

Extracting Navigation States from a Hand-Drawn Map

Marjorie Skubic, Pascal Matsakis, Benjamin Forrester and George Chronis
Dept. of Computer Engineering and Computer Science,
University of Missouri-Columbia,
email: skubic@cecs.missouri.edu

Abstract

Being able to interact and communicate with robots in the same way we interact with people has long been a goal of AI and robotics researchers. In this paper, we propose a novel approach to communicating a navigation task to a robot, which allows the user to sketch an approximate map on a PDA and then sketch the desired robot trajectory relative to the map. State information is extracted from the drawing in the form of relative, robot-centered spatial descriptions, which are used for task representation and as a navigation language between the human user and the robot. Examples are included of two hand-drawn maps and the linguistic spatial descriptions generated from the maps.

1. Introduction

Being able to interact and communicate with robots in the same way we interact with people has long been a goal of AI and robotics researchers. Much of the robotics research has emphasized the goal of achieving autonomous robots. However, this ambitious goal presumes that robots can accomplish human-like perception, reasoning, and planning as well as achieving human-like interaction capabilities.

In our research, we are less concerned with creating autonomous robots that can plan and reason about tasks, and instead we view them as semi-autonomous tools that can assist a human user. The user supplies the high-level and difficult reasoning and strategic planning capabilities. We assume the robot has some perception capabilities, reactive behaviors, and perhaps some limited reasoning abilities that allow it to handle an unstructured and possibly dynamic environment.

In this scenario, the interaction and communication mechanism between the robot and the human user becomes very important. The user must be able to easily communicate what needs to be done, perhaps at different levels of task abstraction. In particular, we would like to provide an intuitive method of communicating with robots that is easy for users that are not expert robotics engineers. We want domain experts to define their own task use of robots, which may involve controlling them, guiding them, or even programming them.

As part of our ongoing research on human-robot interaction, we have been investigating the use of spatial relations in communicating purposeful navigation tasks.

Linguistic, human-like expressions that describe the spatial relations between a robot and its environment provide a symbolic link between the robot and the user, thus comprising a type of navigation language. The linguistic spatial expressions can be used to establish effective two-way communications between the robot and the user, and we have approached the issue from both perspectives.

From the robot perspective, we have studied how to recognize the current (qualitative) state in terms of egocentric spatial relations between the robot and objects in the environment, using sensor readings only (i.e., with no map or model of the environment). Linguistic spatial descriptions of the state are then generated for communication to the user. See our companion paper [1] for details on the approach used.

In this paper, we focus on the user perspective, and offer one approach for communicating a navigation task to a robot, which is based on robot-centered spatial relations. Our approach is to let the user draw a sketch of an environment map (i.e., an approximate representation) and then sketch the desired robot trajectory relative to the map. State information is extracted from the drawing on a point by point basis along the sketched robot trajectory. We generate a linguistic description for each point and show how the robot transitions from one qualitative state to another throughout the desired path. A complete navigation task is represented as a sequence of these qualitative states based on the egocentric spatial relations, each with a corresponding navigation behavior. We assume the robot has pre-programmed or pre-learned, low-level navigation behaviors that allow it to move safely around its unstructured and dynamic environment without hitting objects. In this approach, the robot does not have a known model or map of the environment, and the user may have only an approximate map. Thus, the navigation task is built upon relative spatial states, which become qualitative states in the task model.

The idea of using linguistic spatial expressions to communicate with a semi-autonomous mobile robot has been proposed previously. Gribble *et al* use the framework of the Spatial Semantic Hierarchy for an intelligent wheelchair [2]. Perzanowski *et al* use a combination of gestures and linguistic directives such as “go over there” [3]. Shibata *et al* use positional relations to overcome ambiguities in recognition of landmarks [4]. However, the idea of communicating with a mobile robot

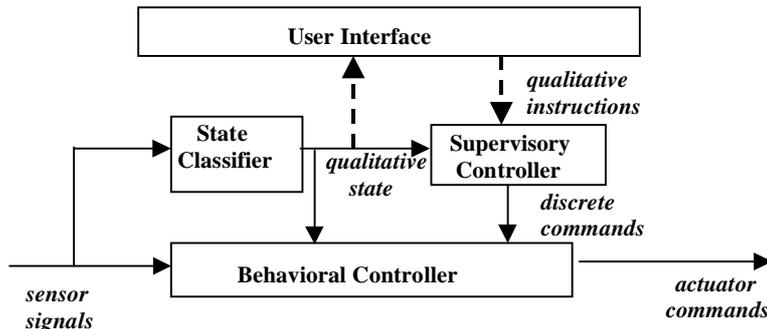


Figure 1. The User Interface and Robot Control Architecture.

via a hand-drawn map appears to be novel. The strategy of using a sketch with spatial relations has been proposed by Egenhofer as a means of querying a geographic database [5]. The hand-drawn sketch is translated into a symbolic representation that can be used to access the geographic database.

In this paper, we show how egocentric spatial relations can be extracted from a hand-drawn map sketched on a PDA. In Section 2, we discuss background material on the human-robot interaction framework. In Section 3, we show the method for extracting the environment representation and the corresponding states from the PDA sketch. Experiments are shown in Section 4 with two examples of hand-drawn maps and the spatial descriptions generated. We conclude in Section 5.

2. Framework for Human-Robot Interaction

Much of our research efforts in human-robot interaction have been directed towards extracting robot task information from a human demonstrator. Figure 1 shows the framework for the robot control architecture and the user interface.

2.1 Robot Control Components

We consider procedural tasks (i.e., a sequence of steps) and represent task structure as a Finite State Automaton (FSA) in the *Supervisory Controller*, following the formalism of the Discrete Event System (DES) [6]. The FSA models behavior sequences that comprise a task; the sensor-based qualitative state (QS) is used for task segmentation. The change in QS is an event that corresponds to a change in the behavior. Thus, the user demonstrates a desired task as a sequence of behaviors using the existing behavior primitives and identifiable QS's, and the task structure is extracted in the form of the FSA. During the demonstration, the QS and the FSA is provided to the user to ensure that the robot is learning the desired task structure. With an appropriate set of QS's and primitive behaviors, the FSA and supervisory controller is straightforward. Also, this task structure is

consistent with structure inherently used by humans for procedural tasks, making the connection easier for the human. We have used this approach in learning force-based assembly skills from demonstration, where a qualitative contact state provided context [7]. For navigation tasks, spatial relations provide the QS context.

With the *State Classifier* component, the robot is provided with the ability to recognize a set of qualitative states, which can be extracted from sensory information, thus reflecting the current environmental condition. For navigation skills, robot-centered spatial relations provide context (e.g., *there is an object to the left front*). Adding the ability to recognize classes of objects provides additional perception (e.g., *there is a person to the left front*).

The robot is also equipped with a set of primitive (reactive) behaviors and behavior combinations, which is managed by the *Behavioral Controller*. Some behaviors may be preprogrammed and some may be learned off-line using a form of unsupervised learning. The user can add to the set of behaviors by demonstrating new behaviors which the robot learns through supervised learning, thus allowing desired biases of the domain expert to be added to the skill set.

Note that this combination of discrete event control in the Supervisory Controller and the "signal processing" in the Behavioral Controller is similar to Brockett's framework of hybrid control systems [8].

2.2 User Interface

As shown in Figure 1, the interface between the robot and the human user relies on the qualitative state for two-way communications. In robot-to-human communications, the QS allows the user to monitor the current state of the robot, ideally in terms easily understood (e.g., *there is an object on the right*). In human-to-robot communications, commands are segmented by the QS, termed qualitative instructions in the figure (e.g., *while there is an object on the right, move forward*).

The key to making the interactive robot training work is the QS, especially in the following ways: (1) the ability to perceive an often ambiguous context based on sensory conditions, especially in terms that are understandable for the human trainer, (2) choosing the right set of QS's so as to communicate effectively with the trainer, and (3) the ability to perform self-assessment, as in knowing how well the QS is identified which helps in knowing when to get further instruction. Spatial relationships provide powerful cues for humans to make decisions; thus, it is plausible to investigate their use as a qualitative state for robot tasks, as well as a linguistic link between the human and the robot.

3. Extracting Spatial Relations States

The interface used for drawing the robot trajectory maps is a PDA (e.g., a PalmPilot). The stylus allows the user to sketch a map much as she would on paper for a human colleague. The PDA captures the string of (x,y) coordinates sketched on the screen and sends the string to a computer for processing (the PDA connects to a PC through a serial port).

The user first draws a representation of the environment by sketching the approximate boundary of each object. During the sketching process, a delimiter is included to separate the string of coordinates for each object in the environment. After all of the environment objects have been drawn, another delimiter is included to indicate the start of the robot trajectory, and the user sketches the desired path of the robot, relative to the sketched environment. An example of a sketch is shown in Figure 2, where each point represents a captured (x,y) screen pixel.

For each point along the trajectory, a view of the environment is built, corresponding to the radius of the sensor range. The left part of Figure 3 shows a sensor radius superimposed over a piece of the sketch. The sketched points that fall within the scope of the sensor radius represent the portion of the environment that the robot can sense at that point in the path.

The points within the radius are used as boundary vertices of the environment object that has been detected. They define a polygonal region (Figure 3, step (a)) whose relative position with respect to the robot (assimilated to a square) is represented by two histograms (Figure 3, step (b)): the histogram of constant forces and the histogram of gravitational forces [9][1]. These two representations have very different and interesting characteristics. The former provides a global view of the situation and considers the closest parts and the farthest parts of the objects equally. The latter provides a more local view and focuses on the closest parts.

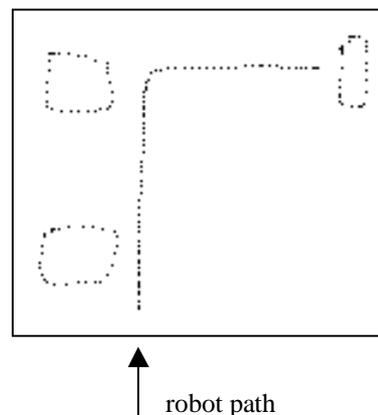


Figure 2. A sketched map on the PDA. Environment objects are drawn as a boundary representation. The robot path starts from the bottom.

The notion of the histogram of forces, introduced by Matsakis and Wendling, ensures processing of raster data as well as vector data, offers solid theoretical guarantees, allows explicit and variable accounting of metric information, and lends itself, with great flexibility, to the definition of fuzzy directional spatial relations (such as *to the right of, in front of, etc.*).

For our purposes, it also allows for a low-computational handling of heading changes in the robot's orientation and makes it easy to switch between a world view and an egocentric robot view. The heading is computed as the direction formed by the current point and the second previous point along the sketched path. A pixel gap in the heading calculation serves to smooth out the trajectory somewhat, thereby compensating for the discrete pixels.

The histogram of constant forces and the histogram of gravitational forces associated with the robot and the polygonal region are used to generate a linguistic description of the relative position between the two objects. The method followed is the method described in [10][11] (and applied to LADAR image analysis).

First, eight numeric features are extracted from the analysis of each histogram (Figure 3, step (c)). They constitute the "opinion" given by the considered histogram. The two opinions (*i.e.*, the sixteen values) are then combined (Figure 3, step (d)). Four numeric and two symbolic features result from this combination. They feed a system of fuzzy rules that outputs the expected linguistic description.

The system handles a set of adverbs (like *mostly, perfectly, etc.*) which are stored in a dictionary, with other terms, and can be tailored to individual users. Each description generated relies on the sole primitive directional relationships: *to the right of, in front of, to the left of, and behind.*

The spatial description is generally composed of three parts. The first part involves the primary direction (e.g., *the object is mostly to the right of the robot*). The second part supplements the description and involves a secondary direction (e.g., *but somewhat to the rear*). The third part indicates to what extent the four directional relationships are suited to describing the relative position between the robot and the object (e.g., *the description is satisfactory*). In other words, it indicates to what extent it

is necessary to utilize other spatial relations (e.g., *surrounds*).

Figure 4 shows the linguistic description generated for some point on the robot path. In this example, a secondary direction is not generated because the primary direction clause is deemed to be adequate. Figure 5 shows a second example along the robot path, with the three-part linguistic spatial description generated for that point.

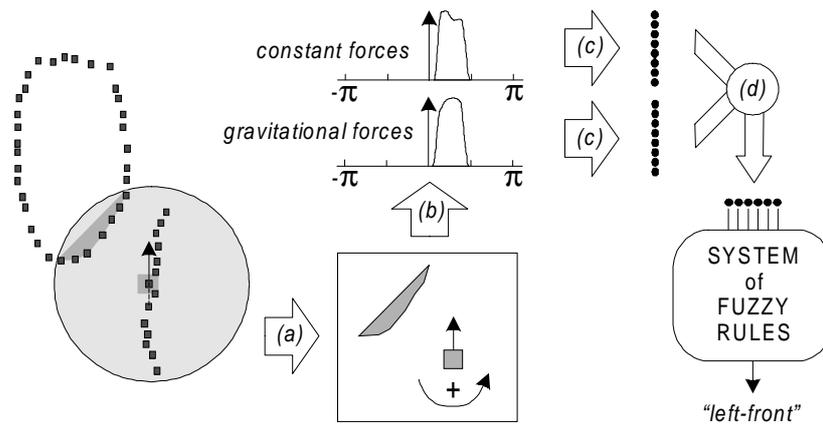
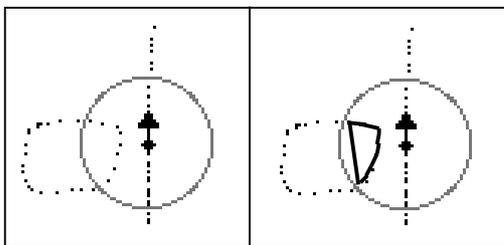
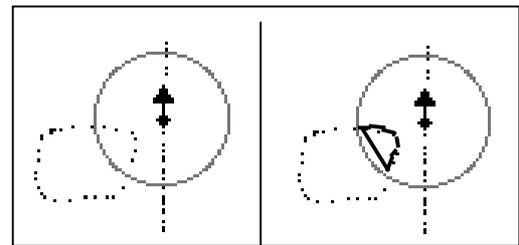


Figure 3. Synoptic diagram. (a) Construction of the polygonal objects. (b) Computation of the histograms of forces. (c) Extraction of numeric features. (d) Fusion of information.



“Object is to the left of the Robot
(the description is satisfactory)”



“Object is mostly to the left of the Robot
but somewhat to the rear
(the description is satisfactory)”

Figure 4. Building the environment representation for one point along the trajectory, shown with the generated linguistic expression.

Figure 5. Another example with a three-part linguistic spatial description generated.

4. Experiments

Experiments were performed on two hand-drawn maps to study the linguistic spatial descriptions generated. The first map is shown in its raw (pixel) state in Figure 2. The user first draws the three objects in the bottom left, top left and top right locations. Then, she draws a desired robot trajectory starting from the bottom of the PDA screen.

Representative spatial descriptions are shown in Figure 6 for several points, labeled 1 through 11, along

the sketched robot trajectory. The assessment was always satisfactory so it is not specified on the figure. Note that the heading is also calculated and used in determining the robot-centered spatial relations. The sensor radius was set to 22 pixels.

At position 1, part of the object A is detected to the left-front of the robot, according to the generated linguistic description. As the robot proceeds through positions 2, 3, and 4, the parts of A that are within the 22 pixel radius are processed and the corresponding

linguistic descriptions are shown in the figure. At position 5, there is nothing within the sensor radius of the robot, so no linguistic descriptions are generated. At points 6, 7, and 8 we observe a sharp right turn. The corresponding parts of the second object **B** that fall within the sensor radius at each point are expressed in linguistic terms. At point 9, the robot is again between objects and nothing is within the sensor radius. Finally, part of the last object **C** is detected to the front of the robot at position 10, and at position 11 an extension of the part of **C** also falls within the radius to the right of the robot.

Figure 7 shows the second map sketched on the PDA. To experiment with a different scaling factor, the sensor radius was set to 30 pixels. Several spatial descriptions are shown in Figure 8. All linguistic descriptions were accepted as satisfactory. An interesting variation in this second experiment is the simultaneous detection of two different objects, namely **A** and **B**. For positions 3, 4, and 5, we show the linguistic descriptions while the robot passes between **A** and **B**.

These experiments indicate the feasibility of using spatial relations to analyze a sketched robot map and trajectory, but much work remains to be done. The limited resolution of the PDA screen results in abrupt changes of the robot heading, which can affect the accuracy of the description generated. The current algorithm for building the object representation for the map cannot handle all cases (e.g., concave objects). Also, we need to study the granularity of the spatial descriptions generated. While they are descriptive for human users, they may be too detailed for use in navigation task representation. The next step is to perform further experiments and extract the corresponding navigation behavior to study the granularity issue.

5. Concluding Remarks

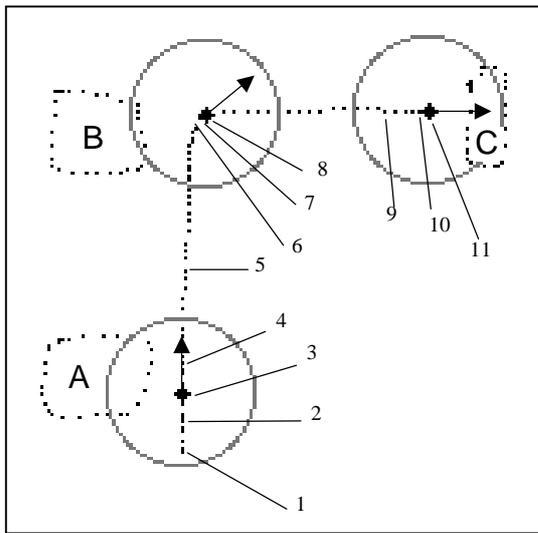
In this paper we have proposed a novel approach for human-robot interaction, namely showing a robot a navigation task by sketching an approximate map on a PDA. The interface utilizes spatial descriptions that are generated from the map using the histogram of forces. The approach represents a first step in studying the use of spatial relations as a symbolic language between a human user and a robot for navigation tasks.

Acknowledgements

The authors wish to acknowledge support from ONR, grant N00014-96-0439 and the IEEE Neural Network Council for a graduate student summer fellowship for Mr. Chronis. We also wish to acknowledge Dr. Jim Keller for his helpful discussions and suggestions.

References

- [1] M. Skubic, G. Chronis, P. Matsakis and J. Keller, "Generating Linguistic Spatial Descriptions from Sonar Readings Using the Histogram of Forces", submitted for the 2001 IEEE International Conference on Robotics and Automation.
- [2] 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*, V. Mittal, H. Yanco, J. Aronis and R. Simpson, ed., Springer Verlag, , 1998, pp. 179-193, Berlin, Germany.
- [3] D. Perzanowski, A. Schultz, W. Adams and E. Marsh, "Goal Tracking in a Natural Language Interface: Towards Achieving Adjustable Autonomy", In *Proceedings of the 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Monterey, CA, Nov., 1999, pp. 208-213.
- [4] F. Shibata, M. Ashida, K. Kakusho, N. Babaguchi, and T. Kitahashi, "Mobile Robot Navigation by User-Friendly Goal Specification", In *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication*, Tsukuba, Japan, Nov., 1996, pp. 439-444.
- [5] M.J. Egenhofer, "Query Processing in Spatial-Query-by-Sketch", *Journal of Visual Languages and Computing*, vol. 8, no. 4, pp. 403-424, 1997.
- [6] P.J. Ramadge and W.M. Wonham, "The control of discrete event systems", *Proceedings of the IEEE*, vol. 77, no. 1, pp. 81-97, Jan., 1989.
- [7] M. Skubic and R.A. Volz, "Acquiring Robust, Force-Based Assembly Skills from Human Demonstration", *IEEE Transactions on Robotics and Automation*, to appear.
- [8] R.W. Brockett, "Hybrid models for motion control systems," in *Essays on Control: Perspectives in the Theory and Its Applications*, H.L. Trentelman and J.C. Willems, Eds., chapter 2, pp. 29--53. Birkhauser, Boston, MA, 1993.
- [9] 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.
- [10] P. Matsakis, J. M. Keller, L. Wendling, J. Marjamaa and O. Sjahputera, "Linguistic Description of Relative Positions in Images", *IEEE Trans. on Systems, Man and Cybernetics*, submitted.
- [11] J. M. Keller and P. Matsakis, "Aspects of High Level Computer Vision Using Fuzzy Sets", *Proceedings, 8th IEEE Int. Conf. on Fuzzy Systems*, Seoul, Korea, pp. 847-852, 1999.



1. **Object A** is to the left-front of the **Robot**.
2. **Object A** is mostly to the left of the **Robot** but somewhat forward.
3. **Object A** is to the left of the **Robot** but extends forward relative to the **Robot**.
4. **Object A** is to the left of the **Robot**.
5. *None*
6. **Object B** is mostly to the left of the **Robot** but somewhat to the rear.
7. **Object B** is behind-left of the **Robot**.
8. **Object B** is mostly behind the **Robot** but somewhat to the left.
9. *None*
10. **Object C** is in front of the **Robot**.
11. **Object C** is in front of the **Robot** but extends to the right relative to the **Robot**.

Figure 6. Representative spatial descriptions along the sketched robot trajectory for the PDA-generated map 1.

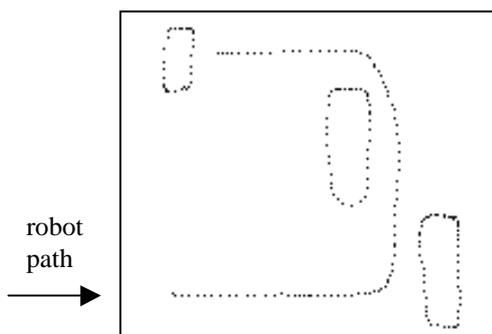
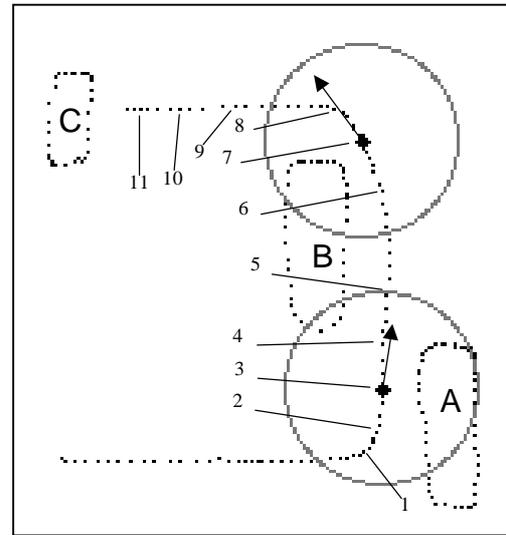


Figure 7. The sketched map used for the second experiment. The robot path starts from the bottom left.



1. **Object A** is in front of the **Robot** but extends to the right relative to the **Robot**.
2. **Object A** is to the right of the **Robot**.
3. **Object A** is to the right of the **Robot** but extends to the rear relative to the **Robot**.
4. **Object A** is mostly to the right of the **Robot** but somewhat to the rear.
Object B is to the left-front of the **Robot**.
5. **Object A** is mostly behind the **Robot** but somewhat to the right.
Object B is mostly to the left of the **Robot** but somewhat forward.
6. **Object B** is to the left of the **Robot** but extends to the rear relative to the **Robot**.
7. **Object B** is mostly to the left of the **Robot** but somewhat to the rear.
8. **Object B** is to the left of the **Robot** but extends to the rear relative to the **Robot**.
9. **Object B** is behind-left of the **Robot**.
10. **Object C** is in front of the **Robot**.
11. **Object C** is in front of the **Robot** but extends to the left relative to the **Robot**.

Figure 8. Representative spatial descriptions along the sketched robot trajectory for the PDA-generated map 2, showing the simultaneous detection of two different objects.