hallway* could possibly refer to a number of locations.

A. Representing natural language phrases

A piece of symbolic spatial information is defined as a quadruple, with each element encapsulating one of the four natural language components discussed in Section II. Figure 4 shows a visual representation of the spatial components in a structured language phrase. In this example, the figure location p_f is described as *"beyond"* the reference location p_r. This spatial relationship is represented by the shaded area, which is defined by the frame of reference provided by the context location p_c.

Simplifying assumptions are made about the nature of the elements in the symbolic spatial information quadruple. In this work, the relationship is always assumed to be a spatial preposition and the figure, reference object, and context are all assumed to be toponyms. Each toponym either refers to the location of a door for a labelled space, or an arbitrary point of reference within a space. For example, p_c could be a robot's current position in a hallway whereas p_r could be the entry to a person's office.

In Figure 2b, five structured language phrases (hereafter referred to as phrases *i* to *v*) are shown which describe the layout of the environment in Figure 2a in relation to the robot's starting position (referred to as *"here"*). Each of the phrases omits an explicit specification of the context toponym. The system assumes that the context location is the robot's current point in space unless the language explicitly specifies otherwise.

B. Interpreting natural language phrases

Prepositions are the component used in structured spatial language to describe the spatial relationship between spaces. This relationship can either be a piece of information about

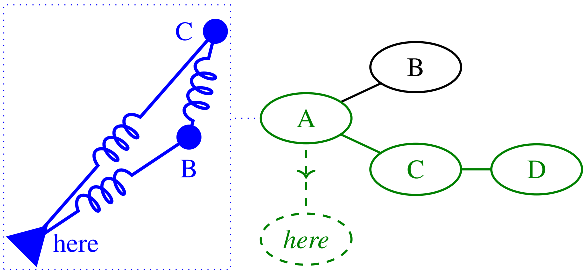


Fig. 3: Sample Environment: Topological (right) and spatial (left) properties inferred by the language phrases

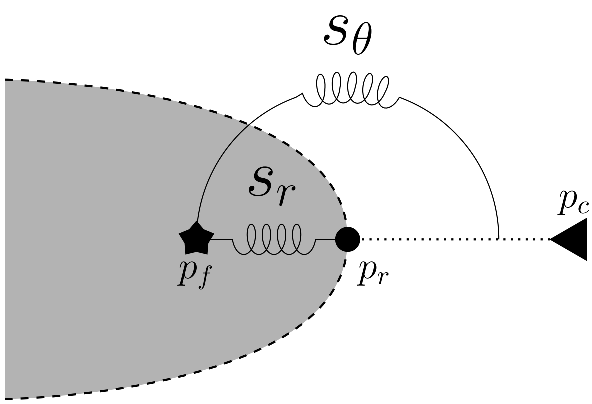


Fig. 4: The figure (place p_f) is described as *"beyond"* the reference place (p_r) from the perspective of p_c (context). Radial (s_r) and angular (s_θ) springs approximate this implied constraint.

Relation	**HL**	**CD**	**Spatial Layout**						
			$\boldsymbol{p_f}$	$\boldsymbol{p_r}$	$\boldsymbol{p_c}$	$\boldsymbol{X_r}$	$\boldsymbol{K_r}$	$\boldsymbol{X_\theta}$	$\boldsymbol{K_\theta}$
beyond	-	$f \leftrightarrow c_p$	f	r	c	1	0.1	π	1
		$r \leftrightarrow c_p$	r	f	c	1	0.1	0	1
in	$f \in r$	-	-	-	-	-	-	-	-
through	-	$f \leftrightarrow r$	-	-	-	-	-	-	-

TABLE I: Interpretation parameters for three prepositions used in the model, defining the three properties: 1) hierarchical link (HL), 2) connectivity declaration (CD), and 3) spatial layout. f, r, c, and c_p refer to figure, reference, context, and context's parent space respectively.

the topological structure of an environment, or a description of the spatial layout. In the example phrases in Figure 2b, both topological structure (right side of Figure 3) and spatial layout (left side of Figure 3) are described.

Phrases *i* and *v* explicitly state information about the environment's topological structure. Specifically, that the robot's current position (*"here"*) is hierarchically included in space A, and there is a navigable connection between space C and D. Similarly, phrases *ii* - *iv* implicitly infer that navigable connections exists between space A, and spaces B and C. These two types of topological relationships between spaces, navigable connectivity and hierarchical membership, are show in Figure 3 as solid and directed dashed lines respectively.

Phrases *ii* - *iv* use prepositions that also explicitly describe the spatial layout of the environment. As the spatial location being described by these prepositions (e.g. the shaded area in Figure 4) lacks a rigid metric definition, it is difficult to parametrically define this location. For example, drawing a cross somewhere to the right of a point is easy, but highlighting the region where crosses can be drawn is a much more challenging exercise. Figure 4 shows an approximation of the region referred to by preposition *"beyond"* when describing the position of p_f, relative to p_r, from the perspective of p_c. The spatial relationship *"beyond"* could be describing a position anywhere inside this indefinitely expanding area.

The area described by a preposition is approximated as a value range along each dimension of the polar coordinate system defined by the three toponyms. This range represents the elasticity implied by the spatial preposition. In Figure 4, for example, the restriction in the radial direction is completely elastic but rigid around $\pm\pi$ in the angular direction. This restriction on value and range is analogous to the restriction imposed by the natural length (X) and stiffness coefficient (K) of a mechanical spring. Consequently, a parametrised spring is used along each of the polar dimensions (s_r and s_θ for r and θ respectively) to represent the spatial layout information inferred by each preposition. A sole layout constraint is shown in Figure 4. In most cases a single preposition can infer more than one of these constraints.

Combining this parametrisation with the topological properties, a spatial preposition is therefore interpreted as the three properties: 1) hierarchical links, 2) declarations of connectivity, and 3) a series of spring-based layout constraints. Each layout constraint is defined by four parameters X_r, K_r, X_θ, and K_θ. An interpretation only requires at least one of the three properties to be defined for the given spatial preposition. Example interpretations are shown in Table I. Currently, these parameters are intuitively selected, but future work could investigate learning them from human data.

C. Semantic reasoning about natural language phrases

A sequence of semantic navigation actions that the robot should carry out in order to complete its symbolic navigation goal are generated by planning through the graph created by the hierarchical and connectivity information. The right side of Figure 3 shows a visual representation of generating a semantic plan from the graph. The system is at the toponym $here$ and trying to get to space D. The backward search finds the path highlighted in green, with the hierarchical link between A and $here$ completing the semantic plan. In this example, the plan consists of two semantic navigation actions: 1) finding and going through the door to C, and 2) finding and going through the door to D.

The set of semantic actions is generated by backward searching from the goal across the connectivity links, while treating hierarchical links as direction-constrained synonyms,

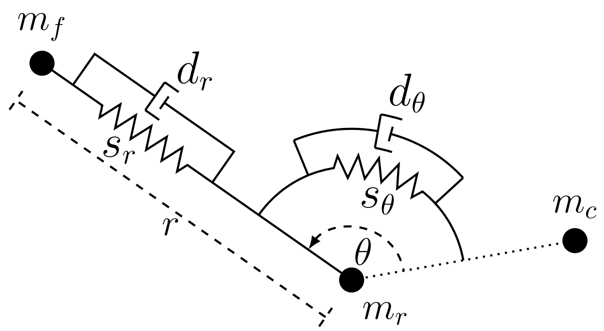


Fig. 5: Mechanical spring-damper system for a spatial constraint

until a path to the robot's current semantic location is found. The direction in which the hierarchy is used to produce spatial synonyms is important. For example, it is correct to infer that being $here$ means the robot is in room A, but incorrect to infer that being in room A means the robot is $here$.

D. Spatial reasoning about natural language phrases

While completing each semantic action, the abstract map uses the spatial constraints on the layout of the current space to guide its metric navigation. The left side of Figure 3 shows the constraints on spatial layout in space A from the natural language phrases in Figure 2b. Phrase *ii* constrains the location of B relative to $here$, phrase *iii* constrains the location of C relative to $here$, and phrase *iv* constrains the location of C relative to B. The location of $here$ is known, but the locations of B and C must settle at a low energy state that best represents the information provided by those constraints.

To reason about multiple spatial constraints, a toponym is treated as a mass in a point-mass system. Furthermore, each constraint is modelled as a force applied to an unfixed mass m_f. The force vector applied by the constraint is determined by the constraint's parameters, and the placement of masses m_r and m_c. At current, there is no preference to specific spaces or locations so every mass in the system is given a unit mass of 1kg.

For each spatial constraint, a mechanical spring-damper subsystem is introduced along each of the two dimensions of the local polar coordinate frame (see Figure 5). The polar axis is defined as the vector from m_r to m_c, with a spring-damper along both the radial and angular axes. The dampers are introduced along each axis to gradually remove energy from the system, allowing it to settle into a minimal-energy state representing the estimated positions of the masses.

The radial spring-damper, consisting of spring s_r and damper s_θ, control the distance r to m_f. Likewise, an angular spring-damper (s_θ and d_θ) controls the azimuth θ to m_f. By approximating the mechanical system in Figure 5 as two independent subsystems, the damping coefficients are set system wide to provide a slightly underdamped frequency response characteristic. This approximation effectively ignores the effect of the centripetal component in Equation 1. Consequently, the radial and angular accelerations acting on m_f within a constraint are defined as:

$$\ddot{r} = \underbrace{-\dot{\theta}^2 r}_{\substack{\text{centripetal}\\ \text{component}}} \underbrace{-K_r(r - X_r) - D_r\dot{r}}_{\text{radial component}} \tag{1}$$

$$\ddot{\theta} = \underbrace{-K_\theta(\theta - X_\theta) - D_\theta\dot{\theta}}_{\text{angular component}} \tag{2}$$

The effects of these accelerations on the point-mass system helps to refine the abstract map's estimated position for the target of the current semantic action. For example in Figure 6a, the navigation system's current semantic action is directing it to *"find C"*. The point mass system for space A (overlayed in blue in Figure 6a) uses its metric estimate for C to guide the robot in completing this action.

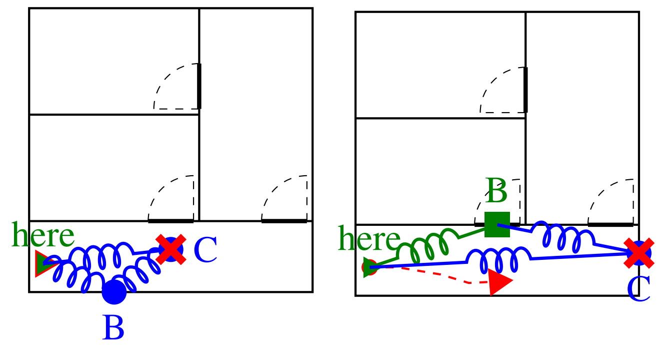


Fig. 6: Sample Environment: a) System's initial starting state without grounded symbolic spatial information and b) after grounding B then adjusting the scaling factor

E. Incorporating grounded symbolic spatial information

The spatial layouts in the abstract map function as a culmination of symbolic spatial information from both structured language and grounded robot observations. The grounded symbolic spatial information that the robot perceives through its observations must be incorporated back into the abstract map to replace and refine its internal representations of the world. By introducing two scaling factors that incorporate grounded spatial information, the system's internal representation can be adapted to provide a more accurate reflection of the robot's environment.

Navigation doesn't always go to plan, and when a goal is not found where expected, the natural human response is to expand the scope of their search. The abstract map replicates this behaviour in each spatial layout by adjusting an exploration scaling factor E_{sc} when a symbolic goal is not found at its estimated position. The factor is incremented by an exploration step Δ_E until a new piece of symbolic spatial information is perceived. At this point, E_{sc} is returned to its resting value of 1, because the system has new information which it can use to re-evaluate its predictions.

For example, the robot in Figure 6a would travel to the location where C is in the spatial layout without receiving

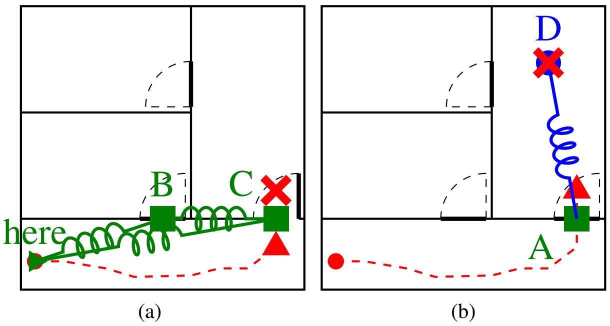


Fig. 7: Sample Environment: a) B and C are observed and grounded, and b) the spatial layout for C is loaded, with the door entered treated as the entrance to A

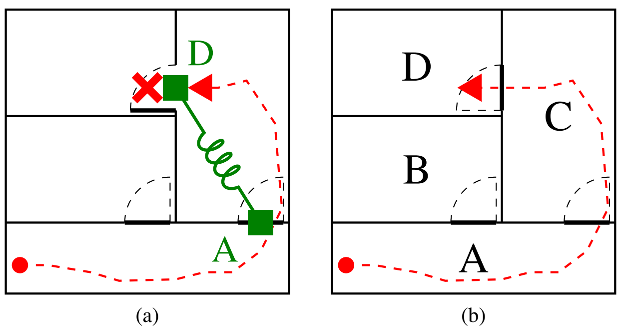


Fig. 8: Sample Environment: a) D is observed and the goal is set to inside the room, and b) the goal is completed

any symbolic spatial information. Without any new information to base its internal representation off, the system is stuck with no course of action. As the exploration scaling factor E_{sc} is increased, the blue spatial layout expands pushing the metric estimate for C further along the corridor. When the robot observes the door label for space B (as shown in Figure 6b), the exploration scaling factor is removed, and replaced with the scaling factor observed between grounded toponyms. This observation scaling factor O_{sc} is defined as

$$O_{sc} = \frac{\sum_{i=1}^{n} K_{r_i} \frac{O_{r_i}}{X_{r_i}}}{\sum_{i=1}^{n} K_{r_i}}, \tag{3}$$

where O_{r_i} is the observed distance between the two spaces referred to in constraint i. This scaling factor is the arithmetic mean of the observed scale shift in natural radial length, weighted by the radial stiffness of each constraint. In practical terms, the observation scaling factor adapts the natural length of the radial springs so that they match the scale observed in grounded symbolic observations.

The two scaling factors, E_{sc} and O_{sc}, are applied to the radial component of each constraint, redefining Equation 1 as:

$$\ddot{r} = -\dot{\theta}^2 r - K_r(r - X_r \times O_{sc} \times E_{sc}) - D_r \dot{r} \tag{4}$$

IV. Experimental Design

A. Environments

The navigation system was tested in simulated environments and a real-world office to evaluate the system efficacy. Simulated experiments were performed in simulated environments generated from floor plans of level 8 and 10 of S Block at QUT Gardens Point campus. The real-world evaluation of the system was performed in level 10 of S Block (Figure 9).

B. Robot platform

In the simulated environment, a simulated robot, created through ROS (Robot Operating System), built a map of its environment with a simulated range sensor. Simulated doors, capable of responding to open and close requests, were placed at the entrance to every room and were perceived with a simulated door label detector.

In the real world environment, these experiments were performed on a GuiaBot from MobileRobots (shown in Figure 1) running ROS, with navigation occupancy grids generated from a laser rangefinder. Door label observations were obtained from a synthetically trained end-to-end text reading pipeline developed for reading navigation cues found in office environments that extends work in [18]. The pipeline uses a small CNN for character recognition and text detection, with no dictionary-based error correction stage, allowing any combination of characters to be detected.

C. Experiment

The symbolic navigation system was tested in ten trials, across two simulated environments, and confirmed by three trials in the real-world environment. The robot was provided with no prior map in either the simulated or real environments - it built a map in real-time as it traversed the area.

From a starting position, the robot was tasked with navigating to a target room using only the provided structured language phrases, and any door label observations it received. The same collection of structured language phrases (like in Figure 2b) were provided for the trials in each environment, with the number of phrases ranging from 58 to 82. While the number of phrases was significant, this was to provide a

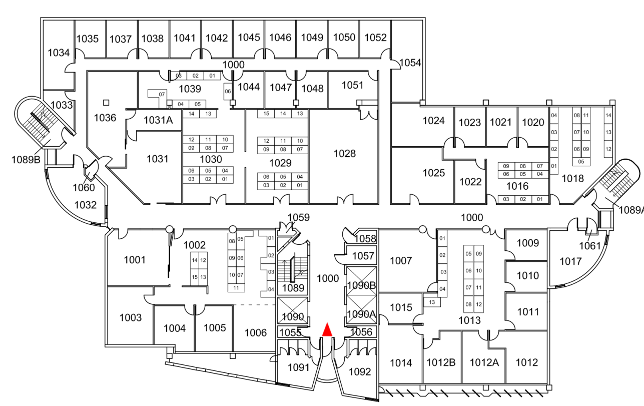


Fig. 9: S Block, Level 10. This environment was used in simulation and real-world validation

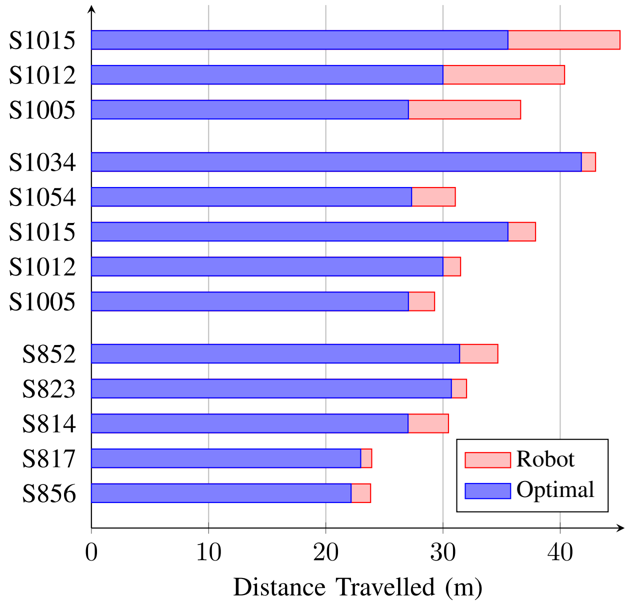


Fig. 10: Comparison of path lengths from S10 real (top), S10 simulated (middle), and S08 simulated (bottom)

description of the location of every room on an entire floor (which would not be necessary for a typical navigation task). This allowed trials of the system to be run for a number of different semantic goals, with the same initial conditions. On completion of the symbolic navigation task, the total distance travelled by the robot in the current trial was recorded.

V. Results

The results demonstrated the abstract map's ability to efficiently complete symbolic goal-directed exploration tasks in unseen environments, with only the aid of structured symbolic phrases and grounded door label observations. On average, the system travelled only 13.3% further than the minimum distance path that was produced by a system with a complete a priori map. In every trial performed, in both the simulated and real-world environments, the system took the most direct route possible to the symbolic destination. The simulated results were between 2% and 13% longer than the optimal path, with the real results ranging from 26% to 35% longer. This difference between simulated and real results is a result of the inefficiency of the robot in executing the planned path in the real environment.

VI. Conclusions

The symbolic navigation system presented in this paper provides an architecture for performing efficient symbolic goal-directed exploration when provided with a structured language description of the environment. The studies presented illustrate the effectiveness of the abstract map in directing a metric navigation system for the completion of goal-directed tasks. The system uses both structured language phrases and grounded symbolic spatial information to reach symbolic goals in unseen environments. In simulation the system completed the exploration tasks by travelling on average only 7.5% further than the optimal path, with real-world results confirming the usefulness of the system.

VII. Acknowledgements

This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP140103216).

References

[1] T. Wolbers and M. Hegarty, "What determines our navigational abilities?" *Trends in cognitive sciences*, vol. 14, no. 3, pp. 138–146, 2010.

[2] J. M. Wiener, S. J. Büchner, and C. Hölscher, "Taxonomy of human wayfinding tasks: A knowledge-based approach," *Spatial Cognition & Computation*, vol. 9, no. 2, pp. 152–165, 2009.

[3] A. Bauer, K. Klasing, G. Lidoris, Q. Mhlbauer, F. Rohrmller, S. Sosnowski, T. Xu, K. Khnlenz, D. Wollherr, and M. Buss, "The autonomous city explorer: Towards natural human-robot interaction in urban environments," *International Journal of Social Robotics*, vol. 1, no. 2, pp. 127–140, 2009.

[4] R. Schulz, B. Talbot, O. Lam, F. Dayoub, P. Corke, B. Upcroft, and G. Wyeth, "Robot navigation using human cues: A robot navigation system for symbolic goal-directed exploration," in *IEEE International Conference on Robotics and Automation 2015*. Washington State Convention Center, Seattle, WA: IEEE, May 2015.

[5] S. Hemachandra, F. Duvallet, T. M. Howard, N. Roy, A. Stentz, and M. R. Walter, "Learning models for following natural language directions in unknown environments," in *IEEE International Conference on Robotics and Automation 2015*. Washington State Convention Center, Seattle, WA: IEEE, May 2015.

[6] R. Schulz, B. Talbot, B. Upcroft, and G. Wyeth, "Constructing abstract maps from spatial descriptions for goal-directed exploration," in *Workshop on Model Learning for Human-Robot Communication at Robotics: Science and Systems 2015*, 2015.

[7] M. MacMahon, B. Stankiewicz, and B. Kuipers, "Walk the talk: Connecting language, knowledge, and action in route instructions," in *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, ser. AAAI'06. AAAI Press, 2006, pp. 1475–1482.

[8] S. Tellex, T. Kollar, S. Dickerson, M. Walter, A. Banerjee, S. Teller, and N. Roy, "Understanding natural language commands for robotic navigation and mobile manipulation," in *AAAI Conference on Artificial Intelligence*, 2011.

[9] T. Williams, R. Cantrell, G. Briggs, P. Schermerhorn, and M. Scheutz, "Grounding natural language references to unvisited and hypothetical locations," in *AAAI Conference on Artificial Intelligence*, 2013.

[10] M. R. Walter, S. Hemachandra, B. Homberg, S. Tellex, and S. Teller, "Learning semantic maps from natural language descriptions," in *Proceedings of Robotics: Science and Systems*, Berlin, Germany, 2013.

[11] C. Matuszek, E. Herbst, L. Zettlemoyer, and D. Fox, "Learning to parse natural language commands to a robot control system," in *Proc. Intl. Symp. on Experimental Robotics (ISER)*, 2012.

[12] F. Duvallet, M. R. Walter, T. Howard, S. Hemachandra, J. Oh, S. Teller, N. Roy, and A. Stentz, "Inferring maps and behaviors from natural language instructions," in *Proc. Intl. Symp. on Experimental Robotics (ISER)*, 2014.

[13] J. Fasola and M. J. Mataric, "Using semantic fields to model dynamic spatial relations in a robot architecture for natural language instruction of service robots," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2013, pp. 143–150.

[14] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward understanding natural language directions," in *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 259–266.

[15] H. A. Taylor and B. Tversky, "Spatial mental models derived from survey and route descriptions," *Journal of Memory and language*, vol. 31, no. 2, pp. 261–292, 1992.

[16] S. C. Levinson, "Frames of reference and molyneux's question: Crosslinguistic evidence," *Language and space*, pp. 109–169, 1996.

[17] B. Landau and R. Jackendoff, "What and where in spatial language and spatial cognition," *Behavioral and Brain Sciences*, vol. 16, pp. 217–238, 6 1993.

[18] O. Lam, F. Dayoub, R. Schulz, and P. Corke, "Text recognition approaches for indoor robotics : a comparison," in *2014 Australasian Conference on Robotics and Automation*, 2014.