TITLE: **Spatial Language for Human-Robot Dialogs**

AUTHORS: **Marjorie Skubic[1]** **(Corresponding Author)**
Dennis Perzanowski[2]
Samuel Blisard[3]
Alan Schultz[2]
William Adams[2]
Magda Bugajska[2]
Derek Brock[2]

[1] Electrical and Computer Engineering Department
349 Engineering Building West
University of Missouri-Columbia, Columbia, MO 65211
skubicm@missouri.edu
Phone: 573-882-7766
Fax: 573-882-0397

[2] Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory, Washington, DC 20375-5337
<dennisp | schultz | adams | bugajska >@aic.nrl.navy.mil / brock@itd.nrl.navy.mil

[3] Computer Science Department
201 Engineering Building West
University of Missouri-Columbia, Columbia, MO 65211
snbfg8@mizzou.edu

Spatial Language for Human-Robot Dialogs

Abstract

In conversation, people often use spatial relationships to describe their environment, e.g., “There is a desk in front of me and a doorway behind it”, and to issue directives, e.g., “Go around the desk and through the doorway.” In our research, we have been investigating the use of spatial relationships to establish a natural communication mechanism between people and robots, in particular, for novice users. In this paper, the work on robot spatial relationships is combined with a multi-modal robot interface. We show how linguistic spatial descriptions and other spatial information can be extracted from an evidence grid map and how this information can be used in a natural, human-robot dialog. Examples using spatial language are included for both robot-to-human feedback and also human-to-robot commands. We also discuss some linguistic consequences in the semantic representations of spatial and locative information based on this work.

Index terms – histogram of forces, human-robot interaction, locatives, multimodal interface, spatial relations

Manuscript received July 15, 2002. This research has been supported by ONR and DARPA.

I. Introduction

In conversation, people often use spatial relationships to describe their environment, e.g., "There is a desk in front of me and a doorway behind it", and to issue directives, e.g., "Go around the desk and through the doorway". Cognitive models suggest that people use these types of relative spatial concepts to perform day-to-day navigation tasks and other spatial reasoning [1], which in part explains the importance of spatial language and how it developed. In our research, we have been investigating the use of spatial relationships to establish a natural communication mechanism between people and robots, in particular, striving for an intuitive interface that will be easy and natural for novice users.

There have been considerable research efforts to study the linguistics of spatial language, e.g., [2,3,4,5,6]. One motivation of this research is the assumption that the cognitive processes humans use to structure language are the same processes used to structure non-linguistic information. In this respect, language provides a window to our cognition. Talmy has discussed the schematic nature of spatial language, i.e., a linguistic description contains only certain characteristics of a scene and discards the rest [2]. Landau and Jackendoff's analysis of spatial language concludes that the cognitive location representation of objects is coarser than the recognition representation [3]. Regier and Carlson assert that the linguistic organization of space provides an interface between language and our perception of the world [4]. We also assert that the ability to use spatial language illustrates a fundamental understanding and reasoning capability. Spatial reasoning is essential for both humans and mobile robots situated in unstructured environments. Our premise is that giving robots the ability to use human-like spatial language will provide an intuitive interface for human users that is consistent with their innate spatial cognition.

In this paper, robot spatial language [7] is combined with a multi-modal robot interface developed at the Naval Research Laboratory (NRL) [8,9]. In [7], spatial language is generated from a static snapshot of sonar sensor readings. Here, we describe a richer set of spatial terminology and extract spatial information from an evidence grid map, which is built from range sensor data accumulated over time [10]. To overcome the object recognition problem, a class of persistent objects has been created, in which objects are given locations in the map (based on sensor readings) and are assigned labels provided by a user. The robot spatial reasoning and the NRL Natural Language Processing system provide the capability of natural human-robot dialogs using spatial language. For example, a user may ask the robot, "How many objects do you see?" The robot responds, "I am sensing 5 objects." The user continues, "Where are they?" The robot responds, "There are objects behind me and on my left." We consider both detailed and coarse linguistic spatial descriptions, and we also support queries based on spatial language, such as "Where is the nearest object on your right?" In addition, spatial language can be used in robot commands, such as "Go to the nearest object on your right." Finally, we consider unoccupied space that is referenced using spatial terms, to support commands such as "Go to the right of the object in front of you."

In each example above, there is a spatial relational comparison between an object or region and some reference point. We will adopt Langacker's term "trajector" [5] to refer to the first object or region and will use the term "referent" to indicate the reference point. Note that the referent is comparable to Langacker's "landmark"[1] [5]. In the examples above, the trajector is often an environment obstacle and the referent is the robot (e.g., "Where is the nearest object on your right?"). However, the trajector may also be an unoccupied region and the referent an environment object (e.g., "go to the right of the object…").

In our use of spatial relations, we will assume an extrinsic reference frame that is based on the robot's viewing perspective [11]. We have not yet explored objects with an intrinsic reference frame, i.e., an inherent front or rear that is defined by the object itself. In this paper, we will use only generic objects or named objects that do not have an intrinsic front or rear.

The paper is organized as follows. Section II provides a discussion of related work. Section III provides an overview of the system and multimodal interface, and Section IV discusses the semantic representation of our spatial language. In Section V, we briefly review algorithms used to process the grid

[1] We prefer to use our own terminology of "referent" as the term "landmark" has other connotations for mobile robots.

map and generate multi-level spatial language. Section VI provides an example of how the spatial language is used in an interactive dialog and includes a discussion of the results and possible evaluation strategies. We conclude in Section VII.

II. Related Work

Although there has been considerable research on the linguistics of spatial language for humans, there has been only limited work done in using spatial language for interacting with robots. Some researchers have proposed a framework for such an interface. For example, Muller et al. [12] describe a control strategy for directing a semi-autonomous wheelchair along a specified route (e.g., in a hospital). The commands take the form of a sequence of qualitative route descriptions, such as “turn left”, “enter right door”, or “follow corridor.” Gribble et al [13] also describe a semi-autonomous wheelchair that users Kuiper’s Spatial Semantic Hierarchy (SSH) [14] to represent and reason about space. The SSH consists of five levels – metrical, topological, causal, control, and sensorimotor. The user interface is discussed for 3 levels – topological (e.g., “go there”), causal (e.g., “go straight” or “turn right”) and control (e.g., “stop”). In this work, the authors set the stage for using spatial language but stop short of illustrating it.

Zelek proposed a lexicon template for incorporating robot commands using spatial references [15]. The template is applied to 2-dimensional robot navigation. Commands are given in the form of a verb, destination, direction, and speed, where destination could be a region relative to a reference object. Here, reference objects were walls and doors identified using 2 laser range-finders, each mounted on a pan-tilt head. Robot navigation was accomplished using a potential field technique; the goal region was given a low potential value and the robot stopped when it approached the edge of the region.

Stopp et al. [16] proposed a two-arm mobile robot designed for assembly tasks. Relative spatial references (e.g., front, right) are used to identify an object in the robot’s geometric world model (i.e., not directly from sensor readings). The user selects an object from the model using a relational expression such as “the spacer on the left.” Elementary spatial relations are computed using idealizations such as center of gravity and bounding rectangle to approximate an object. Spatial relations are modeled using a potential field representation [17].

Moratz et al. [18] investigated the spatial references used by human users to control a mobile robot. They conducted an experiment in which each test subject was asked to direct the robot to a specified location in a field of goal objects situated between the human and the robot. Participants faced the robot and controlled its actions by using natural language sentences typed into a computer. The spatial referencing system, using vision information, was fairly simple, as the goal objects were small blocks and could be modeled as idealized points. Results showed that about half of the subjects directed the robot using a goal object reference. The other half decomposed the control actions into simpler path segments such as “drive a bit forward”, and “come ahead to the right.” The authors hypothesize that the test subjects may have assumed that the robot did not have the capability to understand goal object references. They also apparently found the route decomposition to be a natural interface strategy. An interesting finding is that the test subjects consistently used the robot’s perspective when issuing directives, in spite of the 180-degree rotation. At first, this may seem inconsistent with human to human communication. However, in human to human experiments, Tversky et al. observed a similar result and found that speakers took the listener’s perspective in tasks where the listener had a significantly higher cognitive load than the speaker [19].

Our spatial language dialog with the robot is set in the context of a multimodal interface. From the beginnings of Bolt’s “Put That There” system [20], multimodal interfaces have evolved tremendously and now may incorporate natural language, gesturing and dialog management in addition to the WIMP interface (windows, icons, menus, pointers). Previous gestural interfaces have used stylized gestures of arm and hand configurations [21] or gestural strokes on a PDA display [22]. Other interactive systems, such as [23,24], can process information about the dialog. Our multimodal interface incorporates all of these modalities with some limitations. Our multimodal robot interface is unique in its combination of natural language understanding coupled with the capability of generating and understanding linguistic terms using spatial relations.

The use of linguistic spatial terms in this context requires a computational model for capturing the qualitative character of spatial relations; several models have been proposed, e.g., [25,4,26,27,28,17]. In the work presented here, we use the histogram of forces, developed by Matsakis [29], to model spatial relations, and a system of fuzzy rules to fuse histogram features and generate linguistic terminology [30]. Although previously used for analyzing images, we have adapted the methodology for use on robot range sensor data. The linguistic output of the force histogram rules has not been compared to human responses in a rigorous manner, but informal studies on images have shown close agreement [31]. The information captured by the force histograms is similar to the Attention Vector Sum (AVS) technique proposed by Regier, which has been found to correlate well with human responses [4]. The AVS method sums a weighted set of vectors from each point in the referent to the trajector (considered to be a point). The force histogram method is more general in that it considers a set of vectors from each point in the referent to each point in the trajector and supports any shape or size of trajector or referent.

III. System Overview

In this section, we describe the multimodal interface that provides the context for the human-robot dialog. Fig. 1 shows a schematic overview. Robots used include a Nomad 200, ATRV-Jr, and B21r[2].

A key research goal is to promote natural human-robot interaction (HRI) and provide the user with a rich selection of modalities. For example, humans can issue spoken commands and gesture to direct a robot. Fig. 1 also shows the PDA and touch screen interface called the End User Terminal (EUT). A map representation is available on both the PDA and the EUT screens; the EUT also includes a textual history of the dialog, menu buttons, and a satellite view for outdoor scenarios. The PDA and EUT provide WIMP-type interactions for situations where natural language and gesture may not be appropriate, e.g., due to distance or environmental conditions.

[2] While the Nomad 200 robot is no longer available, both the ATRV-Jr. and the B21r robots are commercially available from iRobot. See http://www.irobot.com.

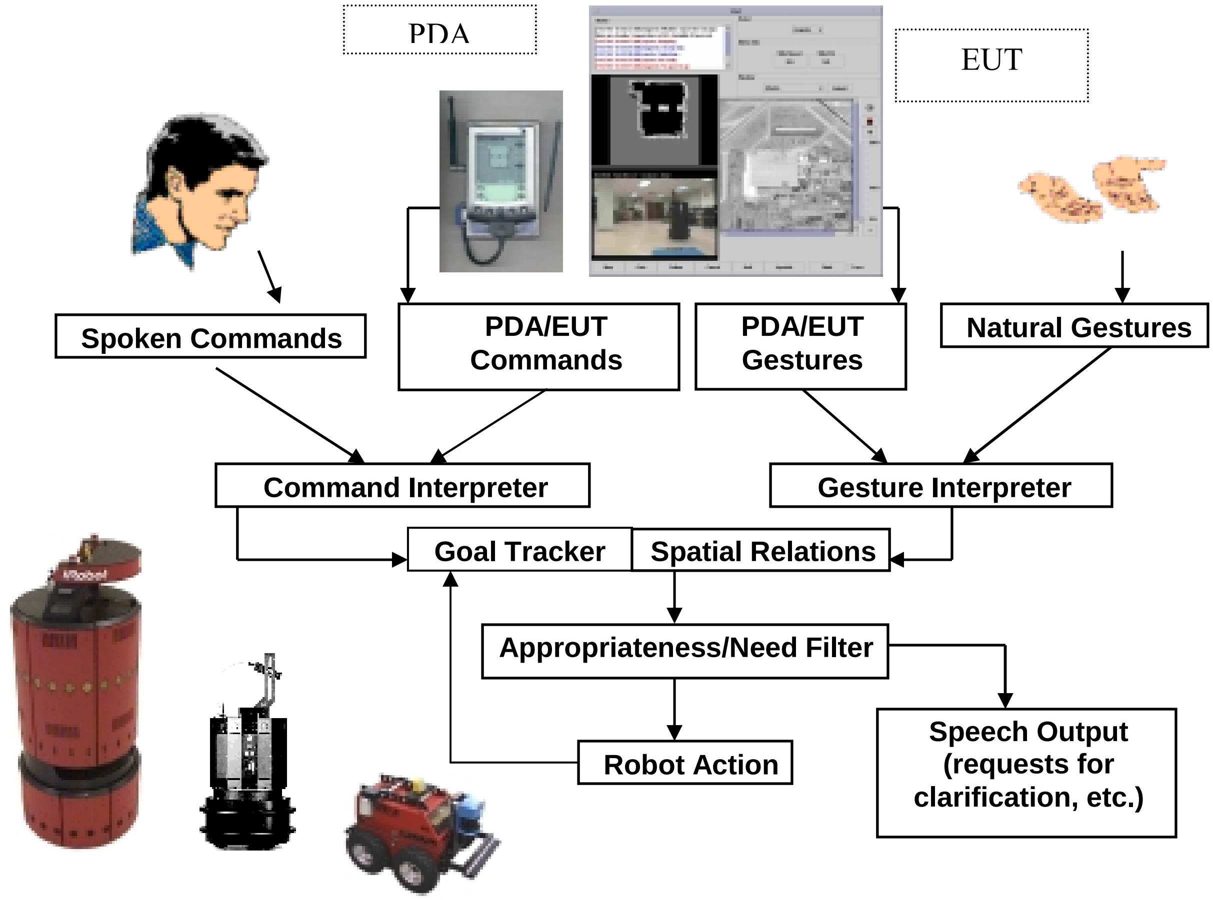
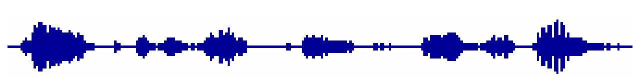


Fig. 1. Schematic overview of the multimodal interface.

Given the components in Fig. 1, we will discuss how the inputs are processed and show how natural language and gestures are combined to produce either a Robot Action or Speech Output. An example mapping of a spoken utterance to the corresponding robot command is shown below in (1) through (6). "Coyote" is the name of a robot.

 (1)

"COYOTE GO OVER THERE" (2)

((ADDRESS (NAME N4 (:CLASS SYSTEM) COYOTE) (3)

```
(IMPER #:V7756 (:CLASS GESTURE-GO)
(:AGENT (PRON N6 (:CLASS SYSTEM) YOU))
(:GOAL (NAME N5 (:CLASS THERE) THERE)))))
```

```
(COMMAND (FORALL X7 (SETOF N4 SYSTEM)                    (4)
              (EXISTS! X8 (SETOF N5 THERE)
                   (GESTURE-GO :AGENT X7 :GOAL X8))))
```

```
(2 45 4590231892.67254106)                               (5)
```

```
"4 45 obstacle"                                          (6)
```

Spoken commands and PDA/EUT generated commands are sent to the Command Interpreter which includes voice recognition and natural language understanding. The ViaVoice speech recognition system[3] analyzes the acoustic signal (1) and produces a text string (2) as output. This string is then analyzed by Nautilus [32], our in-house natural language understanding system, to obtain a semantic representation (3), which is then mapped to a representation (4) similar to a logical form used in propositional logic[4]. Gestures from the different sources (PDA, EUT, and a structured light rangefinder mounted on the robot) are processed by the Gesture Interpreter and provide input to the representation (4). Examples of hand and arm gestures recognized by the robot rangefinder are shown in Fig. 2.

[3] The ViaVoice speech recognition system is sold by IBM.

[4] For expositional purposes, we include the logical representation (4). Although it is not used in checking gestures, it is used for further linguistic analysis where necessary, such as pronominal dereferencing and quantifier scoping. Given (4), therefore, it is possible to interpret such utterances as "Go to the left of it," where *it* is analyzed as a pronoun in a larger discourse, and "How many objects do you see?" where it is necessary to process the number of objects.

(a) (b)

Fig. 2. Examples of gestures given to the robot. (a) The user points to a location with the utterance "Coyote, go over there." (b) The user indicates a distance manually with the utterance "Coyote, back up this far."

The Goal Tracker stores linguistic and state information in a structure we call a *context predicate* [33] where it can be retrieved at a later time if necessary. If an action is stopped, the user can command the robot to continue the stopped action at a later time, using information stored in the Goal Tracker. For example, the user may direct the robot to back up a certain distance and then, before it finishes, issue a stop command. The user might interrupt the robot's movement to ask the robot a question about its environment. After obtaining the requested information, the human then can tell the robot to continue with whatever action it was doing, namely, backing up.

Most of our commands have been direction- or location-oriented, such as "Coyote, go over there," "Coyote, go to the door over there," and "Coyote, go to the left of the pillar." The Spatial Relations component provides necessary object and location information to enable the user and the robot to communicate about those elements in a very natural way. This component extracts spatial relations from sensory information and translates them into linguistic constructs that people naturally use to accomplish navigation tasks. We will consider these in greater detail in Section V.

The Appropriateness/Need Filter determines if an accompanying gesture is necessary and whether or not an appropriate command or query has been issued. This is verified by the first element of the list in (5). Here "2" indicates that a natural vectoring gesture has been perceived. With the linguistic and gestural information, a robot message (6) is sent to the robotics modules for interpretation and mapping into navigation commands. In (6), “4” is arbitrarily mapped to certain functions which cause the robot to move. The second element “45” in (6) indicates that the user vectored 45 degrees from an imaginary line connecting the user and the robot. Finally, the robot will move toward some obstacle in the direction of the human’s gesture, close enough to avoid a collision with the obstacle.[5] This is translated to an appropriate Robot Action.

In the event that the command is not linguistically well-formed, or an appropriate gesture is not perceived, an appropriate message is returned to the user for subsequent error handling in the Speech Output component. These messages are usually synthesized responses informing the user that some error has been detected, e.g., if the user in Fig. 2 does not provide gesture information. In these instances, the robot responds, “Where?" or "How far?" accordingly. Providing information during error handling is an important aspect of the interface, allowing the user to respond intelligently, quickly, and easily to situations.

IV. Semantic Representations for Spatial Language

As noted above, semantic interpretations of the commands are stored in a structure called the *context predicate*. Along with tracking goal states, important spatial information must be obtained and updated, since many utterances involve spatial references. Knowledge of objects and their locations requires a rich semantic representation. Given the sensor information and the linguistic descriptions produced by the Spatial Relations component, we found that the semantic representations we had been using lacked adequate locative and spatial representations to reason about spatial information. Initially, for example, it was sufficient for us to know that commands involving locative information, such as (7), could be represented as (8), a somewhat simplified representation for expositional purposes here.

[5] Currently our vision system does not permit triangulation. Consequently, we simply pass information to the robot module indicating that the robot should move in some general direction specified by the gesture. The use of the “obstacle” element in the string informs the system of the command termination. In the future, we hope to incorporate a more robust vision system where triangulation is possible.

"Coyote, go over there." (7)

```
(imper:                                                   (8)
        ((p-go: go)
              (:agent (system coyote))
              (:loc (location there))))
```

The term *imper* in (8) is an abbreviation of the *imper*ative command of (7). (8) further analyzes the command *go* into a class of semantic predicates *p-go*. *p-go* requires certain semantic roles or arguments to be filled, such as an :*agent* role that is the grammatical subject of the sentence. The :*agent* must be semantically classified as a *system* which is how *Coyote* is defined. Finally, *p-go* requires location information, a *:loc* role, the word *there* that is semantically subcategorized as a *location.*

Given this semantic framework, the commands of (9a,b) generate the same semantic representation (10).

"Coyote, go to the elevator." (9a)

"Coyote, go into the elevator." (9b)

```
(imper:                                                   (10)
          ((p-go: go)
              (:agent (system coyote))
              (:loc (location elevator))))
```

However, (10) misses crucial semantic information, namely, that the ultimate locative goal or location is just in front of the elevator (9a) versus inside it (9b). We, therefore, had to expand our representations.

It is not immediately apparent whether (11a,b) or (12a,b) are adequate representations for the utterances in (9a,b).

```
(imper:                                                   (11a)
        ((p-go-to: go)
              (:agent (system coyote))
              (:loc (location elevator))))
(imper:                                                   (11b)
```

((p-go-into: go)
(:agent (system coyote))
(:loc (location elevator))))

(imper: (12a)
((p-go: go)
(:agent (system coyote))
(:to-loc (location elevator))))
(imper: (12b)
((p-go: go
(:agent (system coyote))
(:into-loc (location elevator))))

(11a,b) compound the number of predicates *go* maps to; namely *p-go-to* and *p-go-into*. (12 a,b) realize only one semantic predicate *p-go* but compound the number of roles of the predicate; namely, *:to-loc* and *:into-loc*.

Both representations capture the locative information for crucially differentiating (9a) and (9b). However, rather than claiming there are several semantic predicates corresponding to the English verb *go,* as realized by the different classes *p-go-to* and *p-go-into* (11a,b), (12a,b) capture the generalization that the English verb *go* maps to a single semantic predicate having multiple roles. Therefore, we choose (12a,b) as adequate semantic representations. Our empirical justification for opting for these representations is simplicity. It seems more intuitively appealing to claim that *go* is a singular semantic verbal concept, taking various locative roles. This conclusion is in keeping with a model-theoretic approach explaining the semantics of locative expressions [34].

Following this line of reasoning, we were able to simplify the representations for sentences like (13) and generalize about other locations, such as elevators and exhibit halls.

"Coyote, the elevator is in front of the exhibit hall." (13)
"Coyote, the elevator is behind the exhibit hall."
"Coyote, the elevator is opposite the exhibit hall."
"Coyote, the elevator is across from the exhibit hall."
"Coyote, the elevator is across the exhibit hall."

Rather than compounding a list of semantic predicates to interpret (13), we map the predicate *be*, syntactically the verb *is* in (13), to a single semantic predicate that we arbitrarily name *be-at-location* having several locative roles (14).

```
(be-at-location: be                                             (14)
        (:theme (location))
           (:in-front-of-loc (location))
           (:behind-loc (location))
           (:relatively-opposite-loc (location))
           (:directly-opposite-loc (location)))
```

In this semantic framework, we maintain the intuitive notion that being in a location is a single semantic concept or predicate, and the actual location is stipulated specifically by a role. In English, this is usually realized as a locative preposition. Therefore, locative and spatial information is mapped to semantic *roles* of predicates rather than to different predicates.

This conclusion may prove to be of interest to other researchers in related fields focusing on spatial relationships. As Tversky and Lee [35] point out, spatial concepts and language are closely related. The semantic representations of locative and spatial information we propose here, therefore, have a direct bearing on the mental models humans employ for communicating spatial relations, as well as on the linguistic structures that people use to communicate that information. While our results are language-specific, namely to English, research in the semantics of locative and spatial expressions in other languages may also show that our claim can be extended to other languages, and to human-robot interfaces employing those languages.

V. Generating Spatial Language from Occupancy Grid Maps

Spatial linguistic terms are extracted directly from range sensor data stored in the evidence grid map. In this section, we discuss the algorithms for translating sensor data into linguistic terms.

A. Preprocessing

The map structure used in this work is a 128 x 128 x 1 cell grid map [10], providing a two-dimensional map of the NRL lab. One cell covers approximately 11cm x 11cm on the horizontal plane.

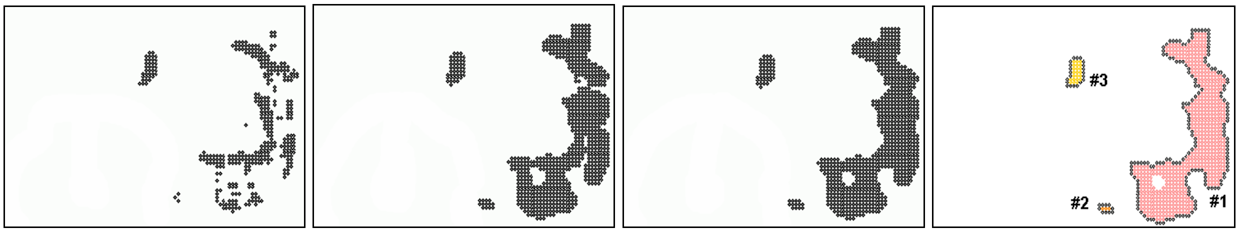
Information from the robot sensors is accumulated over time to calculate probabilities of occupancy for each grid cell; values range from +127 (high probability of occupancy) to -127 (high probability of no occupancy), with 0 representing an unknown occupancy. For the work reported here, these maps are the sensor-fused short-term maps generated by the robot's regular localization and navigation system [36]. An example is shown in Fig. 3(a). A cell with an occupancy value ≥ +1 is considered to be occupied and is shown in black. All other cells are shown in white.

The evidence grid map is pre-processed with a sequence of operations, similar to those used for image processing, to segment the map into individual objects. First, a filter is applied through a convolution operation, using the matrix in (15) as the convolution kernel, *K*.

$$K = \begin{vmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{vmatrix} \qquad (15)$$

This has the effect of blurring the map, filtering single cells and filling in some disconnected regions, as shown in Fig. 3(b). An explicit fill operation is also used to further fill in vacant regions. For each unoccupied cell, if 5 or more of its 8 neighbors are occupied, then the cell status is changed to occupied. Two passes of the fill operation are executed. Results are shown in Fig. 3(c). Finally, spurs are removed. A spur is considered to be an occupied cell with only one occupied neighbor in the four primitive directions (diagonal neighbors are not counted). All spurs, including those with a one-cell length, are removed. At this point, the final cell occupancy has been computed for object segmentation. Objects should be separated by at least a one-cell width.

Next, objects are labeled; occupied cells are initially given numeric labels for uniqueness, e.g., object #1, object #2. A recursive contour algorithm is then used to identify the boundary of the objects. Examples of the final segmented objects, with their identified contours, are shown in Fig. 3(d). See also [37] for more examples.



(a) (b) (c) (d)

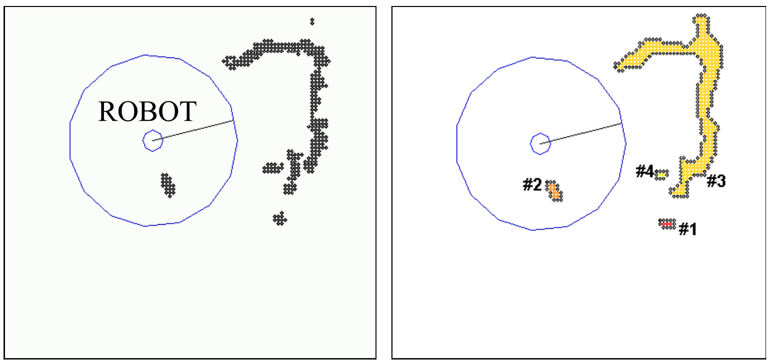
Fig. 3. (a) The southeast part of the evidence grid map. Occupied cells are shown in black. (b) The result of the filter operation. (c) The result of the fill operation. (d) The segmented, labeled map. Physically, object #1 corresponds to a section of desks and chairs, object#2 is a file cabinet, and object #3 is a pillar.

B. Generating Spatial Descriptions of Objects

Spatial modeling is accomplished using the histogram of forces [29,30], as described in previous work [7,37,38,39,40,41]. We first consider the case where the robot is the referent and an environment object is the trajector. For each object, two histograms are computed (the histograms of constant forces and gravitational forces), which represent the relative spatial position between that object and the robot. Computationally, each histogram is the resultant of elementary forces in support of the proposition object #*i* is in direction θ of the robot. For fast computation, a boundary representation is used to compute the histograms. The object boundaries are taken from the contours of the segmented objects in the grid map. The robot contour is approximated with a rectangular bounding box.

Features from the histograms are extracted and input into a system of fuzzy rules to generate a three-part linguistic spatial description: (1) a primary direction (e.g., *the object is in front*), (2) a secondary direction which acts as a linguistic hedge (e.g., *but somewhat to the right*), and (3) an assessment of the description (e.g., *the description is satisfactory*). A fourth part describes the Euclidean distance between the object and robot (e.g., *the object is close*). In addition, a high level description is generated that describes the overall environment with respect to the robot. This is accomplished by grouping the objects into 8 (overlapping) regions located around the robot. An example of the generated descriptions is shown in Fig. 4(c). See [7] for additional details.

One of the features extracted from the force histograms is the main direction, α, of an object with respect to the robot. The main direction, which is comparable to a center of mass, has the highest degree of truth that the object is in direction α of the robot.



(a) (b)

(c)

DETAILED SPATIAL DESCRIPTIONS:
The #1 object is to the right-front of me
-
(the description is satisfactory)
The object is close.
The #2 object is to the right of me
-
(the description is satisfactory)
The object is very close.
The #3 object is in front of me
-
(the description is satisfactory)
The object is close.
The #4 object is mostly in front of me
but somewhat to the right
(the description is satisfactory)
The object is close.
HIGH LEVEL DESCRIPTION:
There are objects in front of me and on my right.

Fig. 4. (a) A robot situated in the grid map. The robot is designated by the small circle with a line indicating its heading. (b) The segmented, labeled map. (c) The generated descriptions. Note the robot heading. Object#2 corresponds to the same pillar in Fig. 2(d).

C. Modeling unoccupied regions for robot directives

To support robot commands such as “Go to the right of the object”, we must first compute target destination points in unoccupied space, which are referenced by environment objects. In this situation, the object is the referent and a destination point in an unoccupied region is the trajector. These trajector points are computed for the four primary directions, left, right, front, and rear of an object, from the robot’s perspective, which is defined by the object’s main direction, α. In keeping with Grabowski’s framework [11], the trajector points are computed *as if* the robot is facing the referent object along its main direction, regardless of the actual heading.

Fig. 5 illustrates the computation of the trajector points. A bounding box is constructed by considering the range of (x, y) coordinates that comprise the object contour.

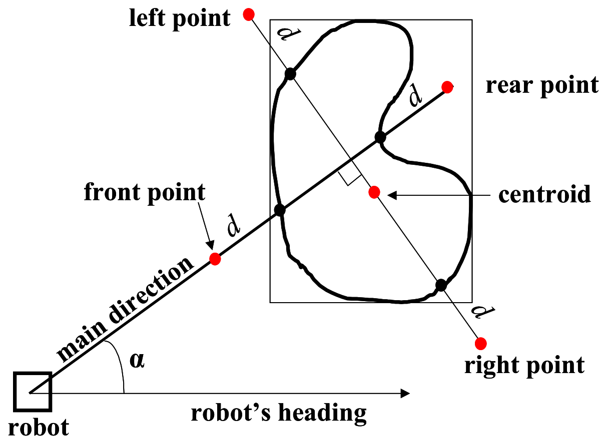


Fig. 5. Computing left, right, front, and rear trajector points in unoccupied space.

The front and rear points are computed to lie on the main direction vector, at a specified distance, *d*, from the object boundary. Consider first the front point. Coordinates are calculated along the main direction vector using the following equations:

$$\begin{aligned} x &= r\cos(\alpha) \\ y &= r\sin(\alpha) \end{aligned} \qquad (16)$$

where α is the main direction, (x,y) is a point along the main direction, and r is the distance of the vector from the robot to the (x,y) point. Coordinate points are computed incrementally, starting from the robot and checked for intersection with the object contour until the intersection point is identified. When the intersection point is found, the front point is computed by subtracting the distance, d, from v_F, the vector length of the front intersection point, and computing a new coordinate.

In computing the rear point, we again search for the intersection point of the contour along the main direction vector, this time starting from behind the object. The algorithm first determines the longest possible line through the object by computing l, the diagonal of the bounding box. The starting vector length used in the search is then $v_F + l$. Once the rear contour intersection point is found, the rear point is computed by adding d to the vector length of the rear intersection point and computing a new coordinate.

The left and right points are computed to lie on a vector that is perpendicular to the main direction and intersects the centroid (x_C, y_C) of the object. Again, a search is made to identify the contour point that intersects this perpendicular vector. The starting point for the search of the right intersection point is shown below:

$$\begin{aligned} x &= x_C + l\cos(\alpha - \frac{\pi}{2}) \\ y &= y_C + l\sin(\alpha - \frac{\pi}{2}) \end{aligned} \tag{17}$$

Once the intersection point is found, a new vector length is computed by adding the distance, d. The left point is found using a similar strategy. Fig. 5 shows some examples. The trajctor points are marked with the diamond polygons around each object; the vertices define the left, right, front, and rear points. More examples can be found in [42].

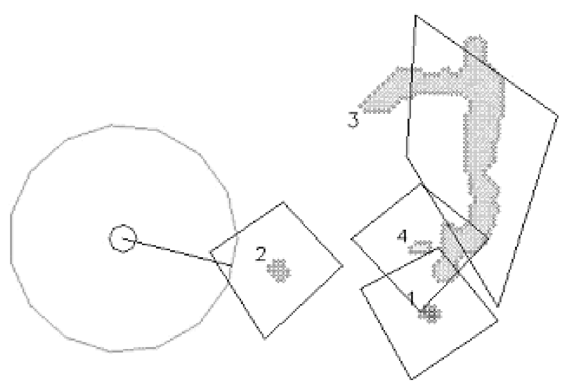


Fig. 6. Computing left, right, front, and rear spatial reference points using the Intersecting Ray Method. The vertices of the polygons mark the positions of the left, right, front, and rear points.

To validate the algorithm, we computed confidence regions using the histogram of forces as described in Sec. V.A, only this time the trajector and referent are switched. For example, a virtual robot is placed at a point in the unoccupied region left of the referent object; we then compute the force histograms to determine whether the robot (now the trajector) really is to the left of the referent object.

The resulting degree of truth is interpreted as a confidence level. In fact, by placing a virtual robot at neighboring positions, this technique can be used to investigate regions that are to the left, right, front, and rear of an object, where the areas are segmented by confidence level.

Fig. 7 shows an example of regions created using this technique, computed for Object 2. Regions for left, right, front, and rear are shown in the figure (from the robot's perspective). The medium gray regions represent a high confidence level, where the cell *i* confidence, $c_i \geq 0.92$. The light gray regions have a medium confidence level ($0.8 < c_i < 0.92$). The dark regions have a low confidence level ($c_i \leq 0.8$). The figure shows that the regions widen as the distance from the object increases. For a relatively small object, the left, right, front, and rear trajector points lie well within the high confidence region, as shown by the polygon vertices. For further analysis and additional examples, see also [42].

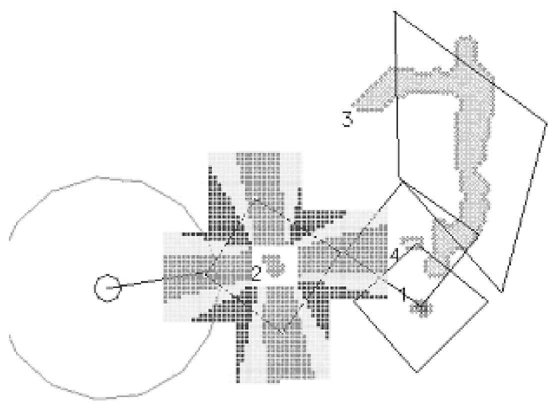


Fig. 7. Confidence Regions around Object 2 for left, right, front, and rear spatial references, from the robot's perspective. Medium gray is high confidence. Light gray is medium confidence. Dark gray is low confidence.

D. Handling Spatial Queries

The spatial language system also supports queries such as, "Where is the nearest object on your left?" To support such queries, 16 symbolic directions are situated around the robot, as shown in Fig. 8. The main direction of each object is discretized into one of these 16 directions. Examples of some corresponding linguistic descriptions are shown in Fig. 8(a). In addition, the 16 symbolic directions are mapped to a set of 8 overlapping regions around the robot (left, right, front, rear, and the diagonals), which are used for queries. Two examples are shown in Fig. 8(b). An object in any of the 5 light gray directions is considered to be in front of the robot. An object in any of the 3 dark gray directions is considered to be to the right rear. Thus, an object that is to the right front (as shown in Fig. 10(a)) would be retrieved in queries for three regions: the front, the right, and the right front of the robot.

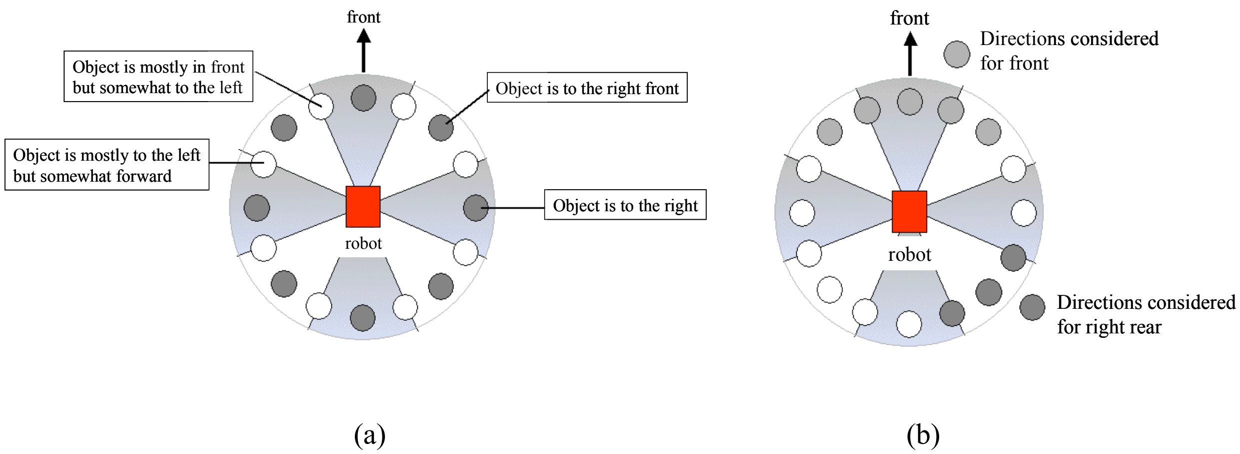


Fig. 8. Sixteen directions are situated around the robot (the small circles). The main direction of each object is discretized into one of these 16 directions. The 8 cone-shaped sections represent the 8 basic regions (front, rear, left, right, and diagonals) used for queries. (a) Examples of the corresponding linguistic descriptions. (b) Examples used for queries. An object is considered to be in front of the robot if it occupies one of the 5 light gray directions. Diagonal directions such as right rear comprise only 3 directions (dark gray).

VI. Integrating Spatial Language into Human-Robot Dialog

With this spatial information and linguistic descriptions, we can now establish a dialog using spatial language. To overcome the object recognition problem (the system does not yet support vision-based object recognition), we have defined a class of persistent objects that are recognized and named by a human user. Persistent objects are created from the segmented objects identified in the grid map through a dialog with the robot. The figures and dialog in this section illustrate how persistent objects are named and present one possible scenario in using sensor-based spatial language within a dialog context. Throughout the dialog, relative spatial terms are consistently given from the robot's perspective, in keeping with the results of experiments with humans [19] and robots [18]. The front of the robot is defined by the placement of the camera and the laser rangefinder.

A. Scenario

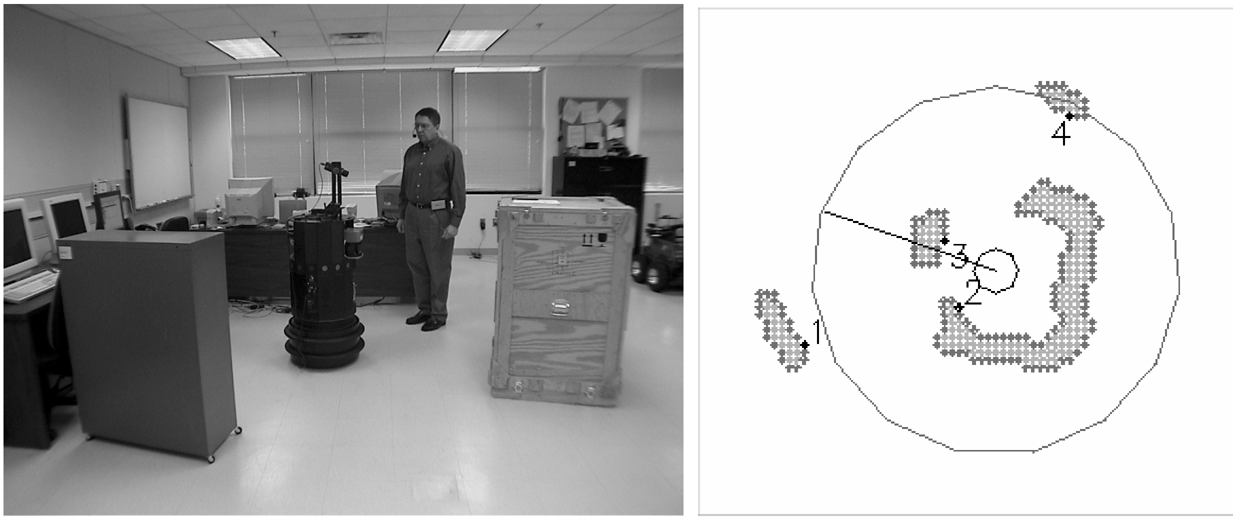
In this example scenario, the user directs the robot from a starting location (scene 1 shown in Fig. 9) to a final destination (scene 4 shown in Fig. 12) through a dialog. To begin, consider the scene in Fig. 9 for the dialog below. Fig. 9(a) shows a picture of the user in the scene, and Fig. 9(b) shows the graphical display presented to the user. The actual display is presented in color and illustrates some features more clearly than the black and white figure shown here. In the scene, object #2 is the group of desks that wrap the corner of the room; the user has been assimilated onto the tip of the object. The scene illustrates an example of the *surrounded* relation. In [7], we introduce 3 levels of surrounded based on the width of the force histograms, e.g., (1) *I am surrounded on the right*, (2) *I am surrounded with an opening on the left*, and (3) *I am completely surrounded.*

The sample dialog shows how the detailed and high level descriptions from Fig. 9(c) are used in a dialog setting. The high level description is used to answer questions about the environment as a whole, whereas the detailed description is used when the question refers to a single object. Note that the user has all of the interface modalities available and can view both the robot and the graphical display of the scene while talking to the robot. The robot responds using synthesized speech output [43].

Human: “How many objects do you see?”
Robot: “I am sensing 4 objects.”
Human: “Where are they?”
Robot: “There are objects in front of me. I am surrounded from the rear. The object # 4 is to the right of me.”
Human: “Where is the nearest object on your left?”
Robot: “The object # 1 is mostly in front of me but somewhat to the left. The object is close.”
Human: “Where is the nearest object in front of you?”
Robot: “The object # 3 is in front of me but extends to the right relative to me. The object is very close.”

The user can also name an object in the scene, thereby creating a persistent object in a location designated by the segmented object in the grid map. In the continuing dialog, the user names object #3 as a “box” and then gives a command to the robot, referencing the box. Fig. 9(d) shows the graphical display after naming the box and issuing the command; the grid cells that comprise the box are shown in black.

Human: “Object #3 is a box.
Robot: “I now know that object # 3 is a box.”
Human: “Go to the right of the box.”



(a) (b)

DETAILED SPATIAL DESCRIPTIONS:
The object # 1 is mostly in front of me but somewhat to the left. The object is close.

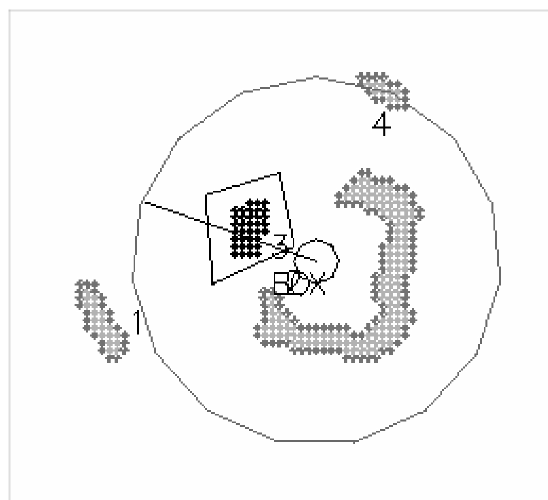
I am surrounded from the rear (surrounded by the object # 2). The object is very close.

The object # 3 is in front of me but extends to the right relative to me. The object is very close.

The object # 4 is to the right of me. The object is very close.

HIGH LEVEL DESCRIPTION:
There are objects in front of me.
I am surrounded from the rear.
The object # 4 is to the right of me.

(c)



(d)

Fig. 9. Scene 1. (a) Querying the robot for the number of objects (b) The graphical display showing the robot situated in the field of obstacles. The robot is designated by the small circle with a line indicating its heading. Object #2 is the group of desks surrounding the robot; the user has been assimilated onto the tip of object. Object #3 is the box. (c) The generated linguistic descriptions before naming the box, corresponding to the display in (b). All descriptions have a satisfactory assessment. (d) The graphical display after naming the box. The box cells are shown in black. The vertices of the diamond show the trajector points for right, left, front, and rear regions relative to the box.

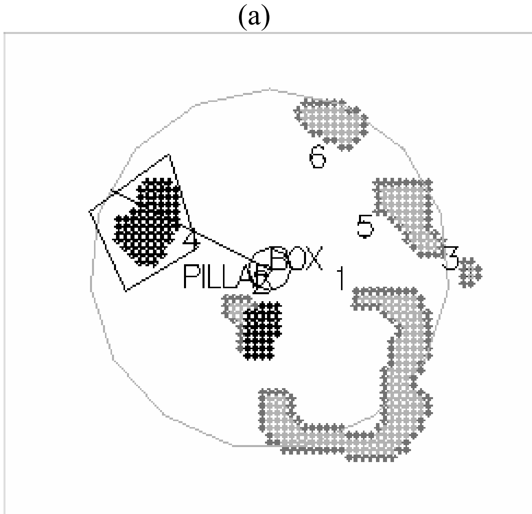
As a result of the command, the robot will move to the trajector point right of the box. The actual route taken is determined by the path planning and obstacle avoidance modules. We are not concerned with the route here, but only whether the robot arrives at a point designated by the dialog. In scene 2 (Fig. 10), the robot has executed the command and has stopped at a point to the right of the box. As the robot moves, new sensor readings are collected and the grid map is updated approximately once a second. A new object segmentation is created for each new map. Fig. 10(b) shows the new graphical display, and Fig. 10(c) shows the linguistic descriptions generated as a result of this updated map. Note that additional grid cells of the box are now shown as occupied; however, the persistent object retains the original set of grid cells as defined in scene 1. When generating the high level descriptions, the system attempts to match new dynamic objects with persistent objects defined earlier by comparing the detailed linguistic descriptions and approximate distances from the robot. In this case, object #2 has been identified as the box and thus does not appear in the high level description. In the continuing dialog, the user queries the robot about the new scene and then defines yet another persistent object and issues another command.

Human: Where is the box?"
Robot: "The box is to the left of me but extends to the rear relative to me. The object is very close.
Human: "The object in front of you is the pillar."
Robot: "I now know the object in front of me is the pillar."
Human: "How many objects do you see?"
Robot: "I am sensing 6 objects."
Human: "Where are they?"
Robot: "There are objects behind me and on my right. The box is to the left of me. The pillar is in front of me."
Human: "Go behind the pillar."

Fig. 11 illustrates scene 3 after the robot has moved behind the pillar. Again the grid map is updated, and a new object segmentation is created. Old objects no longer in view now begin to disappear and new objects emerge. The user can still ask about the new location of persistent objects, as the robot "remembers" which grid cells they occupy. In the dialog below, the user queries the robot and then creates another persistent object and issues the command that takes the robot to its final destination.

Human: “Where is the box?”
Robot: “The box is behind me. The object is close.”
Human: “Where is the pillar?”
Robot: “The pillar is loosely behind-right of me. The object is very close.”
Human: “Object #10 is a table.”
Robot: “I now know object #10 is a table.”
Human: “Go to the front of the table.”

(a)



(b)

DETAILED SPATIAL DESCRIPTIONS:
The object # 1 is behind me but extends to the left relative to me. The object is very close.

The object # 2 is to the left of me but extends to the rear relative to me. The object is very close.

The object # 3 is mostly behind me but somewhat to the right. The object is close.

The object # 4 is in front of me but extends to the left relative to me. The object is very close.

The object # 5 is behind-right of me. The object is very close.

The object # 6 is to the right of me. The object is very close.

The BOX is to the left of me but extends to the rear relative to me. The object is very close.

The PILLAR is in front of me but extends to the left relative to me. The object is very close.

HIGH LEVEL DESCRIPTION:
There are objects behind me and on my right.
The BOX is to the left of me
The PILLAR is in front of me

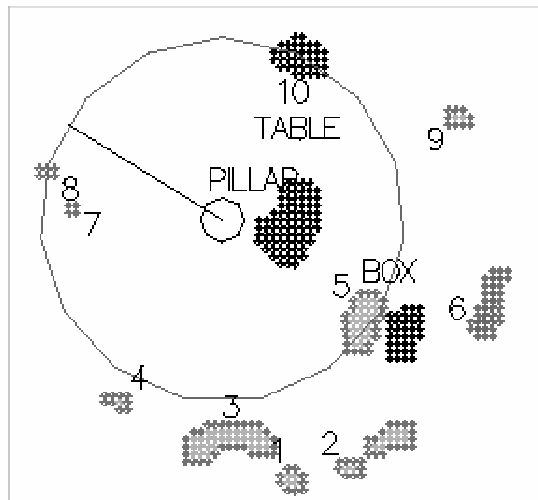
(c)

Fig. 10. Scene 2. (a) The user in the scene. (b) The graphical display showing the robot situated in the field of obstacles after naming the pillar. (c) The generated linguistic descriptions corresponding to the display in (b).

In the final scene (Fig. 12), the robot has arrived at its destination and is situated in front of the table. The dialog below shows the robot's responses to new queries about its surrounding environment.

Human: "How many objects do you see?"
Robot: "I am sensing 7 objects."
Human: "Where is the box?"
Robot: "The box is behind-right of me. The object is close"
Human: Where is the pillar?"
Robot: "The pillar is mostly behind me but somewhat to the right. The object is very close."
Human: Where is the table?"
Robot: "The table is in front of me. The object is very close."

(a)



(b)

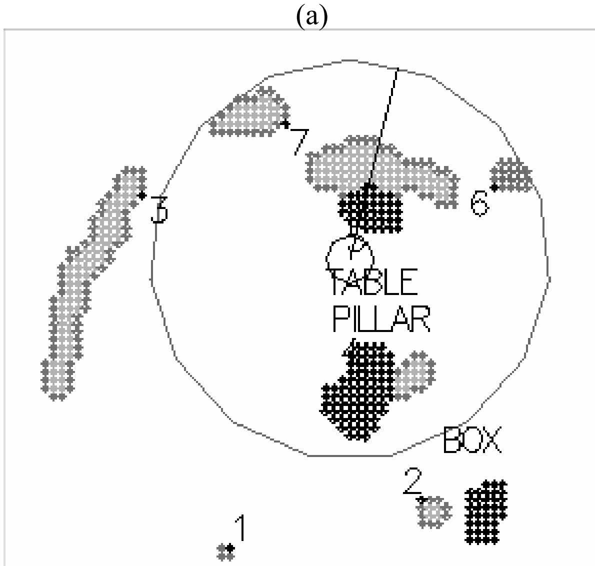
HIGH LEVEL DESCRIPTION:

There are objects in front of me, on my left, behind me and on my right.
The BOX is behind me.
The PILLAR is loosely behind-right of me.
The TABLE is mostly to the right of me.

(c)

Fig. 11. Scene 3. (a) The user in the scene. (b) The graphical display showing the robot situated in the field of obstacles after naming the table. (c) The generated high level description corresponding to the display in (b).

(a)



(b)

DETAILED SPATIAL DESCRIPTIONS:
The object # 1 is mostly behind me but somewhat to the left. The object is close.

The object # 2 is mostly behind me but somewhat to the right. The object is close.

The object # 3 is to the left of me but extends to the rear relative to me. The object is close.

The object # 4 is mostly behind me but somewhat to the right. The object is very close.

The object # 5 is in front of me but extends to the right relative to me. The object is very close.

The object # 6 is to the right-front of me.
The object is very close.

The object # 7 is to the left-front of me.
The object is very close.

HIGH LEVEL DESCRIPTION:
There are objects behind me and on my front right. The object number 7 is to the left-front of me.
The BOX is behind-right of me
The PILLAR is mostly behind me
The TABLE is in front of me

(c)

Fig. 12. Scene 4. (a) The user in the scene. (b) The graphical display showing the robot in its final destination. (c) The generated linguistic descriptions corresponding to the display in (b).

B. Discussion

One of the first questions we considered in this work was which reference frame and perspective should be used? In the case where the robot is the referent, it seems natural to take the robot's perspective. The user asks the robot about its environment, and the robot responds much as a human would (with no object recognition), e.g., there is something in front of me. This is consistent with the literature in cognitive linguistics, e.g., [2, 11]. Queries about specific directions also fall into this category, e.g., where is the nearest object in front?

The second case is more interesting. What perspective should be used when an environment object is the referent, such as in commands issued to the robot, e.g., go to the right of the box? The region to the right of the box will depend on whether we use the robot's perspective or the human user's perspective. The two are inherently different unless the user is situated on top of the robot. This is also true for two people communicating using similar directives. Although the cognitive load for the speaker is higher, experimental results suggest that the speaker will use the hearer's perspective in such situations because it is easier for the hearer and facilitates communication [19]. The limited experimental results with robots confirms this [18]. Once we have decided to use the robot's perspective, we still need to determine a valid robot heading. Again using the literature on humans, we adopt Grabowski's "outside perspective" [11] in which the referent must be in front of the observer (the robot) before computing the spatial relations. Grabowski specifies that the observer must change his orientation if necessary so that the referent is on his front. Thus, our approach of imagining the robot facing the referent object along its main direction is consistent with this outside perspective.

Although we have made some attempts to be consistent with the cognitive linguistics research, the work would benefit from a more rigorous analysis and comparison to human linguistics. In addition, the linguistics discipline may benefit from continued work in the HRI domain. For example, by processing the grid map only (created by range sensors) and not using vision data for recognition, we have separated the location problem from the recognition problem. Landau and Jackendoff suggest that the cognition representation used for location information is much different from the representation used for recognition [3]. The separation here provides a method of studying the two representations independently.

One possible representation of spatial location is the set of force histograms used to generate the spatial language in this work. We have used one set of features extracted from the histograms to generate the linguistic descriptions [30]. However, the force histograms provide a rich selection of possible features that could be examined with cognitive linguistics in mind. The histograms are convenient in our robotics domain because they provide a common representation that can be used with a variety of sensor types and structures. Once the histograms are computed, the spatial linguistic terminology is readily available [38].

Finally, we must consider whether the spatial language is of any value to the user for interacting with robots. We have not yet done a rigorous user study to evaluate the interface techniques described in the scenario above. Such a study would test the overall concept and also guide our refinements, e.g., in adjusting the level of detail provided by the linguistic descriptions. We have done a 4-person pilot study to investigate multimodal robot interface requirements in preparation for a future study. Modalities included live video, PDA, EUT, touch screen, and gestures as well as language. The user could choose any combination of these as desired for performing a remote search task (the user was stationary while the robot was directed to search for a specific target). The study has been done in the form of a wizard of oz experiment; test subjects were told that they were controlling a robot but in reality a human wizard drove the robot.

The preliminary results show a great deal of variety in the interface modalities selected by the test subjects, and it is not possible to draw any conclusions from such a small number of subjects. As noted in the experimental results described in [18], test subjects here often used short commands such as "turn left" or "move slightly forward" instead of using target object references. The test subjects seemed cautious about their use of the interface modalities, so perhaps they too did not believe that the robot could understand target object references. In addition, the users were situated at a distance from the robots which may detract from the type of interaction described in our scenario.

To better test the spatial language component of the interface, we suggest a more constrained evaluation environment in which modalities are tested individually and then compared. For example, one group of test subjects may perform the task using spatial language, another using a touch screen map (perhaps with incomplete information) and another using a joystick. A possible evaluation task is to search for an object in a cluttered environment and then direct the robot to move close to the object (as if to pick it up) and then return back to the starting point (a pseudo fetch task). The interface modalities could be compared by capturing metrics such as success or failure of the task within the allocated time, total time to perform the task if successful, time to reach the target, time to return, time spent searching for the target, time spent in navigating through the environment, total distance traveled, smoothness of motion (acceleration of movement), and a measure of user satisfaction. The searching part of the task and the navigation part of the task are different and may require different interface modalities.

In addition to the formal evaluation, we also see the need for future enhancements. Certainly, the addition of vision for even limited object recognition would be a dramatic improvement in creating a more natural interface. We view the user's assignment of names to persistent objects as a temporary solution that allows us to test some of the other interface issues. With limited recognition capabilities, vision would provide spatial information in the vertical plane and allow us to incorporate additional spatial relations such as *above*, *below*, and *on top*. The use of the force histograms as a representation provides a straightforward framework for investigating additional relations. We are also in the process of extending the relations to include *between* and *through*, and there are others that we might explore, such as *across* (an unoccupied region).

VII. Concluding Remarks

The motivation behind this work was to create an intuitive interface for mobile robots that would be easy and natural for novice users. Both humans and robots need essential skills in spatial reasoning, especially for handling unstructured environments. Our premise was that providing robots with human-like spatial language capabilities would facilitate an intuitive interface for humans that would be consistent with their innate spatial cognition.

To provide robots with this human-like spatial language, we have had to address a number of issues, many of them in the area of representation. First, we looked at spatial semantics and formulated a framework for representing spatial relations, in a simple yet general way. This is a crucial first step and allows us to support a variety of spatial relationships. Next, we considered the representation of spatial information in the environment. The evidence grid map is a common structure for storing an environment map based on range sensor data. In this paper, we presented several algorithms for extracting spatial information from the map and generating spatial language. The language generated includes the situation where the robot is the referent, so that the robot can provide an egocentric description back to the human. We also looked at the case where an environment object was the referent, to support human-to-robot commands such as "Go to the right of the table." We view both the semantic representations and the spatial algorithms as significant contributions of this work.

To investigate our original premise that spatial language would help the user, we have had to integrate the spatial language into a robot interface that supports natural language understanding. This prototype provides a platform in which to further investigate the HRI issues using real robots and real sensor data. In this work, we have illustrated an example of the spatial language in a human-robot dialog. Our intuition tells us that language is a good communication modality for at least some tasks, but the work has not been evaluated with real test subjects. There are still further questions to be answered. Does the general concept work and under what conditions? Does the level of detail in the language need adjusting? Is the perspective the best? We intend to address these questions in the future as we continue to assert that supporting spatial language contributes to a more natural human-robot interface.

Acknowledgements

The authors would like to acknowledge the help of Scott Thomas and Myriam Abramson at NRL, Dr. Pascal Matsakis at the University of Guelph, and. Dr. Jim Keller at UMC.

References

[1] F.H. Previc, "The Neuropsychology of 3-D Space", *Psychological Review*, vol. 124, no. 2, pp. 123-164, 1998.

[2] L. Talmy, "How Language Structures Space", In *Spatial Orientation: Theory, Research and Application*, H. Pick and L. Acredolo (Eds.), New York: Plenum Press, 1983.

[3] B. Landau and R. Jackendoff, "What and Where in Spatial Language and Spatial Cognition", In *Behavioral and Brain Science*, vol. 16, pp. 217-265, 1993.

[4] T. Regier and L. Carlson, "Grounding Spatial Language in Perception: An Empirical and Computational Investigation", *Journal of Experimental Psychology, General*, vol. 130, no. 2, pp. 273-298, 2001.

[5] R. Langacker, *Foundations of Cognitive Grammar: Theoretical Prerequisites*, vol. I, Stanford Univ. Press, Stanford, CA, 1987.

[6] A. Herskovits, “Semantics and Pragmatics of Locative Expressions”, *Cognitive Science*, vol. 9, pp. 341-378, 1985.

[7] M. Skubic, P. Matsakis, G. Chronis, and J. Keller, “Generating Multi-Level Linguistic Spatial Descriptions from Range Sensor Readings Using the Histogram of Forces,” *Autonomous Robots*, vol. 14, no. 1, pp. 51-69, Jan., 2003.

[8] D. Perzanowski, A.C. Schultz, W. Adams, E. Marsh, M. Bugajska, “Building a Multimodal Human-Robot Interface”, *IEEE Intelligent Systems*, pp. 16-20, Jan./Feb, 2001.

[9] W. Adams, D. Perzanowski, A.C. Schultz, “Learning, Storage and Use of Spatial Information in a Robotics Domain”, in *Proc. of the ICML 2000 Workshop on Machine Learning of Spatial Language*, Stanford Univ.: AAAI Press, 2000, pp. 23-27.

[10] M.C. Martin, H.P. Moravec, “Robot Evidence Grids”, Carnegie Mellon University, Pittsburgh, PA, Technical Report #CMU-RI-TR-96-06, Mar., 1996.

[11] J. Grabowski, “A Uniform Anthropomorphological Approach to the Human Conception of Dimensional Relations”, *Spatial Cognition and Computation*, vol. 1, pp. 349-363, 1999.

[12] R. Muller, T. Rofer, A. Landkenau, A. Musto, K. Stein, and A. Eisenkolb, “Coarse Qualitative Descriptions in Robot Navigation”, in *Spatial Cognition II*. Lecture Notes in Artificial Intelligence 1849, C. Freksa, W. Braner, C. Habel and K. Wender (Eds.) Springer-Verlag, Berlin, 2000, pp. 265-276.

[13] W. Gribble, R. Browning, M. Hewett, E. Remolina, and B. Kuipers, “Integrating Vision and Spatial Reasoning for Assistive Navigation”, in *Assistive Technology and Artificial Intelligence*. Lecture Notes in Computer Science, V. Mittal, H. Yanco, J. Aronis and R. Simpson (Eds.), Springer-Verlag, Berlin, 1998, pp. 179-193

[14] B. Kuipers, “A Hierarchy of Qualitative Representations for Space”, in *Spatial Cognition*. Lecture Notes in Artificial Intelligence 1404, C. Freksa, C. Habel, and K. Wender (Ed.), Berlin: Springer-Verlag, 1998, pp. 337-350.

[15] J. Zelek, “Human-Robot Interaction with a Minimal Spanning Natural Language Template for Autonomous and Tele-operated Control”, in *Proceedings of the 1997 IEEE International Conference on Robots and Intelligent Systems*, Grenoble, France, vol. 1, Sept., 1997, pp. 299-305.

[16] E. Stopp, K-P. Gapp, G. Herzog, T. Laengle and T. Lueth, “Utilizing Spatial Relations for Natural Language Access to an Autonomous Mobile Robot”, in *Proceedings of the 18th German Annual Conference on Artificial Intelligence*, Berlin, Germany, 1994, pp. 39-50.

[17] J. Schirra and E. Stopp, “ANTLIMA: A Listener Model with Mental Images”, in *Proceedings of the 13th IJCAI*, pp. 175-180, Chambery, France, 1993.

[18] R. Moratz, K. Fischer and T. Tenbrink, “Cognitive Modeling of Spatial Reference for Human-Robot Interaction”, *International Journal on Artificial Intelligence Tools*, vol. 10, no. 4, pp. 589-611, 2001.

[19] B. Tversky, P. Lee and S. Mainwaring, “Why Do Speakers Mix Perspective?”, *Spatial Cognition and Computation*, vol. 1, pp. 399-412, 1999.

[20] R.D. Bolt, “Put-that-there: voice and gesture at the graphics interface,” *Computer Graphics,* vol. 14, no. 3, 1980, pp. 262-270.

[21] D. Kortenkamp, E. Huber and P. Bonasso, “Recognizing and Interpreting Gestures on a Mobile Robot”, in *Proceedings of the Thirteenth National Conference on Artificial Intelligence and the*

Eighth Innovative Applications of Artificial Intelligence Conference, vol.2, Portland, OR, Aug. 1996, pp. 915-21.

[22] T.W. Fong, F. Conti, S. Grange and C. Baur, "Novel Interfaces for Remote Driving: Gesture, Haptic, and PDA", in *SPIE 4195-33, SPIE Telemanipulator and Telepresence Technologies VII*, Boston, MA, November 2000, pp. 300-311.

[23] C. Rich, C.L. Sidner, and N. Lesh, "COLLAGEN: Applying collaborative discourse theory to human-computer interaction", *AI Magazine* vol. 22, no. 4, pp. 15-25, 2001.

[24] J.F. Allen, D.K. Byron, M, Dzikovska, G. Ferguson, L. Galescu and A. Stent, "Toward conversational human-computer interaction, *AI Magazine* vol. 22, no. 4: 27-37, 2001.

[25] D. Randell, Z. Cui and A. Cohn, "A Spatial Logic Based on Regions and Connections", in B. Nebel, W. Swartout and C. Rich (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the 3rd Intl. Conference, Cambridge, MA, Oct., 1992, pp. 165-176.

[26] I. Bloch, "Fuzzy Relative Position between Objects in Image Processing: New Definition and Properties Based on a Morphological Approach," *Int. J. of Uncertainty Fuzziness and Knowledge-Based Systems*, vol. 7, no. 2, pp. 99-133, 1999.

[27] J. Freeman, "The Modelling of Spatial Relations," *Computer Graphics and Image Processing*, vol. 4, pp. 156-171, 1975.

[28] K.P. Gapp, "Angle, Distance, Shape, and their Relationship to Projective Relations," in *Proceedings of the 17th Conf. of the Cognitive Science Society*, 1995, pp. 112-117.

[29] P. Matsakis and L. Wendling, "A New Way to Represent the Relative Position between Areal Objects", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 634-643, 1999.

[30] P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera, "Linguistic Description of Relative Positions in Images," *IEEE Trans. on Systems, Man and Cybernetics*, *Part B*, vol. 31, no. 4, pp. 573-588, 2001.

[31] R. Bondugula, J. Keller, P. Matsakis and O. Sjahputera, "Learning Spatial Relationships", submitted to *Pattern Recognition Letters*, 2003.

[32] K. Wauchope, *Eucalyptus: Integrating Natural Language Input with a Graphical User Interface,* Naval Research Laboratory, Washington, DC, Technical Report NRL/FR/5510-94-9711, 2000.

[33] D. Perzanowski, A. Schultz, W. Adams, and E. Marsh, "Goal Tracking in a Natural Language Interface: Towards Achieving Adjustable Autonomy," in *Proc. of the 1999 IEEE Intl. Symp. on Computational Intelligence in Robotics and Automation,* Monterey, CA, 1999, pp.208-213.

[34] M. Kracht, "On the Semantics of Locatives," *Linguistics and Philosophy*, vol. 25, pp. 157-232, 2002.

[35] B. Tversky and P. Lee, "How Space Structures Language", in *Spatial Cognition*. Lecture Notes in Artificial Intelligence 1404, C. Freksa, C. Habel, and K. Wender (Ed.), Berlin: Springer-Verlag, 1998, pp. 157-175.

[36] A. Schultz, W. Adams and B. Yamauchi, "Integrating Exploration, Localization, Navigation and Planning with a Common Representation," *Autonomous Robots*, vol.6, no.3, pp. 293-308, June, 1999.

[37] M. Skubic, D. Perzanowski, A. Schultz, and W. Adams, "Using Spatial Language in a Human-Robot Dialog," in *Proceedings of the IEEE 2002 International Conference on Robotics and Automation*, vol. 4, Washington, D.C., May, 2002, pp. 4143-4148.

[38] J. Keller, P. Matsakis, and M. Skubic, “Beyond 2001: The Linguistic Spatial Odyssey”, in *Computational Intelligence: The Experts Speak*, C. Robinson, ed, Wiley, 2003, pp. 11-24.

[39] M. Skubic, G. Chronis, P. Matsakis and J. Keller, “Generating Linguistic Spatial Descriptions from Sonar Readings Using the Histogram of Forces”, in *Proc. of the 2001 IEEE Intl. Conf. on Robotics and Automation*, Seoul, Korea, May, 2001, pp. 485-490.

[40] M. Skubic, P. Matsakis, B. Forrester and G. Chronis, “Extracting Navigation States from a Hand-Drawn Map”, in *Proc. of the 2001 IEEE Intl. Conf. on Robotics and Automation*, Seoul, Korea, May, 2001, pp. 259-264.

[41] M. Skubic, G. Chronis, P. Matsakis and J. Keller. “Spatial Relations for Tactical Robot Navigation”, in *Proc. of the SPIE, Unmanned Ground Vehicle Technology III*, vol. 4364, Orlando, FL April, 2001, pp. 377-387.

[42] M. Skubic and S. Blisard, “Go to the Right of the Pillar: Modeling Unoccupied Spaces for Robot Directives,” Technical Report, AAAI 2002 Fall Symposium, Human-Robot Interaction Workshop, Nov., 2002.

[43] D. Perzanowski , M. Skubic, A. Schultz, W. Adams, M. Bugajska, K. Wauchope, and E. Marsh, “Multi-Modal Navigation of Robots Using Spatial Relations: A Videotaped Demonstration*”, Video Proceedings, IEEE 2002 International Conference on Robotics and Automation*, Washington, D.C., May, 2002