# Mobile robot sensor fusion with fuzzy ARTMAP

SIEGFRIED MARTENS, PAOLO GAUDIANO, AND GAIL A. CARPENTER
Boston University Neurobotics Lab
Department of Cognitive and Neural Systems
677 Beacon St., Boston, MA 02215, USA
{sig,gaudiano,gail}@cns.bu.edu, http://neurobotics.bu.edu

*Abstract* - **The raw sensory input available to a mobile robot suffers from a variety of shortcomings. Sensor fusion can yield a percept more veridical than is available from any single sensor input. In this project, the fuzzy ARTMAP neural network is used to fuse sonar and *visual sonar* on a B14 mobile robot. The neural network learns to associate specific sensory inputs with a corresponding distance metric. Once trained, the network yields predictions of range to obstacles that are more accurate than those provided by either sensor type alone. This improvement in accuracy holds across all distances and angles of approach tested.**

## I. INTRODUCTION

Mobile robots require accurate representations of their surroundings for navigation. Acquiring these representations involves stages of increasing abstraction, transforming analog streams of sensor values into a symbolic view of the world. Sensor fusion is a topic of great current scientific interest (Luo and Kay, 1989; Huntsberger, 1992; Murphy, 1994; Murphy, 1996). Individual sensors tend to have shortcomings limiting their applicability; sensory data can be fused, however, and the fused percept can be more veridical than that provided by any single sensor.

This research employs a B14 mobile robot from Real World Interface, Inc. (Jaffrey, NH), a cylindrical robot measuring 14" in diameter (Figure 1), equipped with a synchro drive that permits forward and reverse translation and rotation in place. Arranged around the B14's surface are sixteen infrared proximity detectors and sixteen sonar range finders, distributed uniformly around the robot's perimeter; and a camera mounted on a pan-tilt platform. This project uses only the frontal eight of the B14's sixteen sonar and infrared sensors.

The robot learns to predict the frontal distance to obstacles using variations of the fuzzy ARTMAP neural network (Carpenter, Grossberg, and Reynolds, 1991; Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992). The training process is self-supervised, i.e., the robot is not provided with the distance to obstacles. A relative distance metric is obtained using odometry, as

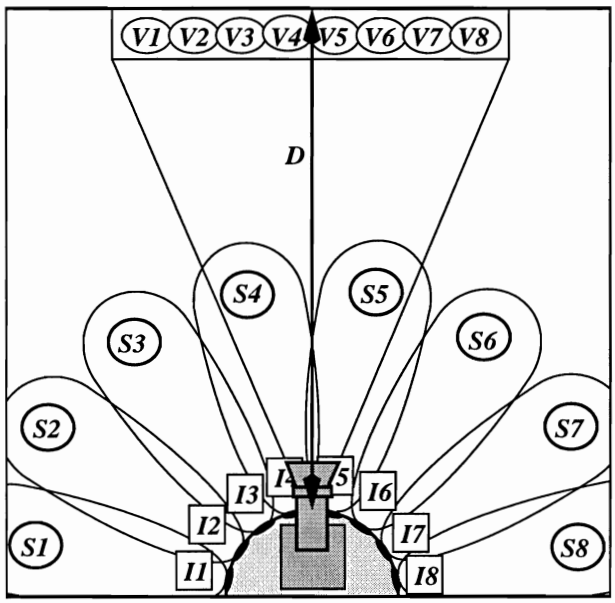


**Figure 1: Overhead view of the B14 robot, showing its ranging sensors. Closest to the robot are infrared sensors, labeled *I1-I8*. These detect obstacles at close range, and specify the angle at which the obstacle is encountered. Sonars are labeled *S1-S8* (beams emanating radially from robot). A camera is mounted on top of the robot, and provides gray-scale images. An edge detection algorithm is applied to these images, yielding *visual sonar*, a distance metric depicted as sensors *V1-V8* (See Fig. 2). Odometry provides the distance *D* to the nearest obstacle.**

the robot randomly explores its training area. Snapshots of the sensory input are recorded as the robot moves in a straight line. When the robot encounters an obstacle detected with the infrared sensors, the on-board odometer provides a relative distance to associate with the sensory snapshots. The neural network is trained to learn the association between the robot's sensory input and the distance to the obstacle. The robot thus learns to interpret its sensory input on its own, without human intervention. This self-supervised learning can allow the robot to explore new environments on its own. This ability is complemented by the ARTMAP network's

(a) (b)

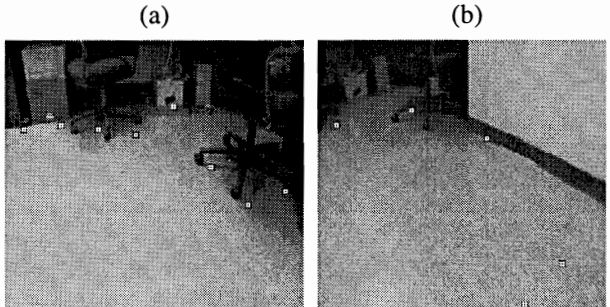

**Figure 2: Visual sonar's view of a room. The height of the points reflect their distance from the robot, and are returned as the values of sensors *V1-V8*. Perception is accurate in (a), but in (b) markings on the floor are mistaken for an obstacle**

capacity for *one-shot learning*, which allows it to form associations between inputs and percepts with a single presentation.

Input vectors to the ARTMAP neural network are collected from the camera and from the sonar sensors. The camera provides a crude estimate of relative distance to objects through *visual sonar* (Horswill, 1994), obtained by dividing the image into eight columns and searching for edges from the bottom of each column. Under the assumptions that the robot is operating indoors and that obstacles are on the ground, the distance of an edge from the bottom of a column is proportional to the distance of the corresponding obstacle from the robot. Figure 2 provides an example of visual sonar working in an office setting. The points plotted on the image represent the lowest edge in each column, and the relative heights of these points are the distance metrics returned by visual sonar. The sensory input from visual sonar is thus a vector of eight real-valued numbers (***V1-V8***), with low numbers indicating a nearby obstacle and high numbers indicating free space. The visual sonar sensor detects the bases of walls well, but is confused by image discontinuities that do not correspond to obstacles, as resulting for instance from textured carpeting, door sills, or markings on the floor (e.g. Figure 2(b)).

Sonars are the other main sensors used. They emit an ultrasonic beam and use the amount of time until the echo to calculate distance to obstacles. The input to the neural network from the sonars consists of a vector of up to eight numbers (***S1-S8***), describing the distance to obstacles as measured by the B14's eight frontal sonars. The raw sonar data suffer from a variety of limitations; the returning echo may have been emitted by a different sensor, or may have bounced off several surfaces before being detected. Ultrasonic and visual sonar both have sensory limitations, but in different circumstances, and so they seem apt choices for sensor fusion.

## II. DATA SET COLLECTION

A data set of 10167 data samples has been collected, each consisting of a set of sensor readings and an associated distance metric obtained through odometry. The values specified are ***S1-S8*** for sonar, ***V1-V8*** for visual sonar, the associated distance value ***D***, and the angle of approach to the obstacle, specified by the infrared sensor $I_x$ which detected it. Samples are recorded every 20 cm as the robot travels in a straight line. When an obstacle is encountered, as measured by the infrared proximity detectors, a new direction is chosen randomly and the process is repeated. The data collection is conducted in an empty area, approximately 2m by 3.5 m, bounded by flat surfaces (walls and styrofoam panels).

The recorded distance ***D*** is specified as a continuous value. As it provides the teaching signal to a classifier, the distances are binned into 20 discrete categories. Binning is nonlinear, with small distance bins up close and bin size proportional to the square root of the distance. This nonlinear binning allows for more accurate predictions at shorter distances. Figure 3(a) shows these 20 distance bins, with the width of the column indicating the range of distances covered by the bin and the height of the column representing the number of samples in the bin. The semicircular sector plot in Figure 3(b) gives a spatial view of the density of the data collected. The radial bins correspond to the different angles of approach specified by the infrared

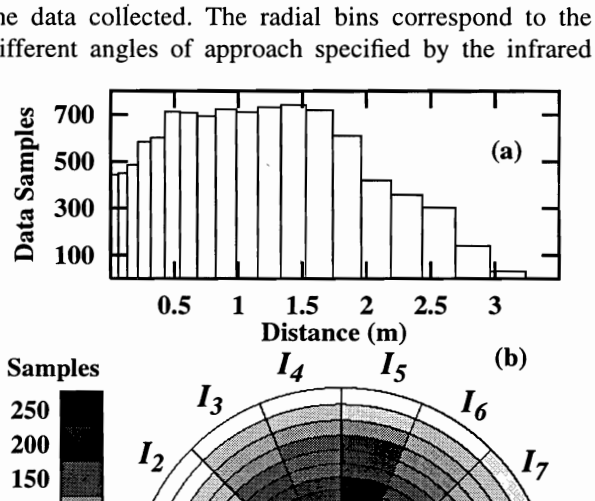


**Figure 3: Density of data sampling. (a) Data samples for each distance bin. (b) Data samples for each distance bin broken down by angle of approach. The non-linear increase in size of distance bins is shown in both plots.**

detector $I_x$ which detected the obstacle. The radial thickness of each cell in the semicircle also reflect the non-linear binning of distances. It is important to note that although in this and subsequent figures we break down the data by the angle of approach for clarity, the neural network is trained by lumping all inputs from a given distance regardless of angle. Figure 3 shows that most encounters are at head-on approaches, though all bins except some of the outermost are represented.

## III. RESULTS

The fuzzy ARTMAP neural network is applied to the data set, learning to associate different combinations of sensory inputs with the distance $D$. For each of the $n$ data samples, the network produces a predicted distance $p_i$, which is compared to the recorded distance $D_i$. Simulations are compared using $\varepsilon$, the average absolute predictive error over the data set:

$$\varepsilon = \frac{\sum_{i=1}^{n} |p_i - D_i|}{n} \quad (1)$$

Fifteen simulations are performed, each with a different combination of inputs provided to the neural network for fusion. Combinations range from sonar only or visual only to mixtures of both sensory modalities. Specifically, the sensor combinations use two, four or eight of the available values (sonar or visual), in each case using the most central sensors, starting with ***S4***, ***S5*** or ***V4***, ***V5***.

Five-fold cross validation ensures that the training and test sets are always disjoint. More specifically, the data set is divided into five partitions. One at a time, each of these data partitions is reserved as the test set while the network is trained with the remaining four. This process is repeated five times, each time using a separate data partition for testing, and the remainder of the data set for training. Moreover, to enhance the accuracy of prediction, and the repeatability of the results, five copies of the neural network are trained, each with a separate ordering of the training set. The final predicted distance $p_i$ is the average of the predictions of these five networks. Thus, for each reported simulation, twenty five copies of the neural network are used, one for each of the five orderings of the five training sets. Table 1 compares the average absolute error for the fifteen simulations performed.

**Table 1: Average absolute error (cm), for all fused combinations of sonar and visual sensors.**

| Visual \ Sonar | 0 | 2 | 4 | 8 |
|---|---|---|---|---|
| **0** | | 31.1 | 15.5 | 13.1 |
| **2** | 34.7 | 12.6 | 11.2 | 11.1 |
| **4** | 25.0 | **10.3** | 10.5 | 10.7 |
| **8** | 16.7 | 10.6 | 10.7 | 10.7 |

The best result, highlighted in boldface, uses the two most central sonar sensors and the four most central visual sensors (***2S+4V***). However, all of the results using any of the sonars and at least four visual sensors are nearly as good. Table 1 demonstrates the advantage provided by sensor fusion, both within and between sensory modalities. When using only one type of sensor, performance is proportional to the number of sensors used. The predictive error is always less, however, when both sonar and visual data contribute to the prediction.

Before showing our results with the ARTMAP network we illustrate the accuracy obtained with our data set using the raw sonar data. Figure 4 shows, for the two frontal sonars ***(S4, S5)***, the average absolute error as given Equation 1, broken down by impact angle ***(I1-I8)***. Below each semicircle, the raw error data are plotted as a function of distance, again broken down by angle of impact, with the leftmost box corresponding to ***I1*** and the rightmost box corresponding to ***I8***. Each point in these scatter plots is the calculated difference between measured and actual distance. If each sonar were "perfect", all the points would be zero, i.e., the data would lie on the horizontal midline. These error data show some interesting trends: the majority of the points fall above the horizontal midline, suggesting that sonar tends to overestimate, probably due to echoes bouncing against more than one surface or cross-talk between sensors. However, the data show a tendency to underestimate at large distances. It is unclear whether this is due to the sonar itself (e.g., through reflected echoes), or to the fact that we are using our odometry to measure "actual" distance, and odometry is likely to contain a systematic error over the range of distances we

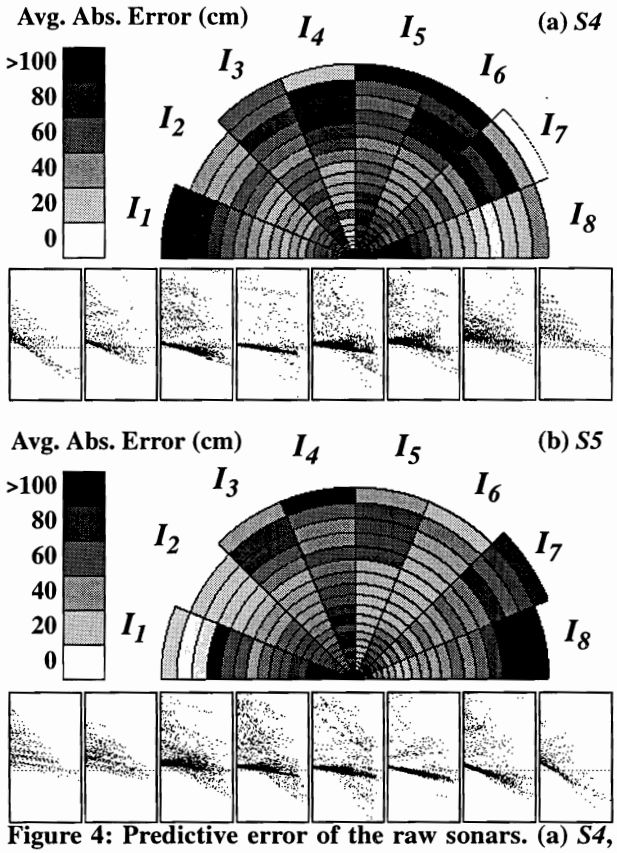


**Figure 4: Predictive error of the raw sonars. (a) *S4*, and (b) *S5*. Semicircular plots show average absolute error level in each radial/distance bin (bins not shown were not sampled in data collection). Small dot plots show error for each datum collected, one plot for each of the eight angles of approach. Y axis thus shows predicted distance minus recorded, and is in the range [-200, 300] (cm). The X axis range is from [0, 325] (cm).**

used. It should also be noted that these results for sonar at high angles of approach are unusually good. One might expect performance to decline in proportion to this angle. One explanation can be seen in the small plots showing the raw data. Note that each sonar displays a trend: underestimating at short distances and overestimating at long distances. Necessarily, there will be a transitional point where predictions are accurate. For example, in (b), prediction at $I_1$, $D \approx 3m$ is unusually good. Looking at the leftmost raw data plot, however, we see that at large distances, i.e. to the right in the plot, the few predictions made fall right on the axis. This is an artifact of the crosstalk, and should not be taken for an accurate prediction. Most of the diagonally oriented trends in the raw data are indicative of noise due to crosstalk between the sonars and complicate the extraction of useful information.

These figures make clear that relying on raw sonar

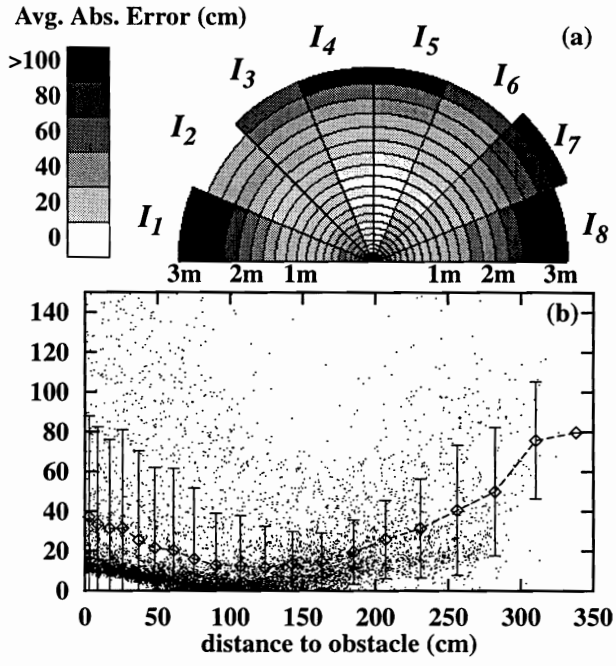


**Figure 5: Predictive error of the *min(S4, S5)* model. (a) Average absolute error in each sampling bin. Model predictions of distance are accurate at short distances and head-on approaches, but performance is impaired at longer distances or more oblique angles of approach (b) Absolute error by distance. Dots show absolute predictive error for each datum, dotted line shows average absolute error for each distance bin, and error bars indicate one standard deviation.**

data is inadequate for safe navigation. A simple way of getting rid of some of the noise is to take the minimum value returned by (***S4***, ***S5***).This assumes that at least one of the pair will register the appropriate echo. The validity of this idea is shown in Figure 5, which shows the predictive ability of a sensor based on ***min (s4, s5)***. Accurate prediction is extended to most of the central angular bins, representing relatively orthogonal angles of approach (Fig. 5(a)). It is interesting that the ***min(S4,S5)*** seems to do poorly for obstacles straight ahead at large distances, as is evident in the outermost cells for sectors ***I4*** and ***I5*** in Fig.5(a). The individual scatter plots in Figure 4 suggest the reason for this problem: notice that both sonars, especially at the central angles, tend to generate a lot of underestimates at larger distances, so that ***min(S4,S5)*** actually worsens the results. In Figure 5(b), the absolute error of the prediction is shown for each sample in the data set (dots). The averages of the absolute error values are plotted, as well as the standard deviation within each of the distance bins. The lowest level of error is seen in the middle range, between one and two meters. This is problematic for sonar-based navigation, as the high

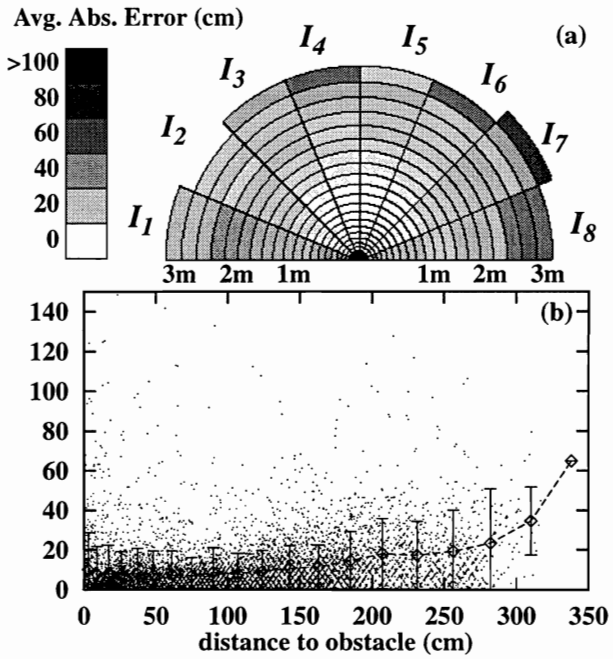


**Figure 6: Predictive error of the fuzzy ARTMAP *(2S+4V)* fusion model. Plot contents and scales are the same as in Figure 5. (a) Predictive accuracy is equal or better at all distance and angle bins. In particular, prediction of distance is improved at high angles as well as at large distances. (b) The average absolute error is less at all distances, and the variance of the error is smaller.**

level of error close to obstacles increases the likelihood of collisions. Figure 5(a) demonstrates that much of the error at short distances occurs during oblique approaches. Neither ***S4*** nor ***S5*** receive the appropriate echo, and so overpredict the distance to the obstacle. Error rises again at higher distances, between two and three meters, but this is less of a problem than the error at short distances. With this illustration of the problems of prediction based on raw sonar, we can look in more detail at the result of fusing sensory information with fuzzy ARTMAP.

Figure 6 illustrates the predictive performance of the neural network fusing ***2S+4V***, i.e. the two most central sonars and the four most central visual sonars. Table 1 showed this to be the combination of sensors yielding the overall lowest average absolute error. Figure 6(a) shows improvements in prediction in nearly every angle and distance bin. Performance at high angles of approach and distances is dramatically improved. Even more important, prediction at close range is now quite accurate at all angles of approach. This is more visible in (b), which shows that average absolute predictive error is a nondecreasing function of distance. Predictive accuracy thus now has the desirable

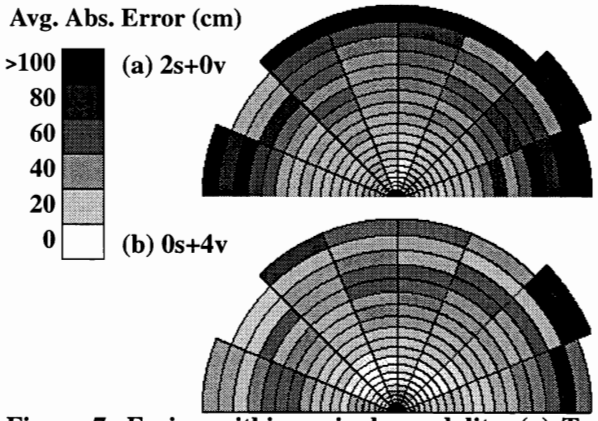


**Figure 7: Fusion within a single modality. (a) Two frontal sonars are fused. (b) Four central visual sensors are fused.**

property of being roughly proportional to distance. Moreover, the standard deviation of the error is significantly smaller than in Figure 5, and it is also proportional to distance. Together, these features make the fused sensors a safer basis for navigation than the raw sonar data.

## IV. Conclusion

These results demonstrate the viability of fuzzy ARTMAP as a methodology for fusing data across and within sensory modalities. Using two sonar and four visual sensors, a fused system has been devised which outperforms the raw predictions of distance of the two frontal sonars. The obtained system is robust in several key aspects. It is more accurate at high angles of approach to an obstacle, compensating for sonar's inability to detect walls at oblique angles. Furthermore, the errors made by the fused system are small at short ranges and larger at long ranges, and the variance of the error follows the same trend. This allows the robot to navigate with confidence at close quarters.

To demonstrate the benefit of fusing sensors across modalities, Figure 7 plots the predictive error obtained by training the network with either (a) two sonars, or (b) four visual sonars. These are the sensory components that were fused to produce the results in Figure 6. In either case, with only one modality available, more errors are made (Table 1). Note, however, that the errors are in different areas. In Figure 7(a), the sonar based fusion system makes most of its errors at large distances, and does fairly well at short ranges. In (b) by contrast performance at a distance is reasonable, and short-range prediction suffers. When the robot is too close to a wall, visual sonar misses the wall entirely. It needs to see the edge at the bottom of the wall to detect it successfully. As the two sensory modalities make errors in different areas, the neural network is able to

combine them to produce a system that exceeds the capability of either separate modality. It was noted that the presence of styrofoam panels helps the performance of the sonar sensors. The improvements from fusing sensor types might be even more pronounced in a more typical indoor environment where specular reflections from walls severely limit the functionality of sonars.

There are other possibilities for reducing the error inherent in the raw sonar signal. Borenstein and Koren (1995) have proposed an "error eliminating rapid ultrasonic firing" algorithm to reduce the number of erroneous readings returned by the sonars. While apparently yielding robust results, this method requires detailed control over the timing sequence of firing the sonars, a degree of control not always available. The method presented here requires nothing more than the sensory data already available.

Others have studied the application of neural networks to the fusion of sensor data (e.g. van Dam, Kröse and Groen, 1996) In particular, Racz and Dubrawski (1995) use the fuzzy ARTMAP network to classify a robot's position within the neighborhood of a door. In their study, however, the position of the robot within its environment is explicitly specified at the start of data collection. In this research no such supervision is necessary, as the robot discovers the retrospective distance to obstacles on its own, through self-guided exploration of its environment. Such a capacity allows the robot to adapt to new environments on its own, a process aided by the fuzzy ARTMAP network's ability to learn incrementally.

All of the results presented here are based on off-line training and testing using a data set. The fact that training is done in batch mode should not be taken as evidence that the fuzzy ARTMAP network cannot learn on-line. Off-line learning is used here so that various fusion models may be tested against the same set of data, yielding comparable results.

Future work on this project will move beyond this off-line paradigm. The best performing algorithms will be put into the robot and used to perform visualization studies. Some work has been done in this area by having the robot rotate and plotting the range predictions, yielding a view of the room in polar coordinates. To make the work more directly comparable to existing work, however, we plan to implement an occupancy grid framework (e.g. Elfes, 1989). Using this framework, range predictions will be used to fill in the cells of the occupancy grid, gradually filling in a picture of the room as the robot moves about.

**Acknowledgments:** This work is supported by DARPA, the Office of Naval Research and the Navy Research Laboratory via grants ONR-00014-96-1-0772, ONR-00014-95-1-0409, and ONR-00014-95-0657.

## V. References

Borenstein, J., and Koren, Y. (1995) Error Eliminating Rapid Ultrasonic Firing for Mobile Robot Obstacle Avoidance, *IEEE Transactions on Robotics and Automation*, February 1995, Vol. 11, No. 1, pp 132-138.

Carpenter, G.A., Grossberg, S., Markuzon, N. Reynolds, J.H., and Rosen, D.B. (1992) Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, 3:698-713.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991) ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, *Neural Networks*, 4:565-588.

Elfes, A. (1989) Using occupancy grids for mobile robot perception and navigation, *IEEE Computer*, June, pp. 46-57.

Horswill, I. (1994) Collision Avoidance by Segmentation, in *Proceedings of the 1994 International Conference on Intelligent Robots and Systems (IROS-94)* IEEE Press.

Huntsberger, T.L. (1992) Data Fusion: A Neural Networks Implementation, in Abidi, M.A. and Gonzalez, R.C., *Data Fusion in Robotics and Machine Intelligence* (1992) Academic Press, San Diego, CA. pp. 507-535.

Luo, R.C. and Kay, M.G. (1989) Multisensor Integration and Fusion in Intelligent Systems, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 5, pp. 901-931.

Murphy, R.R. (1994) *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, MA. pp. 857-860.

Murphy, R.R. (1996) Biological and Cognitive Foundations of Intelligent Sensor Fusion, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 26, No. 1, pp. 42-51.

Racz, J. and Dubrawski, A. (1995) Artificial neural network for mobile robot topological localization, *Robotics and Autonomous Systems*, Vol. 16, pp. 73-80.

van Dam, J.W.M., Kröse, B.J.A., and Groen, F.C.A. (1996) Neural Network Applications in Sensor Fusion For An Autonomous Mobile Robot. *Proceedings of the International Workshop on Reasoning with Uncertainty in Robotics*. Eds. Dorst, van Lambalgen, Voorbraak. Amsterdam. pp. 1-19.