Beyond 2001: The Linguistic Spatial Odyssey

**JAMES M. KELLER, PASCAL MATSAKIS
AND MARJORIE SKUBIC**

*Computer Engineering and Computer Science Department
University of Missouri-Columbia
Columbia, MO 65211-2060
KellerJ@missouri.edu*

ABSTRACT

Why is it so hard to talk to a machine? If only we could communicate in a natural human language with robots, they would be so much more useful. Having machines that can reason spatially and receive and communicate such reasoning linguistically will extend their utility in many more scenarios that are dangerous, tedious, unhealthy, etc. Scene description, involving linguistic expressions of the spatial relationships between image objects, is a major goal of high-level computer vision. People have studied spatial relationships for several years. In a series of papers, we have introduced the use of histograms of forces to produce evidence for the description of relative position of objects in a digital image. There is a parameterized family of such histograms, for example, the histogram of constant forces (much like the earlier histogram of angles) and the histogram of gravitational forces that highlights areas that are close between the two objects. Utilizing the fuzzy directional membership information extracted from these histograms within fuzzy logic rule-based systems, we have produced high-level linguistic descriptions of natural scenes as viewed by an external observer. Additionally, we have begun to exploit the theoretical properties of the histograms to match images that may be the same scene viewed under different pose conditions. In fact, we can even recover estimates of the pose parameters. These linguistic descriptions have then been brought into an ego-centered viewpoint for application to robotics. We describe three initial activities here: production of linguistic scene description from a mobile robot standpoint, spatial language for human/robot communication, and understanding of a sketched route map for communicating navigation routes to robots. These efforts just scratch the surface of the potential applications and we end with future projections.

Keywords: Human/robot interaction, spatial relationships, linguistic scene description, force histograms, fuzzy sets, sketch understanding.

1. INTRODUCTION

Consider the really great computational personalities of science fiction: Robby the Robot, HAL, The Terminator, Commander Data, Andrew the Bicentennial Man to name but a few. Why can we relate more to them than, say, to the computer on the Enterprise or the robot from *The Day the Earth Stood Still*? Even though R2D2 is smarter, C-3PO's actions are easier to interpret. In our opinion, a large part of the reason lies in the (imagined) ability of these machines not only to reason about their environment but to discuss this reasoning with humans in a natural language or through other human-based media. In this chapter, we discuss efforts at the University of Missouri to move towards this goal of efficient linguistic communication between a computer (or a robot) and humans. We will not produce a survey of all research in this field (and apologize *a priori*), but provide our insights into solutions to particular sub-problems and to speculate on the future utility of linguistic communication.

Determination and utilization of spatial relationships among objects in an image has been an active area of research for many years [1-11]. In earlier work, Keller and Wang [10,12] used a fuzzy rule-base to generate linguistic description of relative position between two image objects, and ultimately, to produce a complete description of the scene. The fuzzy rule-base received confidence values of the four main directional relations (LEFT, ABOVE, RIGHT, BELOW) and SURROUND based on the histogram of angles [7,8]. Subsequently, Matsakis et al. [11,13] designed a system for spatial relationship estimation through an axiomatic framework for functions from which "histograms of forces" were generated to represent relative position between a pair of 2D image objects. By selecting particular functions, we can construct various histograms, ranging from the histogram of angles to a histogram of gravitational forces (section 2). In [13], we utilized the histogram of forces to generate numeric features from multiple force histograms that were then used to generate a linguistic description of a scene using a fuzzy rule-base. This approach encompassed the earlier paradigm and led to a richer language for scene description. Due to the complementary nature of the histograms of forces, it was even possible to construct a self-assessment measure for each linguistic description between a pair of image objects, as developed in section 3. Next, we used the properties of these histograms along with fuzzy similarity measures to match scenes while recovering the camera pose parameters [14,15] (section 4). We also applied our system for linguistic description generation to human-robot communication (section 5) [16-18] and sketched route map understanding (section 6) [19, 20], moving from external views of scenes to an ego-centered perspective.

2. FORCE HISTOGRAMS

The fuzzy relative position between 2D objects is often represented by a histogram of angles [7,8,10]. The histogram of angles associated with any pair (A,B) of crisp and digitized objects is a function Ang^{AB} from $\mathbf{R}$ into $\mathbf{N}$. For any direction θ, the value $Ang^{AB}(\theta)$ is the number of pixel pairs (p,q) belonging to $A \times B$ such that p is

in direction θ of q. In [11], Matsakis and Wendling introduced the notion of the histogram of forces. It generalizes and supersedes that of the histogram of angles. It ensures rapid processing of raster data as well as of vector data, and of crisp objects as well as of fuzzy objects. It also offers solid theoretical guarantees, and allows explicit accounting of metric information. The histogram of forces associated with (A,B) via F, or the F–histogram associated with (A,B), is a function F^{AB} from $\mathbf{R}$ into $\mathbf{R}_+$. Like Ang^{AB}, this function represents the relative position of A with regard to B. For any direction θ, the value $F^{AB}(\theta)$ is the total weight of the arguments that can be found in order to support the proposition "A is in direction θ of B." More precisely, it is the scalar resultant of elementary forces. These forces are exerted by the A points on those of B, and each tends to move B in direction θ, as depicted in figure 1. Actually, the letter F denotes a numerical function. Let r be a real. If the elementary forces are in inverse ratio to $d^{\,r}$, where d represents the distance between the points considered, then F is denoted F_r. For instance, the F function associated with the universal law of gravitation is F_2. The F_2–histogram and F_0–histogram (histogram of constant forces) have very different and very interesting characteristics. The latter, very similar to the histogram of angles, gives a global view of the situation. It considers the closest parts and the farthest parts of the objects equally, whereas F_2–histogram focuses on the closest parts.
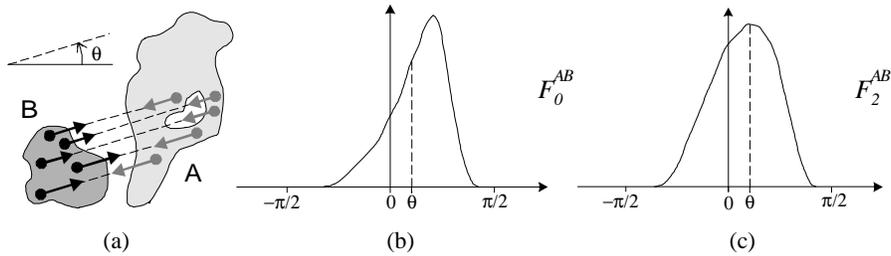

Figure 1. Force histograms. (a) $F^{AB}(\theta)$ is the scalar resultant of forces (black arrows). Each one tends to move B in direction θ. (b) The histogram of constant forces associated with (A,B). It represents the position of A relative to B. (c) The histogram of gravitational forces associated with (A,B). It is another representation of the relative position between A and B.

The forces can be classified in different types. Consider, for instance, the proposition "A is in direction 0 of B" (which will be read "A is to the RIGHT of B"). First, the set of directions is divided into four quadrants as shown in figure 2. The forces $F_r^{AB}(\theta)$ of the outer quadrants ($\theta \in [-\pi,-\pi/2] \cup [\pi/2,\pi]$) are elements which, to various degrees, weaken the proposition "A is to the RIGHT of B"; the forces of the inner quadrants ($\theta \in [-\pi/2,0] \cup [0,\pi/2]$) are elements which support the proposition. Some forces of the third quadrant are used to compensate—as much as possible—the contradictory forces of the fourth one. The proportion of these compensatory forces is defined by some angle α_+. Forces of the second quadrant are used in a similar way to compensate the contradictory forces of the first one.

The amount of these compensatory forces is defined by α_-. The remaining forces are called the effective forces. A threshold τ divide them into optimal and sub-optimal components. The optimal components support the idea that A is

"perfectly" to the right of *B*: whatever their direction, they are regarded as horizontal and pointing to the right. The "average" direction α_0 of the effective forces is then computed, in conformity with this agreement.
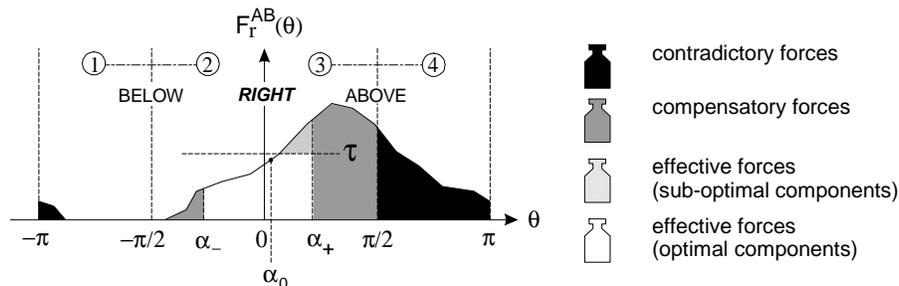

Figure 2. Force typology associated with the proposition "*A* is to the *RIGHT* of *B*."

3. LINGUISTIC SCENE DESCRIPTION

We produce a linguistic description of the relative position between any 2D objects *A* and *B* based on the sole primitive directional relationships: "to the right of," "above," "to the left of," and "below." First, eight values are extracted from the analysis of each histogram F_0^{AB} and F_2^{AB} relying upon the categories of forces described in section 2 (see [13] for a detailed exposition of the feature extraction). These values are: $a_r(\text{RIGHT})$, $b_r(\text{RIGHT})$, $a_r(\text{ABOVE})$, $b_r(\text{ABOVE})$, $a_r(\text{LEFT})$, $b_r(\text{LEFT})$, $a_r(\text{BELOW})$ and $b_r(\text{BELOW})$. They represent the "opinion" given by the considered histogram. For instance, according to F_2^{AB}, the degree of truth attached to the proposition "*A* is to the right of *B*" belongs to the interval $[a_2(\text{RIGHT}), b_2(\text{RIGHT})]$. Then, the two opinions—16 values—are combined using a heuristic scheme. We work on the principle that F_0^{AB} is never too optimistic, but is often too cautious. We attribute the previous drawback to the fact that F_0^{AB} only has a global view of the situation, and we correct it considering F_2^{AB}, which focuses on the closest parts of the objects. However, just because of this characteristic, F_2^{AB}'s opinion may be excessive: sometimes excessively pessimistic, and sometimes excessively optimistic. Four numeric and two symbolic features result from the heuristic combination. They feed a system of fuzzy rules that finally outputs the expected description.

In [13], we used LADAR (Laser Radar) range images of a power-plant at China Lake, CA provided by the Naval Air Warfare Center. They were processed by applying first a median filter, and then the pseudo-intensity filter $1/\sqrt{1+G_x^2+G_y^2}$, where G_x and G_y are the Sobel gradient magnitudes in a 3×3 window. Finally, the filtered images were segmented and labeled manually. Wang and Keller used the same real data to test a fuzzy rule-based approach for linguistic scene description [12]. Figure 3 shows a typical LADAR scene after preprocessing, followed by the results of segmentation and labeling.
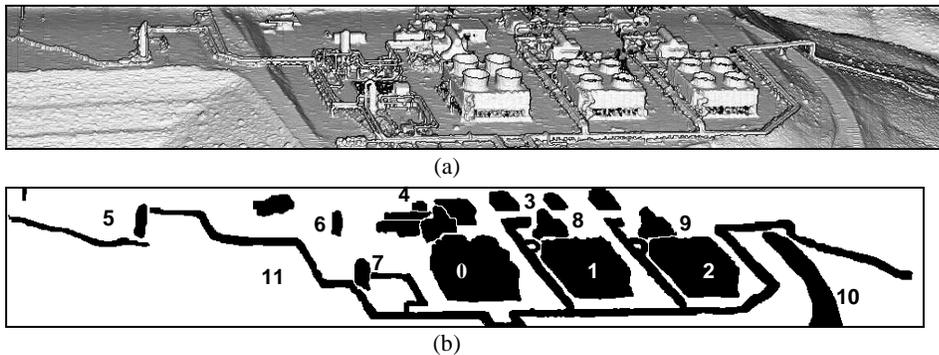
(a)

(b)

Figure 3. LADAR range image NAWC 20675; (a) after pseudo-intensity filtering, (b) after hand-segmentation and labeling.

As a simple representative example, consider the two configurations in figure 4. The resultant linguistic description of part (a) is: "The stackbuilding 1 is *perfectly to the right of* the reference stackbuilding 0, but *slightly shifted upward*. The description is *satisfactory*." and "The stackbuilding 2 is *perfectly to the right of* the reference stackbuilding 0. The description is *satisfactory*." whereas that of part (b) is: "The group 4 of storehouses is *loosely above-left of* the group of stackbuildings. The description is *satisfactory*." System details and many other cases are shown in [13].
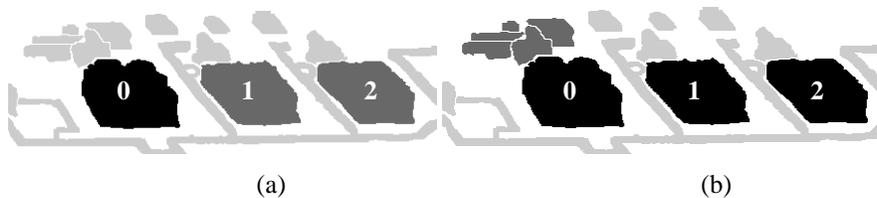

(a)   (b)

Figure 4. Two configurations from Figure 3. For each image, the reference object is in black, and the argument(s) in dark gray. The light gray objects are ignored.

4. SCENE MATCHING

In [14,15] we considered the problem of matching two views of the same scene from the information contained in the force histograms and LADAR range data. In this case, the views were reconstructed so that they were independent of the third dimension and the tilt (or declination) of the camera. In other words, the three-dimensional information contained in the range data was used to determine the declination angle, and to transform the segmented scene to a position as viewed from above. Matching was then performed by comparing the relationships between the objects in the transformed scenes. The force histograms were treated as fuzzy sets and similarity measures were employed to determine rotation and scale for each object pair match and to provide a ranking of the matches.
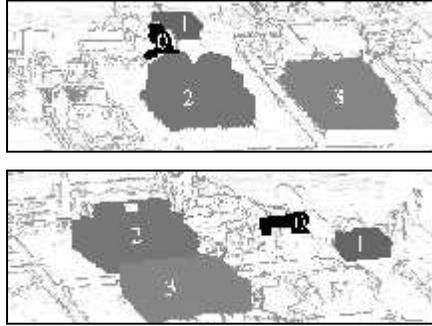
Figure 5. These two images are from the same scene shown from two different viewpoints.

Figure 5 depicts an example of this matching. Only the four labeled objects were considered. The shapes of the objects and the distance between a given pair in these scenes vary greatly. By "eliminating" the declination, these factors are ameliorated. The similarity measures from [15] are shown in Table 1. In that table, the symbol a_b corresponds to the force histogram of the object pair (a,b) in the top image (rows) and in the bottom image (columns).

TABLE 1
SIMILARITY MEASURES OF OBJECT PAIRS IN FIGURE 5

	0_1	0_2	0_3	1_2	1_3	2_3
0_1	0.484	0.484	0.925	0.736	0.928	0.363
0_2	0.52	0.525	0.808	0.817	0.824	0.386
0_3	0.457	0.451	0.896	0.689	0.86	0.345
1_2	0.329	0.322	0.699	0.514	0.665	0.255
1_3	0.308	0.3	0.644	0.479	0.613	0.242
2_3	0.733	0.701	0.384	0.491	0.407	0.696

These results are not sufficient to match object pairs, let alone the entire scene. In every row, except for 0_3, the highest similarity measure does not correspond to the true matching. Thus, it is necessary to use additional information, such as the rotational differences and the scaling ratios calculated from the histograms. There are 720 possible ways to assign object pair labels between the two images. Using all the information available (histogram similarity, recovered rotation angle similarity and estimated scale consistency), an overall matching degree is computed for each case. The highest degree was found to be the true matching, with a value of 0.997. The next closest matching swaps the relationships of objects 0 and 3, with objects 1 and 3, and only earns a matching degree of 0.976. The high value is because objects 0 and 1 are both about the same size and the relationships between them and object 3 only differ by a few degrees. Only 8 of the 720 matching possibilities achieved a matching degree above 0.9.

5. HUMAN-ROBOT DIALOG

We now move from external views of scenes to an ego-centered perspective. People often use spatial relationships in a conversation to describe their environment, e.g., "There is a pillar in front of me and a doorway behind it," and to give instructions such as "Go around the pillar and through the doorway." Recent cognitive models suggest that people use these types of relative spatial concepts to perform day-to-day navigation tasks and other spatial reasoning [21,22], which may explain the importance of spatial language and how it developed. We have initiated work, using the histograms of forces, to embed this capability into a mobile robot. Ego-centered descriptions are produced based on both direct sensor readings [16,17] and on sensor readings integrated over time into an occupancy grid map [18]. In figure 6, the robot and six segmented and labeled objects extracted from the current state of a sensor-based digital grid map are shown for scene 1. The robot heading specified by the arrow from the center (looking "towards" object 5). The fuzzy logic system using force histogram inputs produces the detailed and high-level descriptions shown in the textbox. The linguistic nature of this description could be used as a concise means of communication with a person in an interactive fashion. In fact, as demonstrated in [18] the robots seem human-like because we can interact with them in the same way we would interact with another person. This computation is fast enough on a PC so that it can be accomplished in a real-time mode, allowing for interactive conversations.
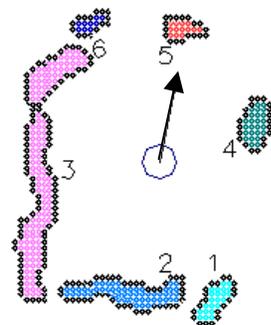
Scene 1

DETAILED SPATIAL DESCRIPTIONS for 6 OBJECTS IN SCENE 1:

Object number 1 is mostly behind me but somewhat to the right (the description is satisfactory). The object is very close.

Object number 2 is behind me (the description is satisfactory) The object is very close.

Object number 3 is to the left of me but extends to the rear relative to me (the description is satisfactory). The object is very close.

Object number 4 is mostly to the right of me but somewhat forward (the description is satisfactory). The object is very close.

Object number 5 is in front of me (the description is satisfactory). The object is very close.

Object number 6 is to the left-front of me (the description is satisfactory). The object is close.

HIGH-LEVEL DESCRIPTION:

There are objects in front of me and behind me.

Object number 3 is to the left of me.

Object number 4 is mostly to the right of me.

Figure 6. Robot scene 1. Objects were extracted from an actual robot-sensed grid map and the detailed and high-level descriptions were produced by the fuzzy logic system. Robot heading is shown by the arrow.

Figure 7 contains a second example of the descriptive power of this robot language. The environment is the same as in scene 1 but the robot is headed in a different direction. Note in particular that the high-level description is very concise. From it, the robot or a person controlling the robot could deduce easily that there is an opening to the rear.
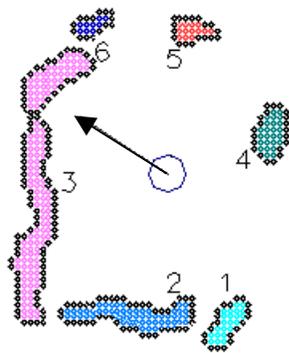
DETAILED SPATIAL DESCRIPTIONS for 6 OBJECTS IN SCENE 2:

Object number 1 is mostly behind me but somewhat to the left (the description is satisfactory). The object is very close.

Object number 2 is mostly to the left of me but extends to the rear relative to me (the description is satisfactory). The object is very close.

Object number 3 is in front of me but extends to the left relative to me (the description is satisfactory). The object is very close.

Object number 4 is mostly to the right of me but somewhat to the rear (the description is satisfactory). The object is very close.

Object number 5 is mostly to the right of me but somewhat forward (the description is satisfactory). The object is very close.

Object number 6 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, on my rear left and on my right.

Scene 2

Figure 7. Robot scene 2. The basic scene is the same as in figure 6. The robot heading, and hence, the generated detailed and high-level descriptions are different.

Of course, it is difficult to get a real feeling for where the robot is from this information alone since the objects can have an arbitrary numeric labeling. Since the goal of this particular project is to have the robot interact with a person within a given environment over time, some objects in the grid map will always be present and could be labeled by the robot or the human. Figure 8 shows an example where on one time snapshot, the pillar was identified and labeled by a human. On the current frame, the robot sensor-based grid map generates object 3 that overlaps the pillar object. From a simple algorithm that matches linguistic spatial descriptions and the distance values, the final merged explanation equates the pillar and object 3. This capability enhances the information content of a two-way communication. Much more work is needed to allow the robot to perform object recognition. Here, though, the advantage is that the environment is reasonably stable, and so the number and type of objects can be restricted. Additionally, the ability for the human to easily interact with the robot to specify

names for particular objects lessens the burden on fully autonomous recognition in uncertain environments. Recognition is difficult for robots but usually easy for humans, so the human provides the difficult recognition for the robot and names landmarks that can be used later as references in the dialog.

DETAILED SPATIAL DESCRIPTION FOR OBJECT 3:

Object number 3 is mostly in front of me but somewhat to the left (the description is satisfactory). The object is close.

HIGH-LEVEL DESCRIPTION OF SCENE 3:

I am surrounded from the rear. Object number 2 is to the left of me. Object number 3 is mostly in front of me.

DESCRIPTION OF PERSISTENT OBJECT, "PILLAR" (AS DEFINED BY USER ON EARLIER FRAME):

The pillar is mostly in front of me but somewhat to the left (the description is satisfactory). It is close.

MERGED HIGH-LEVEL DESCRIPTION:

I am surrounded from the rear. Object number 2 is to the left of me. The pillar is mostly in front of me
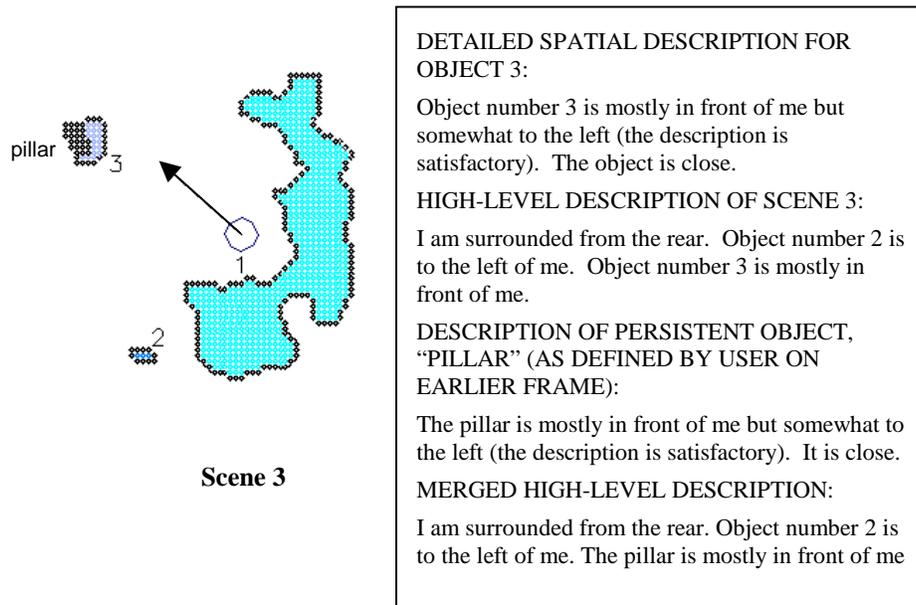
Scene 3

Figure 8. Demonstration of a merged high-level description when a "persistent" object is given a label by a human.

With this spatial information and linguistic descriptions, in [18] a dialog using a spatial language was established with the robot. As an example, consider the scene in figure 7 for the dialog below. This figure is similar to those shown in [18] and utilizes the multimodal robot interface developed at the Naval Research Laboratory [23]. The robot responds using synthesized speech output.

User: "How many objects do you see?"
Robot: "I am sensing 6 objects."
User: "What objects do you see?"
Robot: "There are objects in front of me, on my rear left and on my right."
User: "Where is the nearest object on your right?"
Robot: "Object number 4 is mostly to the right of me but somewhat to the rear. It is very close."

Motion directives can also be given using spatial language, as shown below:

User: "Go to the nearest object in front of you."
Robot: "Looking for the nearest object in front of me.
           I found it. Going to the object."

6. SKETCHED ROUTE MAP UNDERSTANDING

When someone invites you to his house and you're not quite sure where it is, often times you ask your friend to draw a map. Certainly these maps are not precise and to scale, but if there are sufficient landmarks, we normally don't have too much trouble navigating to the destination. A human-friendly method to communicate qualitative route information to a robot would be the same: hand-drawn route maps in which the user sketches an approximate representation of the environment and then draws the desired robot path within that scene. The communication media can be a PDA, such as a PalmPilot, as in [19,20] and the information extracted from a sketch includes spatial information about the map and a qualitative path through the indicated landmarks. The stylus interface of the PDA allows the person to sketch a map much as he would on paper for a friend. The PDA captures the string of (x,y) coordinates sketched on the screen which forms a digital representation suitable for processing. This information is used to build a task representation for the robot, which operates as a semi-autonomous vehicle. In this approach, all information and the resultant representation is based on robot sensing and *relative* instead of *absolute* position.

Figure 9 displays 3 particular locations on a path that was drawn by a user along with the representation of the principle landmarks. The circle around the robot symbol delimits the sensed region at each particular instance.
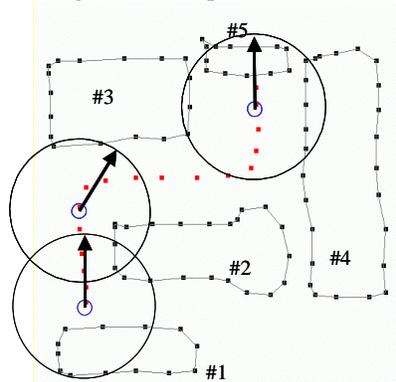

Figure 9. The PDA sketch. The original sketch with an overlay of the robot's sensory radius for several points along the route.

In addition to extracting spatial information with respect to the labeled objects as before, we also extract the movement of the robot along the sketched path. The computation of the robot's heading provides an instantaneous orientation. However, we also want to track the change in orientation over time and compute what would correspond to robot commands, e.g., move forward, turn right, make a "hard" left. The turning rate is determined by computing the change in instantaneous heading between two adjacent route points and normalizing by the distance between the points. A positive rate means a turn to the left, and a negative rate means a turn to the right.

In figure 10, the main direction of each object is plotted for the route steps in which the object is "in view"; labels of the corresponding directions are displayed on the graph to show the symbolic connection. The normalized turning rate that tracks the robot movement along the trajectory is also shown. The turning rate, although not translated into discrete robot commands, shows the general trend in the robot movement along the route and the correlation with relative positions of the environment landmarks. At the beginning of the route, when object #1 is behind the robot, the robot's movement is generally straight ahead (slightly to the left). When object #3 is in view, the robot turns to the right until the object is mostly on the left. When object #4 is in view to the front, the robot turns left and stops when object #5 is in front and very close. In this way, we can extract the key points along the route where a change in direction is made.
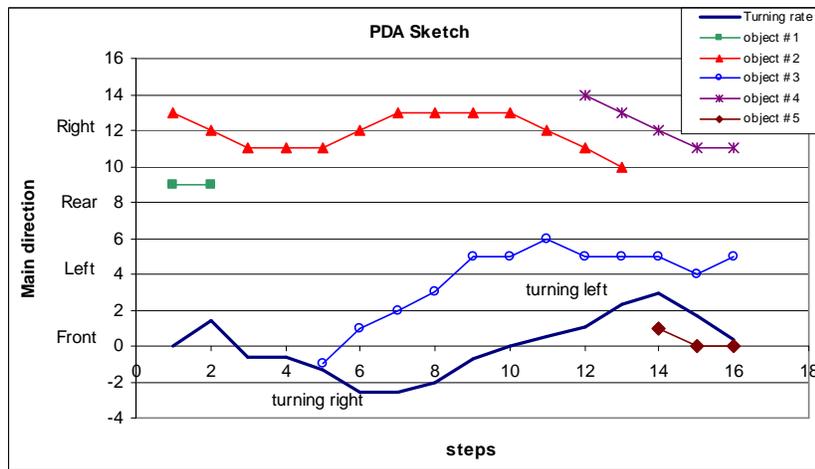

Figure 10. Normalized turning rate of the robot along the sketched route with the corresponding discrete main directions of the objects.

7. THE FUTURE

The ability of interact with computers and particularly with robots in natural human ways will make these devices immeasurably more useful. The research towards that goal presented in this chapter we think is pretty impressive. But, we've just scratched the surface and much more work needs to be done to realize the goal. Let's speculate for a while. One of the tools that will make a major impact on getting robots to behave in more human-like fashion is recognition technology. As defined by L. Zadeh [24], recognition technology refers to current or future systems that have the potential to provide a "quantum jump in the capabilities of today's recognition systems." Zadeh claims that this can occur as a result of three converging developments: (a) major advances in sensor technology; (b) major advances in sensor data processing technology; and (c) the use of soft computing techniques to infer a conclusion from observed data. Exploring new soft computing recognition techniques coupled closely with advances in sensors

and signal/image processing will significantly enhance the ability of robots to become increasingly autonomous. Handling the uncertainty and ambiguity in object recognition is a requirement to intelligent-like behavior. From the current trends, we see this aspect of the bigger problem to steadily improve.

The key ingredient, though, in any application is the ability of the robot or the computer system to interact with people. So, what's coming down the road? In the short term, we think that significant advances will be made in fairly (though not completely) structured environments. Besides getting better with the communication as indicated in this chapter, robots will be able to interpret dynamic behavioral commands, such as "Go to the right of object #2," or "Continue moving generally to the south until you pass behind object #3." Figure 11 shows some preliminary work using histograms of forces on defining positions that could be considered right of, left of, in front of, and behind the three objects in scene 4. The vertices of the quadrilaterals mark these potential spots assuming that the robot turns toward each object along its main direction.
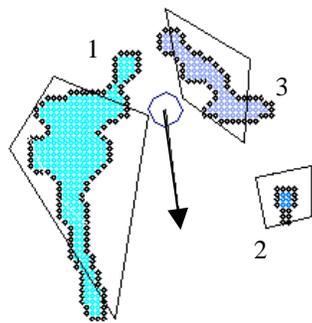
DETAILED SPATIAL DESCRIPTIONS for 3 OBJECTS IN SCENE 4:

Object number 1 is to the right of me (the description is satisfactory). The object is very close.

Object number 2 is to the left-front of me (the description is satisfactory). The object is close.

Object number 3 is loosely behind-left of me (the description is satisfactory). The object is very close.

Scene 4

Figure 11. Robot scene depicting potential places for the robot to move right of, left of, in front of, and behind each object. Places were determined by the relative position information, and assumes a perspective in which the robot faces the object along its "main direction."

In the realm of automated surveillance, we see big advances coming in the way of temporal processing. For example, systems that can report activities like "Object 1 is moving mostly north but a little east" or "The car is executing a left turn" are being investigated. In this case, we envision systems that can perform such analysis based only on the linguistic relative positions of objects on a sequence of frames without resorting directly to the image data. These applications will expand both in scope and complexity. Full 3-dimensional spatial reasoning is also on the horizon. Besides the obvious use in medical and biological imagery, coupling (almost) 3-D reasoning to range images will provide better static and temporal descriptions of natural scenes.

Down the road a bit, we foresee greater application of robotics to the increasingly important task of search and rescue. Here, the environment is very uncertain and perhaps dynamically changing. It will be essential for our machines not only to be able to react to the environment, but to communicate with people and to receive and interpret possibly complicated linguistic commands. Flexibility and robustness will be at a premium. Not only research, but intense engineering development will be required to make the machines reliable enough to bet human life on. On the lighter side, personal assistant robots that can recognize particular people and interact with them would be a great benefit to, say, bed-ridden children or adults.

Is the Bicentennial Man on the horizon? Probably not, but the future of linguistic human/machine interaction is bright. Let's communicate.

ACKNOWLEDGMENTS

REFERENCES

1. P. Winston and "Learning Structural Descriptions from Examples," in P. Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill , New York, 1975.

2. J. Freeman, "The Modeling of Spatial Relations," *Computer Graphics and Image Processing*, vol. 4, pp. 156-171, 1975.

3. R. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial Intelligence*, vol. 17, pp. 285-348, 1981.

4. S. Dutta, "Approximate Spatial Reasoning: Integrating Qualitative and Quantitative Constraints," *Int. J. of Approximate Reasoning*, vol. 5, pp. 307-331, 1991.

5. E. Walker, M. Herman and T. Kanade, "A Framework for Representing and Reasoning about Three-Dimensional Objects in Vision," *Artificial Intelligence*, vol. 9, no. 2, pp. 47-58, 1988.

6. R. Antony, "A Hybrid Spatial/Object-Oriented DBMS to Support Automated Spatial, Hierarchical, and Temporal Reasoning," in S. Chen(ed.), *Advances in Spatial Reasoning*, *Ablex Publishing*, Norwood, NJ, vol. 1, pp. 63-132, 1990.

7. R. Krishnapuram, J. M. Keller and Y. Ma, "Quantitative Analysis of Properties and Spatial Relations of Fuzzy Image Regions," *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 3, pp. 222-233, 1993.

8. K. Miyajima and A. Ralescu, "Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations," *Fuzzy Sets and Systems*, vol. 65, no. 2/3, pp. 225-236, 1994.

9. J. M. Keller and X. Wang, "Comparison of Spatial Relation Definitions in Computer Vision," *ISUMA-NAFIPS'95*, College Park MD, pp. 679-684, 1995.

10. X. Wang and J. M. Keller, "Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches," *Fuzzy Sets and Systems*, vol. 101, no. 1, pp. 5-20, 1999.

11. 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.

12. J. M. Keller and X. Wang, "A Fuzzy Rule-based Approach for Scene Description Involving Spatial Relationships," *Computer Vision and Image Understanding*, vol. 80, pp. 21-41, 2000.

13. P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera, "Linguistic Description of Relative Positions of Objects in Images," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 31, No. 4, 2001, pp. 573-588.

14. O. Sjahputera, J.M. Keller, P. Matsakis, P. Gader, and J. Marjamaa, "Histogram-Based Scene Matching Measures," *Proceedings of NAFIPS'2000*, Atlanta, GA, July, 2000, pp. 392-396.

15. J. Marjamaa, O. Sjahputera, J. Keller, and P. Matsakis, "Fuzzy Scene Matching in LADAR Imagery," *Proceedings, Tenth IEEE International Conference on Fuzzy Systems*, Melbourne, Australia, December, 2001.

16. M. Skubic, G. Chronis, P. Matsakis and J. Keller, "Generating Linguistic Spatial Descriptions from Sonar Readings Using the Histogram of Forces," in *Proceedings of the IEEE 2001 International Conference on Robotics and Automation*, vol. 1, pp. 485-490, Seoul, Korea, May, 2001.

17. M. Skubic, P. Matsakis and J. Keller. Spatial Relations for Tactical Robot Navigation, in *Proceedings of the SPIE Conference Unmanned Ground Vehicle Technology III*, April, 2001, Orlando, FL.

18. Skubic M, Perzanowski D, Schultz A and Adams W, "Using Spatial Language in a Human-Robot Dialog," accepted to the IEEE 2002 International Conference on Robotics and Automation, Washington, D.C., May, 2002.

19. M. Skubic, P. Matsakis, B. Forrester and G. Chronis, "Extracting Navigation States from a Hand-Drawn Map," in *Proceedings of the IEEE 2001 International Conference on Robotics and Automation*, vol. 1, pp. 259-264, Seoul, Korea, May, 2001.

20. M. Skubic, S. Blisard, A. Carle, P. Matsakis, "Hand-Drawn Maps for Robot Navigation," accepted for the AAAI Spring Symposium, Sketch Understanding Session, March, 2002.

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

22. C. Schunn, T. Harrison. Personal communication. 2001.

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

24. L. Zadeh, "Soft Computing, Fuzzy Logic and Recognition Technology," *Proceedings, IEEE International Conference on Fuzzy Systems, Anchorage,* AK, May, 1998, pp. 1678-1679.