Generating Linguistic Spatial Descriptions from Sonar Readings Using the Histogram of Forces

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

Abstract

In this paper, we show how linguistic expressions can be generated to describe the spatial relations between a mobile robot and its environment, using readings from a ring of sonar sensors. Our work is motivated by the study of human-robot communication for non-expert users. The eventual goal is to use these linguistic expressions for navigation of the mobile robot in an unknown environment, where the expressions represent the qualitative state of the robot with respect to its environment, in terms that are easily understood by human users. In the paper, we describe the histogram of forces and its application to sonar sensors on a mobile robot. Several environment examples are also included with the generated linguistic descriptions.

1. Introduction

Our work is motivated by the study of human-robot interaction and, in particular, the investigation of human-robot communication. The ultimate goal is to provide easy and intuitive interaction by naïve users, so that they can guide, control, and/or program a robot to perform some purposeful task. We consider the communication between the human user and the robot to be crucial to intuitive interaction by users that are not robotics experts. We further argue that good communications is essential both from the human to the robot (to command the robot to perform purposeful tasks) and also from the robot to the human (so that the user can monitor the robot's current state or condition). See also [1] and [2] for examples and further motivation on task-oriented dialogues between a robot and a human user.

In this paper, we show how linguistic expressions can be generated to describe the spatial relations between a mobile robot and its environment, using readings from a ring of sonar sensors. The eventual goal is to use these linguistic descriptions for navigation of the mobile robot in an unstructured, unknown, and possibly dynamic environment. We are not attempting to build an exact model of the environment, nor to generate a quantitative map. However, we do want to generate linguistic descriptions that represent the qualitative state of the robot with respect to its environment, in terms that are easily understood by human users.

The linguistic spatial descriptions provide a symbolic link between the robot and a human user, thus comprising a navigation language for human-robot interaction. The linguistic expressions can be used for two-way communications with the robot. First, in robot-to-human communication, they provide a qualitative description of the robot's current state (e.g., *there is an object to the left*, or *there is an object to the right front*).

Second, in human-to-robot communication, the human can command the robot to perform navigation behaviors based on the spatial relations (e.g., *while there is an object on the left, move forward,* or *if there is an object on the right front, turn left*, or even a high-level and very human-like directive such as *turn left at the second intersection*). A task can be represented and described as a sequence of qualitative "states" based on spatial relations, each state 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.

To accomplish both cases of communication, the robot must be able to recognize its state in terms of egocentric spatial relations between itself and objects in its environment, and it must be able to generate a linguistic description of the spatial relations. The main focus of this paper is the creation of these linguistic spatial descriptions from a ring of sonar sensors.

The idea of using linguistic spatial expressions to communicate with a semi-autonomous robot has been proposed previously. Gribble *et al* use the framework of the Spatial Semantic Hierarchy for an intelligent wheelchair [3]. Perzanowski *et al* use a combination of gestures and linguistic directives such as "go over there" [4]. Shibata *et al* use positional relations to overcome ambiguities in recognition of landmarks [5]. In [6], Stopp *et al* use spatial expressions to communicate with a 2-arm mobile robot performing assembly tasks. Spatial relations are used as a means of identifying an object in a geometric model. That is, the robot has a model of its environment, and the user selects an object from the model using relational spatial expressions.

The work presented here is an extension of spatial analysis previously applied to image analysis. Background material on the spatial analysis algorithms is included in Section 2. In Section 3, we show how the robot's sonar readings can be used to generate inputs for the spatial analysis algorithms. Specific test cases are shown in Section 4 along with a discussion of future work. Concluding remarks are found in Section 5. The interested reader is also referred to a companion paper on using spatial analysis to extract navigation states from a hand-drawn map [7].

2. Background on Spatial Relations

Freeman [8] proposed that the relative position of two objects be described in terms of spatial relationships (such as “above”, “surrounds”, “includes”, etc.). He also proposed that fuzzy relations be used, because “all-or-nothing” standard mathematical relations are clearly not suited to models of spatial relationships. Moreover, “although the human way of reasoning can deal with qualitative information, computational approaches of spatial reasoning and object recognition can benefit from more quantitative measures” [9]. By introducing the notion of the histogram of angles, Miyajima and Ralescu [10] developed the idea that the relative position between two objects can have a representation of its own and can thus be described in terms other than spatial relationships. However, the representation proposed shows several weaknesses (*e.g.*, requirement for raster data, long processing times, anisotropy).

In [11][12], Matsakis and Wendling introduced the histogram of forces. Contrary to the angle histogram, it ensures processing of raster data as well as of vector data. Moreover, it 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, the histogram of forces also allows for a low-computational handling of heading changes in the robot’s orientation and also makes it easy to switch between a world view and an egocentric robot view.

2.1. The Histogram of Forces

The relative position of a 2D object A with regard to another object B is represented by a function F^{AB} from $\mathbb{R}$ into $\mathbb{R}_+$. 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 points of A on those of B, and each tends to move B in direction θ (Fig. 1). F^{AB} is called the *histogram of forces associated with* (A,B) *via* F, or the F–*histogram associated with* (A,B). The object A is the *argument*, and the object B the *referent*. Note that throughout this paper, the referent is always the robot. 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 by F_r. The F_0–histogram (histogram of constant forces) and F_2–histogram (histogram of gravitational forces) have very different and very interesting characteristics. The former coincides with the angle histogram—without its weaknesses—and provides a global view of the situation. It considers the closest parts and the farthest parts of the objects equally, whereas the F_2–histogram focuses on the closest parts. Details can be found in [11][12].

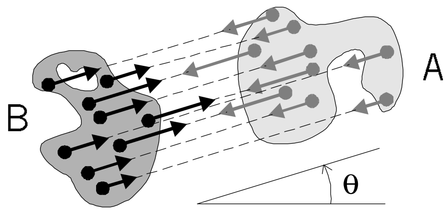


Figure 1. Computation of $F^{AB}(\theta)$. It is the scalar resultant of forces (black arrows). Each one tends to move B in direction θ.

2.2. Handling of Vector Data

In previous work, we generated the F_0 and F_2 histograms using raster image data. In this paper, we present the first application of histograms that uses vector data, *i.e.*, a boundary representation based on the objects’ vertices.

In practice, the F-histogram associated with a pair (A,B) of objects is represented by a limited number of values (*i.e.*, the set of directions θ is made discrete). For any θ considered, the objects are partitioned by sorting both A and B vertices, following direction $\theta+\pi/2$. The computation of F^{AB} is of complexity $O(n \log(n))$, where n denotes the total number of vertices. It is translated into a set of assessments of predetermined algebraic expressions. Each assessment corresponds to the process of a pair of trapezoids. In the case illustrated by Figure 2, the scalar resultant of the forces represented by black arrows is Γ_0 for constant forces and is Γ_2 for gravitational forces:

$$\Gamma_0 = \varepsilon[(x_1+x_2)(z_1+z_2)+x_1z_1+x_2z_2] / [6\cos^2(\theta)]$$
$$\Gamma_2 = \varepsilon[f(x_1+y_1,x_2+y_2)-f(y_1,y_2)+f(y_1+z_1,y_2+z_2) - f(x_1+y_1+z_1,x_2+y_2+z_2)]$$

where f denotes the function defined by:

$$\forall(r,s)\in \mathbb{R}^*_+\times\mathbb{R}^*_+,\ r\neq s \Rightarrow f(r,s) = [s\ln(s)-r\ln(r)]/(s-r)$$
$$\text{and}\quad \forall r\in \mathbb{R}^*_+,\ f(r,r) = \lim_{s\to r} f(r,s) = 1+\ln(r)$$

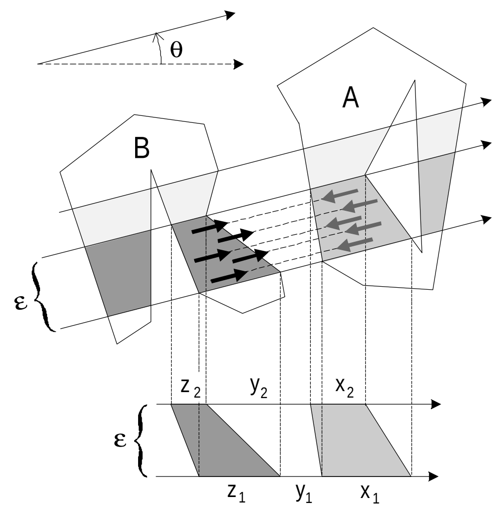


Figure 2. The evaluation of $F^{AB}(\theta)$ is based on the partitioning of the objects.

2.3. Linguistic Description of Relative Positions

In [13][14], Matsakis *et al.* present a system that produces linguistic spatial descriptions. The description of the relative position between any 2D objects A and B relies on the sole primitive directional relationships: "to the right of", "above", "to the left of" and "below" (imagine that the objects are drawn on a vertical surface). It is generated from F_0^{AB} (the histogram of constant forces associated with (A,B)) and F_2^{AB} (the histogram of gravitational forces). First, eight values are extracted from the analysis of each histogram: a_r (RIGHT), b_r (RIGHT), a_r (ABOVE), b_r (ABOVE), a_r (LEFT), b_r (LEFT), a_r (BELOW) and b_r (BELOW). They represent the "opinion" given by the considered histogram (*i.e.*, F_0^{AB} if r is 0, and F_2^{AB} if it is 2). For instance, according to F_0^{AB} the degree of truth of the proposition "A is to the right of B" is a_0(RIGHT). This value is a real number greater than or equal to 0 (proposition completely false) and less than or equal to 1 (proposition completely true). Moreover, according to F_0^{AB} the maximum degree of truth that can reasonably be attached to the proposition (say, by another source of information) is b_0(RIGHT) (which belongs to the interval [a_0(RIGHT),1]). F_0^{AB} and F_2^{AB}'s opinions (*i.e.*, the sixteen values) are then combined. Four numeric and two symbolic features result from this combination. They feed a system of 27 fuzzy rules and meta-rules that outputs the expected linguistic description. The system handles a set of 16 adverbs (like "mostly", "perfectly", etc.) which are stored in a dictionary, with other terms, and can be tailored to individual users. A description is generally composed of three parts. The first part involves the primary direction (*e.g.*, "A is mostly to the right of B"). The second part supplements the description and involves a secondary direction (*e.g.*, "but somewhat above"). The third part indicates to what extent the four primitive directional relationships are suited to describing the relative position of the objects (*e.g.*, "the description is satisfactory"). In other words, it indicates to what extent it is necessary to turn or not to other spatial relations (*e.g.*, "surrounds").

3. Egocentric Spatial Relations from Sonar Readings

In this section, we describe the application of the F_0 and F_2 histograms for extracting spatial relations from the sonar ring of a mobile robot. In our work, we have used a Nomad 200 robot with 16 sonar sensors evenly distributed along its circumference. The sensors' readings are used to build an approximate representation of the objects surrounding the robot. The vertices of each object are extracted and used to build the F_0 and F_2 histograms, as described in Section 2.2, which are then used to generate linguistic descriptions of relative positions between the robot and the environment objects (see Figure 3).

The first step in recognizing spatial relations from sonar readings is to build objects around the robot from the sonar readings. Let us consider a simple case of the robot and a single obstacle, shown in Figure 4. The sonar sensor S returns a range value (which is less than the maximum), indicating that an obstacle has been detected. In the case of Figure 4, all sonar sensors except S return the maximum value, which means that no other obstacle was detected. In this case, a single object is plotted as a trapezoid in the center of cone S. The depth of the obstacle cannot be determined from the sonar reading; thus, we use a constant arbitrary depth when building objects. We also represent the cylindrical robot as a rectangular object, because it is easier to process using vector data, since there are only 4 vertices in a rectangle. The bounding rectangle we build around the robot is also shown in Figure 4.

In the case of multiple sonar returns, we examine the sonar readings that are adjacent to each other. There is a question on whether adjacent sonar readings are from a single obstacle or multiple obstacles. Our solution to this issue is to determine if the robot can fit between the points of two adjacent sonar returns. If the robot cannot fit between two returns, then we consider these returns to be from the same object. Even if there are actually two objects, they may be considered as one for robot navigation purposes. In the case that the distance between the two points of the sonar returns is big enough to allow the robot to travel through, we consider separate objects. To form objects from multiple sonar returns we join the centers of the corresponding sonar cones.

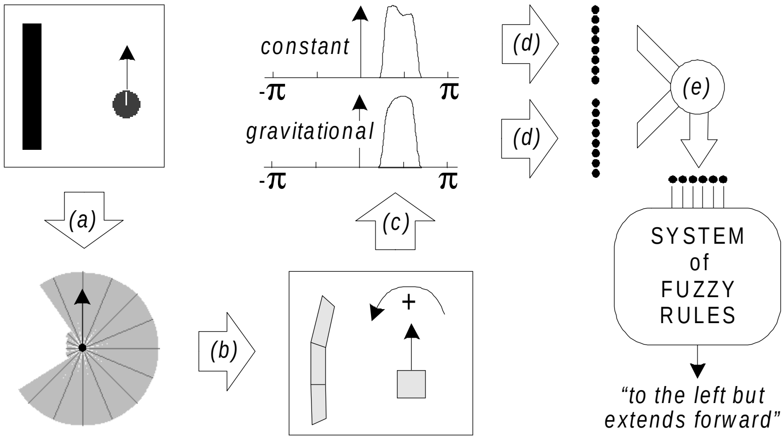


Figure 3. Synoptic diagram. (a) Sonar readings. (b) Construction of the polygonal objects. (c) Computation of the histograms of forces. (d) Extraction of numeric features. (e) Fusion of information.

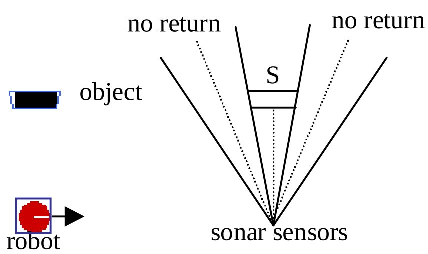


Figure 4. A single object is formed from a single sonar reading.

The distance we compute to determine if two adjacent sonar returns are "close" or not can be expressed by the following formula (distance between two points in polar coordinates):

$$\sqrt{s_1^2 + s_2^2 - 2s_1s_2\cos(2\pi / c)}$$

where: s_1 is the return of sonar sensor S_1,
s_2 is the return of sonar S_2, adjacent to S_1,
c is a constant that determines the angle between the two sonar sensors S_1 and S_2.

For c = 16, the angle between the two sonar sensors is set to the real angle between them ($2\pi/16$), and the formula returns the exact distance between the points of the two sonar returns. However, for our application we used c = 24, for which the distance computed between the points of the adjacent sonar readings is shorter than the actual one.

This way, when the robot diameter is compared to the distance between two obstacles, the distance will be big enough for the robot to easily travel between the obstacles. Thus, we allow extra clearance to make sure that the robot can easily fit between two obstacles.

For example, consider the obstacle in Figure 5. Since the obstacle is relatively far from the robot, the distance between the sonar returns is rather big, and we cannot determine whether the obstacle continues between the three sonar readings, or we have three different obstacles. In this case, we plot three different objects until the robot gets closer to the obstacle and we have a better resolution of the obstacle, since more sensors would detect its presence. In the same figure we show the distance computed for c = 16, which is the distance between A and B, and for c = 24, which is the distance between C and D.

In Figure 6, we show the same obstacle at a closer distance to the robot. There are five adjacent sonar sensors that have returns from the obstacle in this case. The distance measure determines that all sonar returns are close together, for the object to be considered as one.

After building the objects around the robot based on the sonar sensor readings, we represent the relative position between each object and the robot by the histograms of constant and gravitational forces associated with the robot/object pair, as described in Section 2. We then generate an egocentric linguistic description, *i.e.*, from the robot's point of view. Thus, the descriptions also depend on the robot's orientation or heading. A change in robot heading is easily accomplished by shifting the histogram along its horizontal axis. In the next section we show some test cases that illustrate the function of the approach.

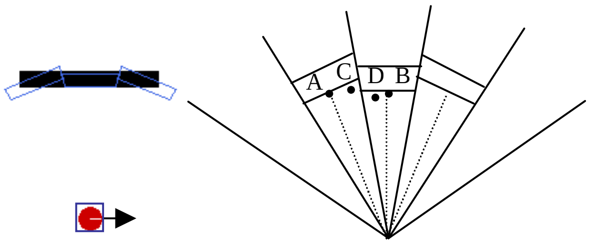


Figure 5. Three different objects are formed from 3 different sonar readings, if the readings are not "close" enough, according to the distance measure.

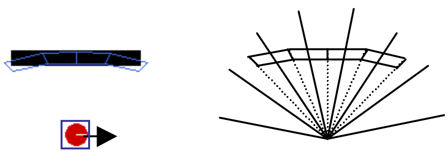
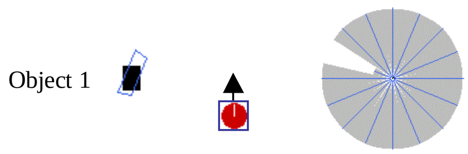


Figure 6. A single object is formed from 5 different sonar readings, if the readings are "close" enough.

4. Experiments and Discussion

The experiments included in this section were generated using the Nomad simulator. The program ran on the simulator at real-time speed. Processing of all obstacles, plotting of objects, processing of histograms and linguistic description generation is done faster than the robot can move, so there are no "delayed" results.

A simple case that demonstrates the functionality is shown in Figure 7. The sonar sensor readings are displayed on the right, the robot is shown as a circular model and an obstacle is drawn as a solid rectangle. For illustration, the software plots a hollow trapezoid based on the sonar readings, which should roughly coincide with the real obstacle, and it also plots the bounding rectangle that represents the robot. The software outputs the linguistic description, after executing the spatial analysis algorithm for all generated objects with respect to the robot. As described in Section 2, the linguistic expressions are generated in a three-part form: (1) "Object 1 is mostly to the left of the robot" (the primary direction), (2) "but somewhat forward" (the secondary direction), and (3) "the description is satisfactory" (the assessment indicating an adequate description).

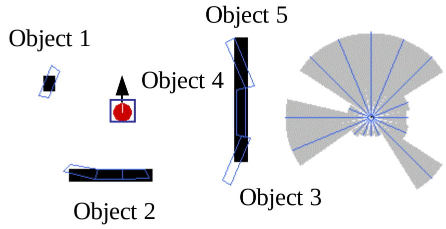


"*Object 1* is mostly to the left of the robot
but somewhat forward (the description is satisfactory)"*

Figure 7. The robot detects one obstacle. The sonar sensor readings are shown on the right. The generated linguistic expression is shown in italics.

In Figure 8, we show a more complex case. Object 1 from Figure 7 remains at the same position. A new obstacle is introduced behind the robot, which is recognized as a single object (Object 2). The obstacle to the right of the robot however, is plotted as three different objects. Since there are only three sonar readings from the right obstacle, and they are far apart according to the distance measure, the readings may not be from a single obstacle. Hence, three different obstacles are plotted. If more detail is needed, the robot may approach these three plotted objects to the right, to get a better resolution from more sonar sensors. This action may indeed reveal a passage through two of the three plotted objects or, if all sensors get returns that are close according to the distance measure, the three objects will prove to be the same one. Figure 8 shows the linguistic description generated for each object detected; in all cases, the assessment shows an adequate description.

Figure 9 shows the detection of two objects. The two obstacles to the left of the robot are so close together, that the robot cannot travel through them. Therefore, for navigation purposes these two obstacles are considered to be one object. Figure 9 shows the description generated, including a satisfactory assessment.



*"**Object 1** is mostly to the left of the Robot
but somewhat forward (the description is satisfactory)"*

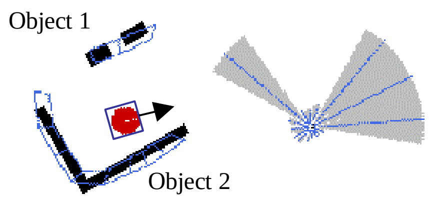
*"**Object 2** is behind the Robot but extends to the left
relative to the Robot (the description is satisfactory)"*

*"**Object 3** is mostly to the right of the Robot
but somewhat to the rear (the description is satisfactory)"*

*"**Object 4** is to the right of the Robot
(the description is satisfactory)"*

*"**Object 5** is mostly to the right of the Robot
but somewhat forward (the description is satisfactory)"*

Figure 8. The robot detects 5 obstacles.



*"**Object 1** is to the left of the Robot
(the description is satisfactory)"*

*"**Object 2** is loosely to the right of the Robot
and extends to the rear relative to the Robot
(the description is rather satisfactory)"*

Figure 9. The robot detects 2 obstacles.

The L-shaped object behind and to the right of the robot is an example of a rather satisfactory (*i.e.*, less satisfactory) linguistic description. The algorithm determines that for such a relative position there is not a really good description in terms of the four primitive directions only. It introduces the term "loosely" together with the classification of the whole description as "rather satisfactory" as opposed to "satisfactory" in all previous examples. This assessment indicates that we may need additional spatial relations (like "surrounds").

In the future, we plan to use more spatial relations for descriptions to include situations such as the one of Figure 9. A higher level of processing may generate such descriptions after considering the outputs of our current algorithm. For example, if there is an object to the right and an object to the left of the robot, then the robot is between the two objects.

We are also planning to introduce descriptions that indicate distance, in addition to relative position, such as *close* or *far*. These descriptions may be generated after processing the distance information that the sonar sensors return. Information from the robot's camera may also be combined with the sonar data to achieve more complete linguistic descriptions of the robot's environment (*e.g.*, recognize and label objects).

Temporal data may also be used for realization of corridors, rooms, etc. For example, if we have many consecutive linguistic descriptions of being between objects, then the robot could be traveling in a corridor. If we have consecutive descriptions of being surrounded, this could mean that the robot is in a room of a certain size.

5. Concluding Remarks

In this paper, we have shown how the histogram of forces can be used to generate linguistic spatial descriptions representing the qualitative state of a mobile robot in an unknown environment. Using the robot's sonar readings, a boundary approximation of the obstacles is made, and their vertices are used as input to the histogram of forces. The usage described in this paper represents the first application of F_0 and F_2 histograms that uses vector data instead of raster data.

Several examples have been presented which illustrate the linguistic expressions automatically generated. The approach is computationally efficient so that the spatial descriptions can be generated in real time. Note that although we have assumed an unknown environment and therefore must build an approximation of the environment from the sonar readings, the approach could also be used to generate linguistic descriptions for a robot in a known environment using a map. In either case, the linguistic expressions can be used to facilitate natural communication between a robot and a human user.

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.

References

[1] K. Morik, M. Kaiser and V. Klingspor, ed., *Making Robots Smarter*, Kluwer Academic Publishers, Boston, 1999.

[2] M. Skubic and R.A. Volz, "Acquiring Robust, Force-Based Assembly Skills from Human Demonstration", *IEEE Transactions on Robotics and Automation*, to appear.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] M. Skubic, P. Matsakis, B. Forrester and G. Chronis, "Extracting Navigation States from a Hand-Drawn Map", submitted for the *2001 IEEE International Conference on Robotics and Automation*.

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

[9] 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.

[10] K. Miyajima, A. Ralescu, "Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations", *Fuzzy Sets and Systems*, vol. 65, iss. 2/3, pp. 225-236, 1994.

[11] P. Matsakis, *Relations spatiales structurelles et interprétation d'images*, Ph. D. Thesis, Institut de Recherche en Informatique de Toulouse, France, 1998.

[12] 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.

[13] 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.

[14] 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.