Modeling Spatial Referencing Language for Human-Robot Interaction

Samuel N. Blisard
Electrical and Computer Engineering Dept.
University of Missouri - Columbia
Columbia, Missouri 65211
Email: snbfg8@mizzou.edu

Marjorie Skubic
Electrical and Computer Engineering Dept.
University of Missouri - Columbia
Columbia, Missouri 65211
Email: skubicm@missouri.edu

***Abstract*— It has long been a dream of science fiction to have a robot which understands the richness of spoken language. Part of attaining this goal is to exploit the ways that language structures space which will lead to a more natural way to interact with our robots. By investigating the application of spatial language as a modality of interacting with robots, we can create an interface that is more intuitive for a novice user. In this paper, we outline a method for computing target points to the FRONT, LEFT, RIGHT, and BEHIND segmented objects in Evidence Grid maps built using range data on a mobile robot. This method uses the segmented objects and their *contour points* to calculate eigenvectors which can be used to calculate the LEFT, RIGHT, FRONT, and BEHIND points. This allows the user to issue spatial referencing commands, such as "go behind the desk" or "look to the left of the table." We also show results for a human-subject experiment which provides validation of the algorithm.**

I. Introduction

Computational Linguistics is generally defined to be the scientific study of language from a computational perspective. In this field of study, there are two areas that need to be investigated, the first being the cognitive side that deals with how humans acquire, produce, and understand language, and the second being how humans understand the relationship between linguistic utterances and the world.

Several researchers have proposed various ways to deal with the organization and structure of space and its connection to natural language for communication with robots. Kuiper's [1] Spatial Semantic Hierarchy (SSH) is one such concept which separates space into five levels: metrical, topological, causal, control and sensorimotor. The SSH allows for a categorization of space using differing levels of resolution for the task at hand. Gribble et al. [2] used the SSH in a robotic wheelchair demonstrating the practicality of the SSH. Zelek [3] proposed the use of a lexical template in order to tie together spatial references and robot commands for use in 2D robot navigation. In his templates, commands are given in the form of a verb, destination, direction, and speed with the spatial reference tied to the destination reference object. Reference objects were walls and doors that were identified utilizing two laser range finders, each mounted on a pan-tilt head. Through the use of a potential field technique, goal regions were assigned to allow the robot to stop.

Stopp et al. [4] developed a two-arm mobile robot designed for assembly tasks that utilized relative spatial references. The references (e.g. FRONT, LEFT) were used with the robot's geometric world model for the identification of objects. In the world model, a user selected an object by using a relational expression, such as "the spacer on the right." The actual computation of the spatial relations used the center of gravity and a bounding box to approximate an object.

In our previous work, Skubic et al. [5] created a system which could extract linguistic descriptions from sonar returns for use in mobile robots. This provided spatial language from the robot to a human user. Later, we added the ability to issue commands from a human to a robot and introduced a simple algorithm to model the primary directions of LEFT, RIGHT, FRONT, and BEHIND (LRFB), as referenced by an environment object [6]. Here, we extend that work by introducing a new algorithm that matches human behavior in the case of elongated reference objects. Results are included for a human study which validates the algorithm [7].

We are also interested in using these spatial concepts in conjunction with working memory. Storing precise points in space where objects are located does not match well with the human experience. Rather than give precise GPS-style coordinates, we live in a world where objects are to the left, right, on top, in front, to the north, etc., of other objects. If we truly want to model the way a biological system approaches where objects are in the environment, then a reasonable "chunk" in the working memory would be the qualitative spatial relationships, some of which could be LEFT, RIGHT, FRONT, or BEHIND, to name a few. Additional information on our ideas about working memory and its application to robotics can be found in Skubic et al. [8].

In this paper, we will first discuss the procedure for the human subject experiment. We then discuss the eigenvector method for determining points to the LRFB Points of objects in an Evidence Grid (EG) map built using range data. Finally, a comparison of results of the human study experiment and the performance of the eigenvector algorithm is presented.

II. The Human Study

A. Organization

To ensure that the robot's algorithms for spatial languaage accurately model human behavior, an IRB-approved human study was performed. The study was broken into two parts:

1) Directing subjects to go to an interesting area in the environment;
2) Having subjects give directions to get to an interesting area in the environment, designated by marked points on the floor.

The main idea behind these tasks is that if one should ask an average person on the street to "Go to the right of the mailbox," we want to determine where in the environment he would go. If we direct a person to go to that location and he arrives at reasonably the same spot that the robot does, there will be some validation for the spatial language algorithms.

For the two phases of this experiment, the subjects are randomly chosen to do either phase one or phase two first so that we can ameliorate any effects of tiredness or interaction between the two phases on the results of this study. With a total of at least 20 suitable test subjects divided into the two groups, we have a statistically significant number of test subjects so that we are able to draw valid conclusions about the results of the study.

Also, there is a short period of training to acquaint each participant with the tasks. This short training period includes an explanation of what we mean by spatial language and some examples of tasks that they will be performing in the experiment. Here, we present the procedure and results for Phase I only using spatial references.

B. The Participants

Participants were divided into two groups, one that would perform Phase I followed by Phase II (Group A) and another group that would perform Phase II followed by Phase I (Group B). Test subjects were found by asking people who work in the Computational Intelligence Research Laboratory (CIRL) to announce the experiment during their classes. The incentive for participating was an opportunity to earn extra credit in the participant's course, which they were given regardless of the quality of their performance.

There were a total of 24 participants tested for this study. This group of 24 consisted of 14 males and 10 females. Out of the group of 24, a total of 4 had to be rejected due to performance that was considered to be atypical of the norm. One American male was consistently confused about RIGHT and LEFT. One American female was an actress and considered FRONT and BEHIND to be in terms of stage directions. Finally, two Chinese females' interactions showed that we must be careful in regards to cultural differences in how language structures space. Regier's [9] work explores a range of different cultural issues that can arise when looking into spatial language. This left a total of 20 subjects for consideration during the experiment.

C. Pre-experiment Questionnaire

All test subjects were given a pre-experiment questionnaire to fill out. The subjects were asked the following questions:

- Age Range
- Gender
- Handedness (left or right)
- Q1: If someone asked you to "pick up the wrench on your left," would you know where to look for the wrench? Rating: 1 (strong no) - 5 (strong yes)
- Q2: Do you have trouble giving people directions (e.g. to your house)? Rating: 1 (difficult) - 5 (easy):
- Q3: Do you have difficulty following other people's directions to an interesting place? Rating: 1 (difficult) - 5 (easy)
- Q4: Rate how hard it is for you to use a quality map (such as one from AAA or Rand McNally) to get to a place of interest. Rating: 1 (difficult) - 5 (easy)
- Pretend this picture shows your hands. Circle the one that would be your "right" hand.

TABLE I
TEST SUBJECT SURVEY RESULTS MEANS AND STANDARD DEVIATIONS

Group	Q1	Q2	Q3	Q4
All Means	4.46	4	2.88	4.17
All Std. Dev.	1.02	0.98	1.26	0.87
A Means	4.5	3.92	3.08	4
A Std. Dev.	0.9	1.16	1.16	0.95
B Means	4.42	4.08	2.67	4.33
B Std. Dev	1.16	0.79	1.37	0.78

For questions Q1-Q4, the subjects were asked to rate their abilities on a scale from 1 to 5. This was done to get a feeling for their abilities to use spatial language in day to day interactions. Relevant statistics on the answers are shown in Table I.

D. Training

As part of the initial discussions that were done in preparation for the study, it was decided that some preliminary training would be necessary in order to ensure that the subjects understood how to perform the tasks which they were directed to perform. This training consisted of familiarization with the environment and some example tasks for them to perform. A script was employed in order to minimize differences in the subjects' initial experiences with the experiment.

E. Phase I - Directing Subjects using Spatial References

The setup for the first phase of the experiment was conducted at the CIRL at the University of Missouri-Columbia. Fig. 1 shows a scale drawing of this room and the experimental setup. The room is approximately 4.8x6.1 meters. The items of note are the long table, the faux tree, and the "L" shaped configuration of chairs. The reason for choosing these objects is that they each exemplify different ways of mentally calculating the spatial reasoning. The table, while it has no intrinsic front and back, does have intrinsic major and minor axes. The tree was chosen as a generic point-like object with no intrinsic FRONT or BACK information. Finally, the arrangement of chairs was chosen to be an odd shape that was a conglomerate of smaller units, but with no intrinsic FRONT or BACK.

The table was used because, although there is no intrinsic front, a subject could possibly use either the major or minor axis of the table to orient himself in the environment. The

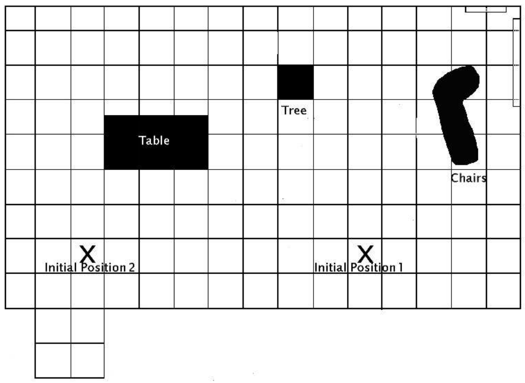

Fig. 1. The Environment of the Experiment

tree is an irregularly shaped blob that is one floor tile in size. Finally, the L-shaped configuration of chairs was used as an object for which there is no real definition as to its major and minor axes.

For the first part of the experiment, each participant was individually brought into the room, where he/she filled out the initial survey form and we answered any questions or concerns about the experiment. The participant was then placed on the start point for phase one which is denoted by the large X taped onto the floor (Fig. 1. The participant started at either starting position 1 or 2 and was asked to go to various places in the environment (such as "left of the tree" or "in front of the table"). The student investigator stood at the unoccupied starting location.

During the course of this experiment, at each point where the test subjects stopped, we measured where in the environment they placed themselves and marked the place on a scaled grid map of the room. In Fig. 2, we see the results of where the test subjects placed themselves from starting position 2 relative to the table.

III. Algorithms for Calculating the LRFB Points

These results could not be modeled with our current system which made necessary the development of a new algorithm which could exploit this tendency of humans to use the major and minor axes of the reference object for alignment. A key part of our method is Matsakis' *Histogram of Forces* (HoF) [10][11] concept. In Fig. 3 we show the LRFB points which our old method, the HoF Main Direction method [6], would calculate. In this figure, the LRFB points lie at the vertices of the trapezoid drawn on top of the table. The arrow shows the *Constant Forces Main Direction* (CFMD) which is calculated from the HoF and passes through the FRONT and BEHIND points. The LEFT and RIGHT points are at the vertices of the trapezoid to the left and right of the CFMD. Fig. 3 shows a segmented EG cell occupancy map as used in previous research [6]. This method works well for many segmented objects but not for elongated objects such as the rectangular table.

For generalization, we need a method to determine whether an object has a dominant major or minor axis, and if so, the LRFB points need to be defined such that they match human behavior. To compute these points, we consider the object's EG map cells NOT as boundaries, but rather as data points.

We use the gridcells to calculate the covariance matrix for the reference object. Eigenvectors are then extracted to point in the direction of the major and minor axes of the data distribution (in this case, it will be 2D vectors). One could also calculate the eigenvectors from the second moment of the object; however, the covariance matrix and the second moment differ by constants, which will be normalized when the eigenvectors are calculated. Detailed descriptions of using the covariance matrix to generate eigenvectors can be found in [12] [13].

By looking at the eigenvectors calculated from the covariance matrix and placing them at the means of the distribution, we can tell the direction of the bases for the data and, hence, the major and minor axes of the object. However, at this point we do not know which is the major or minor eigenvector (as the eigenvectors are by definition unit vectors), so they need to be scaled by their corresponding eigenvalues.

Once we have the scaled eigenvectors, we can then use this information to determine whether we have an elongated object. This is simple, as objects which have eigenvectors where the primary eigenvector's magnitude is more than twice that of the secondary eigenvector's magnitude are suitable candidates for using this method (this restriction is not necessary but is a useful way to select candidates for this method). Once these vectors are assigned, the dot product between a unit vector in the direction of the HoF's Constant Forces Main Direction (CFMD) and the two eigenvectors can be calculated. By looking for the greatest absolute value of the dot product, the search directions for the HoF Main Direction method [6] can be altered to use the eigenvectors rather than perpendiculars based on the CFMD. The detailed algorithm is shown below.

The end result of this process is that for elongated objects, the intrinsic major and minor axes can be detected, and the scaled eigenvectors can be used automatically to alter the calculations of the LRFB points to ensure that they are placed at positions which more closely match human expectations.

A. Algorithm for Calculating Eigenvectors of Segmented Evidence Grid Objects

The steps in the algorithm for this method are as follows:

1) Calculate the covariance matrix for the object;
2) Calculate the eigenvectors and their associated eigenvalues;
3) If the largest eigenvalue is more than twice the smallest, then continue to 4; otherwise start the HoF MD method;
4) Create a unit vector for the CFMD;
5) Calculate the dot product between the cross-scaled eigenvectors and the CFMD unit vector;
6) Set the search direction for the HoF MD method to be in the direction of the greatest dot product calculated in step 5.

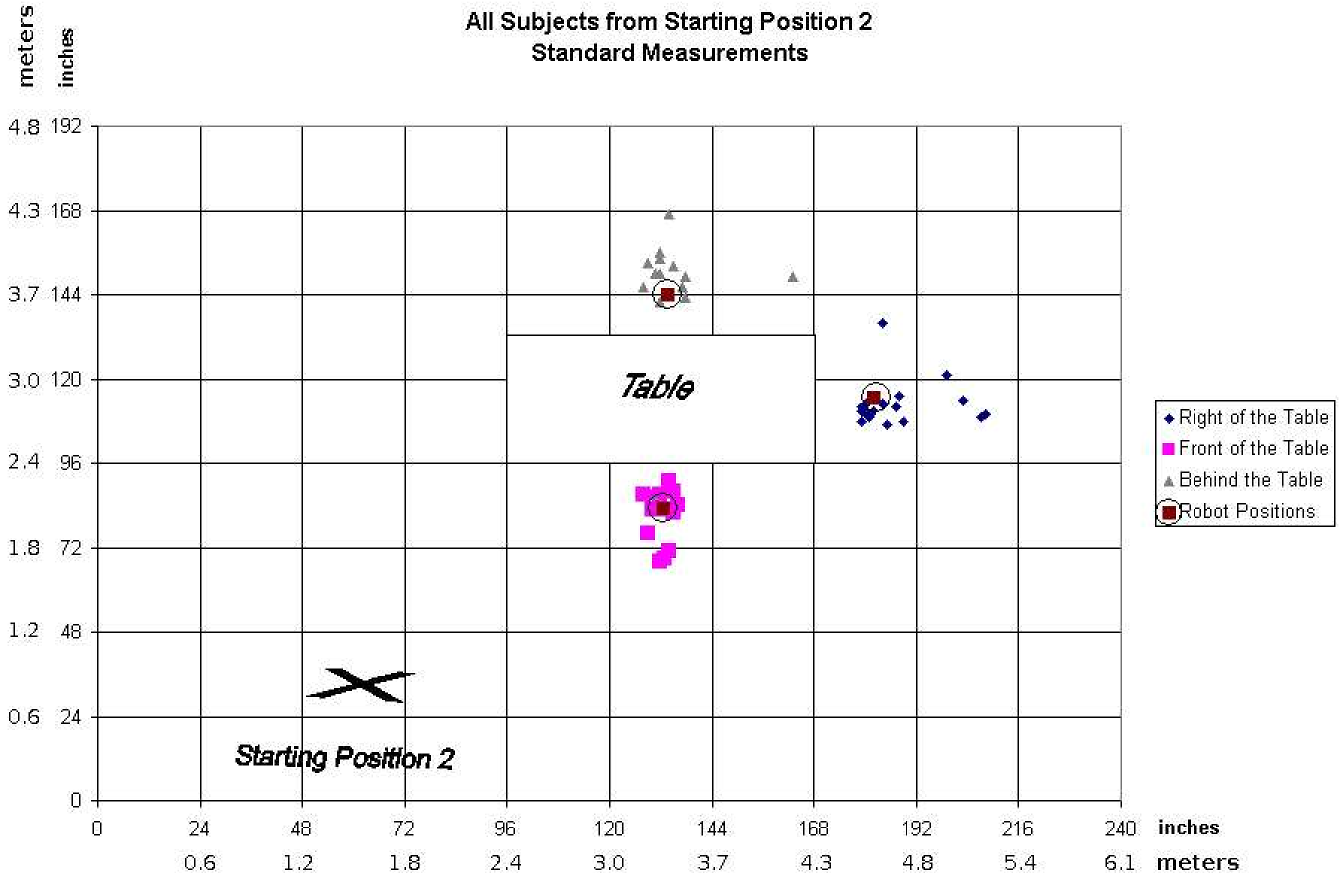

Fig. 2. Results of the Human Subject Experiment around the Table with Eigenvector Method Robot Points

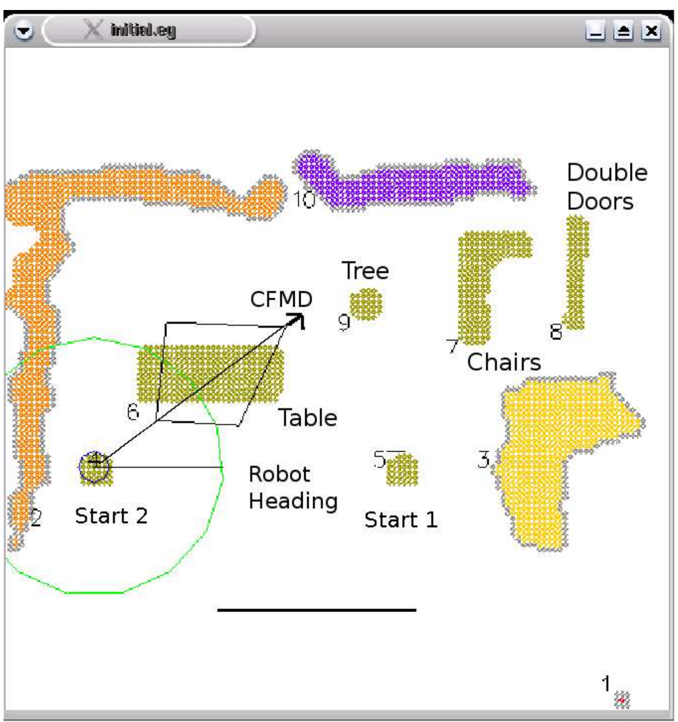

Fig. 3. HoF Main Direction Method Attempting to Model the Points around the Table

In step 1, the covariance matrix is calculated from the contour points of the object by equation 1.

$$Covariance = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x_i}-\mu)^T(\mathbf{x_i}-\mu) \qquad (1)$$

where μ is the mean vector of the contour points for the reference object and $\mathbf{x}$ is a nx2 vector of the contour points that comprise the object.

To compute the eigenvectors and the associated eigenvalues for step 2, equation 2 relates the eigenvalues λ of the matrix A to the eigenvectors $\overrightarrow{x}$.

$$A\overrightarrow{x} = \overrightarrow{x}\lambda \qquad (2)$$

To solve this equation, we used a linear algebra package called the Template Numerical Toolkit (TNT) developed by the National Institute of Standards and Technology (NIST)[14].

In step 3, we can now determine whether the reference object has an intrinsic major or minor axis. If the eigenvalue associated with the primary eigenvector is twice as large or larger than the secondary eigenvalue, then the data has a definite skewing in the direction of the primary eigenvector and, hence, an intrinsic axis. If this is not the case, then we assume a blob-like object (such as a square or circular shape), where the data is more or less equally distributed.

If it is the case that we have intrinsic major and minor axes, then we must determine which of the eigenvectors is most closely associated with the current CFMD between the robot and the object. To calculate this, we must first find a unit vector called the Main Direction Vector ($\overrightarrow{V}_{MD}$) based off the current CFMD in direction θ with respect to the robot with the equation 3, which yields the (x,y) components of $\overrightarrow{V}_{MD}$.

$$\overrightarrow{V}_{MD} = (a{\cdot}cos(\theta), a{\cdot}sin(\theta)) \tag{3}$$

Note that there is a length a in equation 3; however, since we want a unit vector, this value is 1. The dot product is then taken between $\overrightarrow{V}_{MD}$ and each of the eigenvectors. Suppose that the dot product is taken between the raw eigenvectors and the $\overrightarrow{V}_{MD}$. This will lose any information about which direction is greater as the eigenvectors are orthonormal and do not have the eigenvalues associated with them. Therefore, we need to multiply the eigenvectors by the associated eigenvalues, but in this there is also a problem. When the dot product is taken, the bias is in the direction of the primary eigenvector. At first, this would seem reasonable; however, consider its effect in relation to where the robot is positioned. If the eigenvalue is particularly large, in order for the $\overrightarrow{V}_{MD}$ and secondary scaled eigenvector dot product to be the greater value, it must be almost perfectly aligned with the secondary eigenvector. This is not what our subjects showed and is also unreasonable. What you would expect is that as a table becomes more and more elongated, the front bias would grow for the long side of the table. To get this bias, the eigenvalues are switched for the dot product. This has the effect of biasing the calculation of front (and the subsequent LRFB points) in the direction that corresponds to the human behavior observed. This explains the term *cross-scaled* in reference to the eigenvector dot product in step 5.

Finally, once the greatest dot product has been found, the search direction for the FRONT and RIGHT vectors into the object are set to the eigenvectors rather than using the CFMD in the HoF MD method.

TABLE II
ROBOT PERFORMANCE, ALL SUBJECTS, X VALUES, METERS

Robot Performance All Subjects X Values				
Mean Test Subjects	Std. Dev. Test Subjects	Previous HoF MD Method		New Eigenvector Method
5.03	0.40	5.11	LEFT of Tree	5.11
7.39	0.50	7.59	FRONT of Chairs	7.52
8.43	0.41	8.76	RIGHT of Chairs	8.76
5.83	0.12	5.94	FRONT of Tree	5.94
7.53	0.51	7.24	LEFT of Chairs	7.24
4.54	0.66	3.86	RIGHT of the Table	4.85
3.35	0.35	2.72	FRONT of the Table	3.40
3.41	0.40	2.79	BEHIND the Table	3.40

TABLE III
ROBOT PERFORMANCE, ALL SUBJECTS, Y VALUES, METERS

Robot Performance All Subjects Y Values				
Mean Test Subjects	Std. Dev. Test Subjects	Previous HoF MD Method		New Eigenvector Method
3.83	0.18	3.81	LEFT of Tree	3.81
3.95	0.67	2.82	FRONT of Chairs	2.74
3.68	0.62	3.96	RIGHT of Chairs	3.96
3.53	0.66	3.3	FRONT of Tree	3.3
4.03	0.58	4.57	LEFT of Chairs	4.5
2.77	0.21	2.13	RIGHT of the Table	2.82
2.53	0.79	2.13	FRONT of the Table	2.13
3.48	0.64	3.66	BEHIND the Table	3.66

IV. Comparison of Results and Discussion

The results for this experiment are shown in Figs. 2 and 4. Fig. 2 which dramatically shows that people tend to orient themselves with the major and minor axes of objects, even when there is no intrinsic FRONT or BACK. Fig. 4 is presented to show the full results of our algorithm for all objects considered in the environment. In Tables II-IV, the statistics are presented (after removing 4 outliers in the data) for the test subjects. The column for HoF MD stands for *Histogram of Forces Main Direction* and is from previous work we had performed [6] as a comparison of the improvement. As shown, the Eigenvector Method places the X and Y coordinates to within one standard deviation of our human study data. The only exception to this is the Y coordinate for the position "In front of the chairs" which shows the difficulty in modeling human behavior. As shown in Fig. 4, we can see that many people went close to the same spot for "left of the chairs" and "in front of the chairs." This anomaly requires further investigation and most likely arises from the fact that we are dealing with a collection of similar objects to form a larger composite object as opposed to a single object. Many participants in the experiment expressed a look of confusion when asked to go to those locations.

V. Concluding Remarks

During the course of the human subject experiment, the revelation that people use the major and minor axes of objects without intrinsic information was very interesting. It also illustrates the difficulty of attempting to model something we, as humans, have a relatively easy time performing.

In this paper, we have shown a new algorithm for computing target points to the LEFT, RIGHT, FRONT, and BEHIND of an object using the Histogram of Forces and eigenvectors. This algorithm facilitates the use of commands such as "Go to the

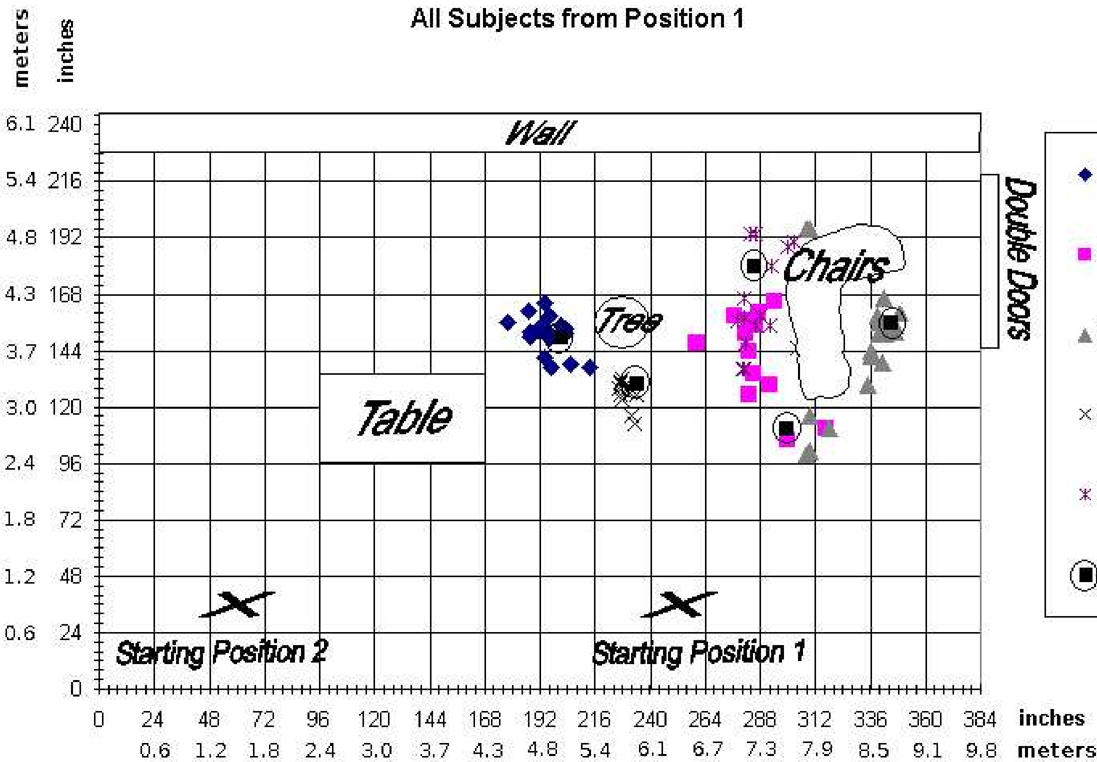

Fig. 4. All subjects from starting position 1

left of the table" with results that place the robot in a location consistent with human expectations.

In the future, we plan to use these qualitative locations in space as possible "chunks" in the working memory of our robots, to test the use of relational, qualitative locations vs. the representation of absolute position for a set of test scenarios. Using an adaptive working memory toolkit, we will compare the efficiency of these different spatial representations for specific types of tasks. For example, in a scenario where the robot is asked to search for an object in the environment, the human collaborator might tell the robot that he left his coffee cup near a computer in the lab. With prior knowledge that the person is right handed, the robot could then try looking in the area that is to the RIGHT of each computer to limit the search region. Such scenarios will be investigated using the algorithms presented in this paper.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the Naval Research Laboratory and NSF (EIA-0325641) for their support of this work.

REFERENCES

[1] B. Kuipers, "A hierarchy of qualitative representations for space," in *Spatial Cognition*, C. Freska, C. Habel, and K. Wender, Eds. Berlin: Springer-Verlag, 1998, pp. 337–250.

[2] W. Gribble, R. Browning, M. Hewitt, E. Remolina, and B. Kuipers, "Integrating vision and spatial reasoning for assistive navigation," in *Assitive Technology and Artificial Intelligence*, V. Mittal, H. Yanco, J. Aronis, and R. Simpson, Eds. Berlin: Springer-Verlag, 1998, pp. 179–193.

[3] J. Zelek, D. Bullock, S. Bromley, and H. Wu, "What the robot sees and understands facilitates dialog," in *Human-Robot Interaction, Papers from the AAAI Fall Symposium*, vol. FS-02-03. AAAI, 2002, pp. 120–136.

[4] 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 13th IJCAI*. Berlin, Germany: IJCAI, 1994, pp. 39–50.

[5] 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, 2003.

[6] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock, "Spatial language for human-robot dialogs," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. Special Issue on Human-Robot Interaction, pp. 154–167, 2004.

[7] S. N. Blisard, "Modeling spatial references for unoccupied spaces for human-robot interaction," Master's thesis, University of Missouri-Columbia, 2004.

[8] M. Skubic, D. Noelle, M. Wilkes, K. Kawamura, and J. Keller, "A biologically adaptive working memory for robots," in *AAAI Fall Symposium Workshop on the Intersection of Cognitive Science and Robotics: From Interfaces to Intelligence, Washington D.C.* AAAI, 2004.

[9] T. Regier, *The Human Semantic Potential*. Cambridge, MA: Bradford Book, MIT Press, 1996.

[10] P. Matsakis and L. Wendling, "A new way to represent the relative position between areal objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 634–643, 1999.

[11] P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera, "Linguistic description of relative positions in images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B.*, vol. 31, no. 4, pp. 573–588, 2001.

[12] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Second Edition*. San Diego, California: Academic Press, 2003.

[13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, Second Edition*. New York City, New York: John Wiley and Sons, 2001.

[14] R. Pozo. (2004) Nist template numerical toolkit. [Online]. Available: http://math.nist.gov/tnt/