Neurobotics Lab Research: Learning, Vision and Sonar Recognition with Mobile Robots

†‡Paolo Gaudiano, ‡Carolina Chang, ‡Ihsan Ecemis, †‡Siegfried Martens, ‡Erol Şahin, ‡William Streilein and ‡Robert Wagner

†Artificial Life, Inc., Four Copley Place, Boston, MA 02116
‡Boston University Neurobotics Lab, 677 Beacon St., Boston, MA 02215

Abstract

This article provides an overview of research projects undertaken in the Neurobotics Laboratory at Boston University. We focus on applications of neural networks and other biomimetic techniques in sensory processing, navigation, and other tasks using mobile robots. These applications share some central themes: the inclusion of minimal assumptions about the robots and the environment; cross-validation of modules on a variety of robotics platforms and environments; and real-time operation using real robots.

Keywords: Mobile robots, looming, mobile robots, robot learning, Neural networks, ARTMAP, sensor fusion

1 Introduction

The Neurobotics Laboratory was founded in 1996 with the goal of applying neural networks and other biomimetic techniques to the control and guidance of wheeled mobile robot. Research in the lab covers various problems in the general area of autonomous mobile robotics, with an emphasis on navigation and control using biomimetic algorithms that operate in real-time with only minimal assumptions about the robots or the environment, and that can learn, if needed, with little or no external supervision.

In the following sections we describe several research projects that we have investigated over the last three years. In the first project, a neural network model of classical and operant conditioning learns to generate avoidance and approach movements in a mobile robot without external supervision. The second project involves the use of *visual looming* as a simple, efficient technique for ranging and localization. The third project utilizes neural networks based on Adaptive Resonance Theory (ART) for ranging and localization using sensor fusion between sonar and visual estimates of distance. The fourth project also uses an ART-based neural network, but this time to perform real-time object recognition using spectral information from ultrasonic sensor returns.

2 Learning approach and avoidance behaviors without supervision

When an animal has to operate in an unknown environment it must somehow learn to recognize informative cues in the environment, and to predict the consequences of its own actions. This learning is possible for organisms in spite of what seem like insurmountable difficulties from a standard engineering viewpoint: noisy sensors, unknown kinematics and dynamics, non-stationary statistics, and so on.

Psychologists have identified classical and operant conditioning as two primary forms of learning that enable animals to acquire the causal structure of their environment. *Classical conditioning* refers to the act of learning to recognize informative stimuli in the environment; for instance, a dog can be trained to learn that a ringing bell precedes the arrival of a shock, until the bell itself will cause the dog to be afraid. In the case of *operant conditioning*, an animal learns the consequences of its actions. More specifically, the animal learns to exhibit more frequently a behavior that has led to a reward, and to exhibit less frequently a behavior that has led to punishment. For example, a pigeon can be trained to peck at an illuminated key in order to receive a small food reward.

In 1971, Grossberg proposed a model of classical and operant conditioning, which was designed to account for a variety of behavioral data on learning in vertebrates. The model was refined in several subsequent publications. In 1987, Grossberg & Levine described a computer simulation of a major component of the conditioning circuit. This model was used to explain a number of phenomena from classical conditioning. Our implementation of Grossberg's conditioning circuit, which follows closely that of Grossberg & Levine [11], is shown in fig. 1.

In the present model the nodes at the upper left of Fig. 1 receive activation from the robot's range sensors. Note that there is no knowledge built into the network about (1) the kind of sensor information (e.g., infrared or sonar), or (2) the position of the sensor on the robot's body.

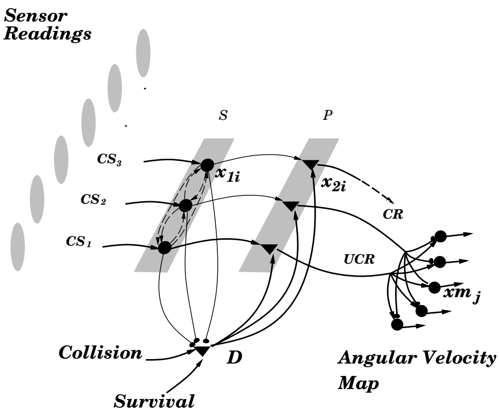


Figure 1: Conditioning model for obstacle avoidance. The range sensor activities represent the CSs. A collision detector activates the UCS. Motor learning occurs at a population coding the robot's target angular velocity. After conditioning, the pattern of activity across the range sensors can predict a collision and modify the robot's angular velocity to avoid the obstacle.

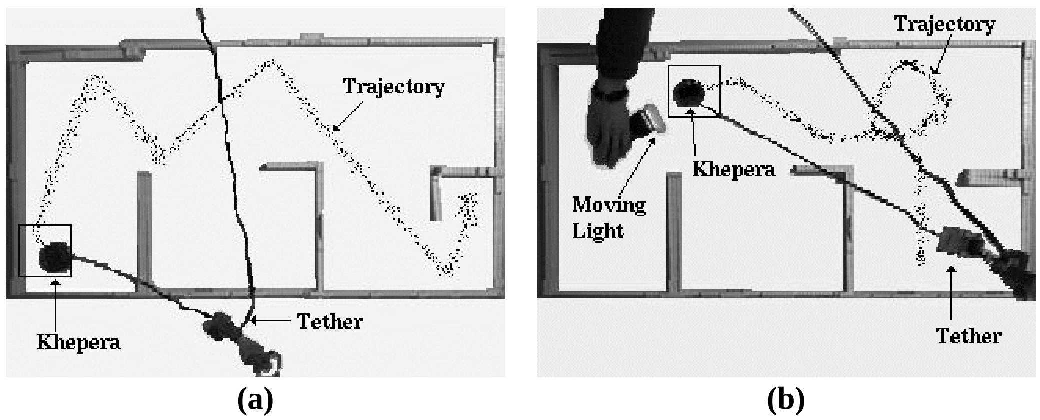


Figure 2: Overhead view of the Khepera in its environment during obstacle avoidance (a) and light approaching (b) behaviors.

At the bottom of Fig. 1 the *drive node* D is represented: conditioning can only occur when the drive node is active. In our model this happens when the robot collides with an obstacle, which could be detected through a bump sensor, or when any one of the range sensors indicates that an obstacle is closer than the sensor's minimum range.

Finally, the neurons at the far right of the figure represent the response (conditioned or unconditioned), and are thus connected to the motor system. In our model the motor population consists of nodes (i.e., neurons) encoding desired angular velocities, i.e., the activity of a given node corresponds to a particular desired angular velocity for the robot. For instance, the leftmost node corresponds to turning left at the maximum rate, the central node corresponds to straight ahead, and so on.

We have used the neural network of Fig. 1 to train a Khepera miniature mobile robot (K-Team SA, Preverenges, Switzerland) to avoid obstacles and approach lights. The robot is trained by allowing it to make random movements in a cluttered environment. Whenever the robot collides with an obstacle during one of these movements (or comes very close to it), the nodes corresponding to the largest (closest) range sensor measurements just prior to the collision will be active. Activation of the drive node (which in this case represents aversion) allows two different kinds of learning to take place: the learning that couples sensory nodes (infrared or ultrasound) with the drive node (the collision), and the learning of the angular velocity pattern that existed just before the collision. Likewise, the robot's light sensors can be used to learn light approaching behavior by activating a second drive node (appetitive drive) whenever the light sensors read a significant increase in light level. The only difference between approach and avoidance learning is that activation of the approach drive node leads to excitatory connections to the population that generates angular velocities, while activation of the avoidance drive node leads to inhibitory connections. Hence, as a result of its own experiences, the robot learns to generate more often those movements that lead to increased light levels, while suppressing those movements that lead to collision.

Additional details on this neural network can be found elsewhere [7]. Figure 2 shows some of the results. Each panel shows the Khepera robot in its environment as seen from an overhead camera. A workstation tracks the robot's

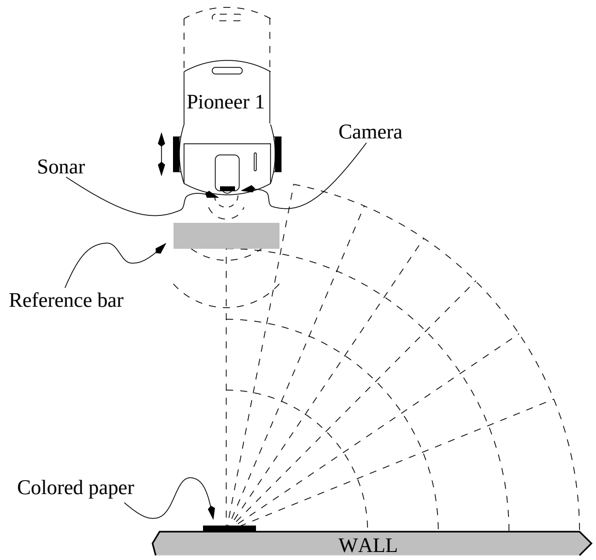


Figure 3: Top view of the experimental setup, showing the Pioneer and the grid of points used for the experiments.

movements through the camera, as shown in the figure by the small dots leading up to the robot's current position. Figure 2(a) shows a typical path followed by the Khepera as it avoids obstacles in its way. Figure 2(b) shows the path followed by the Khepera as it tries to track a flashlight moved around the environment.

The model we have summarized here has proven to be robust, fast, and independent of the type of robot on which it operates: we have been able to replicate our results on the Pioneer 1 (ActivMedia Inc., Peterboro, NH) without any modifications to the neural network. Furthermore, the robot can learn simultaneously the approach and avoidance behaviors, sorting out which sensors are associated with range or light sensing [7].

3 Visual looming

Visual looming is a method for extracting the depth of an object visualized through a camera or a biological vision system. The looming algorithm, which requires only a single camera (or eye), is based on the relationship between displacements of the observer relative to an object, and the resulting change in the size of the object's image on the focal plane of the camera (or retina).

The looming algorithm has been described in prior reports [13, 19, 8]. However, its usefulness for inexpensive, robust ranging has not been realized widely. This is partly because looming requires tracking of an object, which in turn requires segmentation and tracking in a real image, a challenging problem to solve in real-time on a mobile platform. Another probable reason for the relative anonymity of this algorithm is because ranging can be performed cheaply and easily with sonar. We have studied both of these issues, and found that under certain conditions looming is a useful complement to sonar ranging.

As a first step, rather than attempting segmentation on full images, we have used a real-time color-tracking algorithm to track objects of a user-selectable color. Although this solution places some restrictions on the robot's environment, it has allowed us to side-step a number of tangential issues while concentrating on the core features of the looming algorithm itself.

Our results have shown [20] that visual looming can be especially useful for real-time localization of objects that are slanted relative to the robot's line-of-sight, a condition that is notoriously difficult for the commonly used ultrasonic range sensors. To determine this, we have designed an experimental setup as shown in Fig. 3. A Pioneer 1 mobile robot (ActivMedia Inc., Peterboro, NH) is placed at one of several points located at various distances and angles from a colored piece of paper that is attached to the wall at the same height as the robot's camera. Starting at each point, the robot moves backward and forward toward the paper to calculate the position of the paper with the looming algorithm. At the same time, the distance to the wall is estimated using the Pioneer's frontal sonar. The sonar and looming measurements are repeated twenty times at each point to obtain a meaningful average.

Figure 4 illustrates representative results, comparing the distance estimates obtained with looming (using either the width or the height of the tracked object) and with sonar. It is clear that when the wall (and thus the paper on

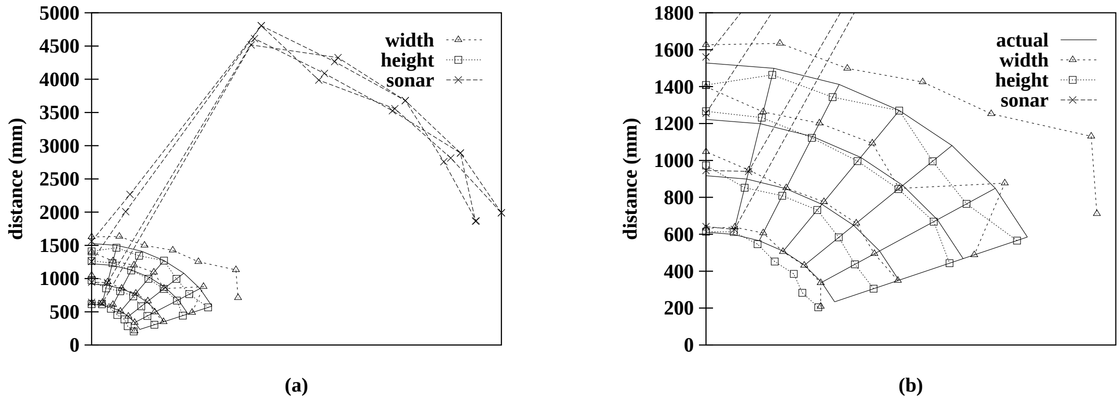


Figure 4: Sample results from a comparison of looming and sonar. Each plot shows (in polar coordinates) the distance between the Pioneer and the object being tracked. The solid grid in each figure shows the points at which the Pioneer was located, at each of seven angles (0–67.5 deg in 11.25 deg increments) and distances of 611, 917, 1223 and 1528 mm (see Fig. 3). At each point, the distance is measured with sonar and with looming using either the width or height measurement of the tracked object. Each data point is the average of twenty measurements. The plot in part **(b)** shows the same data without the sonar to provide additional details on the accuracy of the looming measurement.

the wall) is slanted relative to the robot's line-of-sight, the sonar readings become completely unreliable, while the looming algorithm maintains a high degree of accuracy. This is especially true for the distance measurements obtained with the height, which is not subject to foreshortening effects as the angle between the robot's heading and the object becomes steeper.

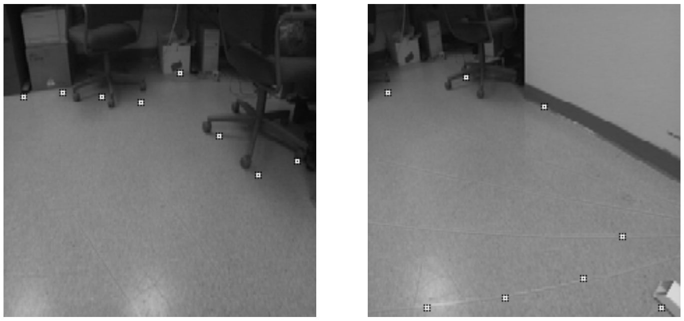
More recently, we have carried out noise analysis [21] to determine the theoretical limits of the accuracy of the looming algorithm in the presence of realistic noise. This analysis, which was corroborated by experimental data, shows that the localization accuracy increases rapidly as a function of the displacement toward or away from the object being tracked.

4 Sensor fusion for ranging and localization

Humans and animals rarely use a single type of sensor. Even the simplest types of recognition and localization appear to rely on information from a variety of sensors, as demonstrated from both behavioral and anatomical studies. The ability to fuse information from multiple sensors would be highly beneficial to the field of robotics. In fact, sensor fusion is a topic of great current scientific interest [9, 17, 18]. It has been shown recently that the ARTMAP neural network can be adapted to accommodate inputs from multiple sensors of the same or different modalities. The ARTMAP neural network [5, 4] combines two unsupervised ART networks, one of which learns to categorize the inputs, while the other represents the desired (output) categories. Supervised learning then takes place between input categories and desired outputs (or output categories). This type of neural network has been shown to perform extremely well under a variety of conditions, typically exceeding the performance (both in terms of speed and memory requirements) of backpropagation networks or traditional classification algorithms such as K Nearest Neighbors.

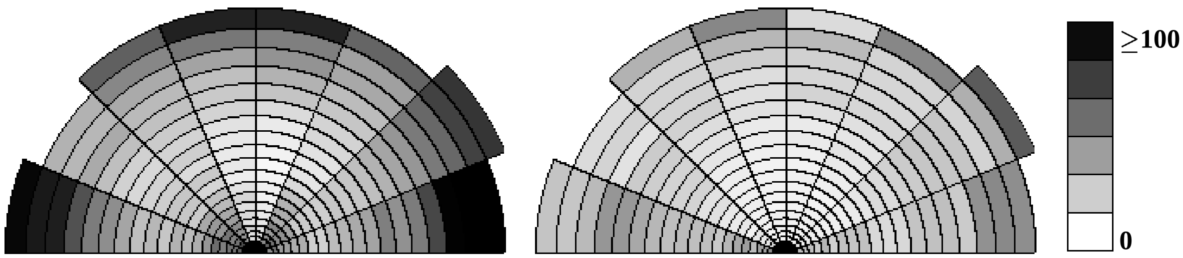
Our goal for this project was to adapt ARTMAP neural networks to perform real-time localization and ranging using multiple sensors on a mobile robot. For this research we used a B14 mobile robot (Real World Interface Inc., Jaffrey, NH), a cylindrical wheeled mobile robot equipped with sonar range finders, infrared proximity detectors, bump sensors and a camera mounted on a pan-tilt platform.

The robot learns to combine visual and sonar information to determine the distance to objects in its environment. The learning process is completely self-supervised: the robot navigates in a room, processing visual and sonar patterns and keeping track of change in the internal odometry signal (i.e., relative displacement). The sensor inputs are categorized in one ART network, while the odometry signal is categorized in another ART network. When a collision is detected through the infrared or bump sensors, ARTMAP learning takes place between the various sensor patterns and the corresponding odometry readings. Hence the ARTMAP network learns to recognize typical sensory patterns that the robot "sees" at regular intervals before colliding with a wall. After sufficient training, activity patterns generated



(a) **(b)**

Figure 5: Localization of obstacles through visual sonar. Each white square on the image shows the point where the visual sonar algorithm detected the nearest edge. **(a)** The visual sonar correctly identifies the location of obstacles in the image. **(b)** Thin pinstriping on the floor confuses the algorihtm, causing false detection.



(a) **(b)**

Figure 6: Localization error as a function of distance and slant angle of objects directly ahead of the robot. **(a)** Minumum return of two frontal sonars. **(b)** ARTMAP-processed. See text.

by the sensors will in turn activate a node corresponding to a certain odometry reading, i.e., the network will recognize the distance of whatever obstacle is in front of the robot. The same scheme can be applied either with multiple sensors of a single modality, or, by adopting the Fusion ARTMAP architecture [2, 1], with sensors from different modalities.

For the visual input, we have implemented a "visual sonar" scheme to obtain an approximate measure of distance through the B14's camera. Visual sonar [12] consists of binning an image into a small number of columns, and then searching for discontinuities in each column starting from the bottom of the image. Under the assumptions that the floor is flat and that obstacles are placed on the floor, the distance of a discontinuity in a given column from the bottom of the image is proportional to the distance of the robot from the corresponding obstacle. Fig. 4 illustrates typical examples in which an image from the B14's on-board camera is passed through the visual sonar algorithm, which identifies the location of obstacles in front of the robot at eight equally-spaced points. In part **(a)** the algorithm correctly identifies all the edges, but in part **(b)** the algorithm detects some spurious edges. These are actually caused by very thin pinstriping (not clearly visible in the printed figure) that was attached to the floor for the looming experiments described earlier.

Visual sonar has the desirable feature of being computationally inexpensive, and of working fairly well in indoor environments. However, it suffers from some shortcomings. First, as illustrated in 4**(b)**, the algorithm is easily fooled by changes in color or texture that may not be generated by obstacles, such as a change in flooring material between a room and the hallway, or sometimes even a cast shadow or light shining through a window. Second, the distance metric obtained through visual sonar is relative, rather than absolute, as it depends on the height of the camera off the floor and on the camera slant.

In some ways, regular sonar and visual sonar have complementary strengths and weaknesses: for instance, if a robot is approaching an open door where there is a change in floor material, the visual sonar will incorrectly detect an obstacle, while the normal sonar will correctly identify the space as being unoccupied. On the other hand, if the robot is approaching a filing cabinet that is slanted diagonally, the regular sonar will not "see" it, while the visual sonar will.

By fusing the information obtained from the sonars and from the visual sonars, the ARTMAP neural network is able to generate correct distance predictions more often than using either sensory modality alone. We have systematically tested sensor fusion by selecting different combinations of sonar and visual sonar sensors, and training ARTMAP on real data collected on the B14 as it roamed around an enclosed area for several hours. We obtained the best results (lowest cumulative error) combining the two frontal sonars with the four central columns of visual sonar.

Figure 6 shows the data obtained with this sensor combination, and compares it with the data obtained with a

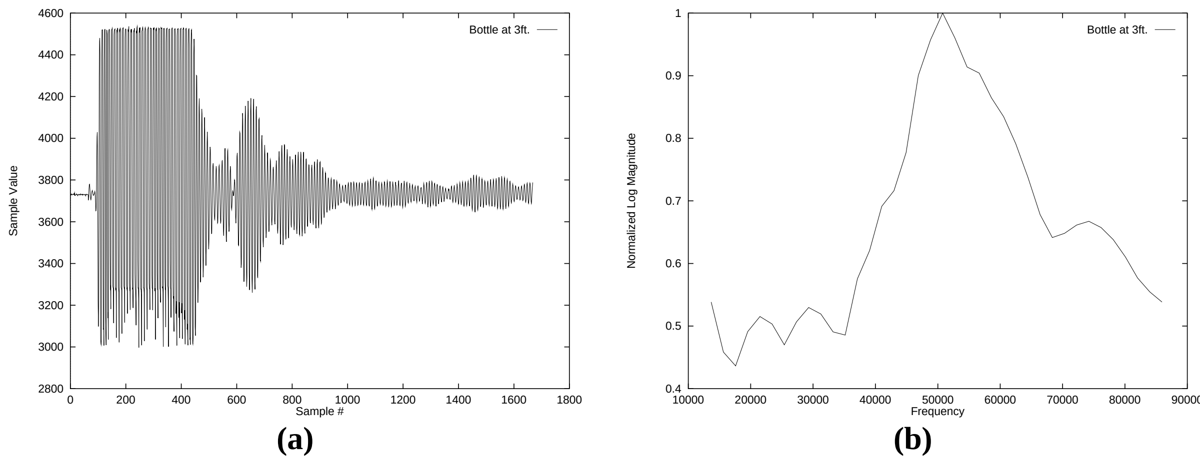


Figure 7: Typical data of the type used in this experiment, including (a) the echo from a water bottle, and (b) the power spectral density function of the echo.

simple model in which the minimum reading from the two frontal sonars is considered to be an accurate reading (based on the observation that sonar often does not "see" an obstacle, thus returning the maximum range value). Approximately 10,000 data points are represented in each semicircular plot. Distance straight ahead from the robot is represented radially, while the slant of the object relative to the robot's heading is represented by the different angular sectors. The shading of each cell corresponds to the error in measuring the distance to objects located straight ahead of the robot, at a distance and slant given by the position of the cell in the semicircle. Lighter shades represent a smaller cumulative error. Figure 6(a) shows that, as expected, sonar ranging is noisy at high slant angles or large distance. Figure 6(b) shows that the ARTMAP pre-processing greatly reduces noise at all distances and angles. Further details, including an application of this sensor fusion scheme to evidence grids, can be found elsewhere [15, 14].

5 Object recognition through sonar

Sonar has been used extensively in robotics for obstacle detection, ranging, and mapping. However, typical applications do not exploit the full range of information carried by sonar echoes, focusing only on the time of arrival of the first detected echo [3]. The remarkable sonar-based perceptual capabilities of animals such as bats and dolphins [10, 16] suggest that sonar can convey detailed information about the environment.

We have developed an object recognition system based on the frequency content of ultrasonic echoes [22]. Our results demonstrate that sonar can be used as a low-cost sensor for real-time object recognition on mobile robots. The system consists of a Polaroid sonar, a data acquisition board and a LINUX-based host computer. The fuzzy ARTMAP neural network [6] is used to recognize objects on the basis of the frequency content of the echoes they generate.

We trained our sonar recognition system to identify five different objects: a wall, an office chair, two cardboard boxes (a small one on top of a bigger one), a metal cylindrical trash can and a 5-gallon plastic water bottle (full of water). Training consisted of placing each object at various distances from the sonar's beam, collecting ten echoes for each object at each position, extracting time-windowed frequency information (power spectral density), and using the frequency information from five (randomly chosen) of the ten samples to train the ARTMAP neural network for each object at each position. The remaining five samples from each object/position are used for testing. Figure 7(a) shows an example of an echo returned from the water bottle at a distance of 3ft, while Fig. 7(b) shows the power spectral density obtained from sampling the echo for approximately 1msec at a frequency of 500KHz.

We have tested the ARTMAP network's ability to recognize the five objects at 11 discrete distances between 2.5 and 7 feet in 0.5-ft increments. This yields a total of 1,100 samples, of which 550 were used for training and the other 550 for testing. Table 5 shows the basic results as a confusion matrix. In this example, classification is 88.4% correct (243 out of 275 correct) over all objects, with individual objects being classified with accuracies ranging between 83.6% (46/55 for the wall and the trash can) and 100% (55/55 for the chair). When we adjusted the ARTMAP's vigilance parameter, we were able to achieve overall accuracy of up to 96.4% on this recognition task.

We have extended these results to test accuracy when training and testing were performed at different distances. As

	wall	chair	boxes	can	bottle	
wall	**46**	0	5	1	1	53
chair	0	**55**	0	4	0	59
boxes	9	0	**49**	2	1	61
can	0	0	1	**46**	6	53
bottle	0	0	0	2	**47**	49
	55	55	55	55	55	**243**

Table 1: Confusion matrix for the basic classification results. The numbers in each column indicate, for each real object, how many times the ARTMAP networks recognized the various objects. For example, when presented with 55 returns from the wall (leftmost column), ARTMAP correctly recognized ”wall” 46 times, but incorrectly recognized ”boxes” 9 times. The boldface numbers along the diagonal thus indicate the number of correct guesses for the five objects, while the number in the lower right corner is the total number of correct guesses (in this case out of a maximum possible total of 275).

expected, recognition accuracy degrades as the distance between training and testing positions increases. For instance, when training at one-foot increments between 2 and 7 feet and testing at 1ft increments between 2.5 and 6.5 feet, overall accuracy decreased to 71.6%. Training at even distances (2, 4, 6ft) and testing at odd distances (3, 5, 7ft) further reduced overall accuracy to 52.0%. This is not surprising if one considers the profound influence that distance has on the characteristics of the echoes: changing distance influences how much and what parts of the object will reflect echoes at what angles, and whether surrounding objects, such as the floor, will influence the reflected ultrasonic signal. Additional details about this project can be found elsewhere [22].

Our results suggest that sonar may be useful as a sensor for simple object recognition. While sonar will never be able to replace vision, we have shown that it is possible to achieve impressive recognition results on simple objects using commonly found hardware, inexpensive components, and low-cost computations that can easily run in real-time on mobile robots.

6 Conclusions

This article has summarized several research projects that have been undertaken in the Neurobotics Laboratory of Boston University. The projects cover a broad range of topics, and utilize a variety of techniques. The common thread to all these projects is our commitment to simple, fast, robust solutions to typical problems in mobile robotics. By minimizing assumptions about the robot or the environment, we are able to achieve a high degree of platform independence, making it possible to evaluate our systems on a variety of robotics platforms.

7 Acknowledgments

Various parts of this work were supported by DARPA, the Office of Naval Research, and the Naval Research Laboratory through grants ONR-00014-96-1-0772, ONR-00014-95-1-0409 and ONR-00014-95-1-0657.

References

[1] Y.R. Asfour, G.A. Carpenter, S. Grossberg, and G.W. Lesher. Fusion ARTMAP: a neural network architecture for multi-channel data fusion and classification. In *Proceedings of the World Congress on Neural Networks*, volume II, pages 210–215, 1993.

[2] Y.R. Asfour, G.A. Carpenter, S. Grossberg, and G.W. Lesher. Fusion ARTMAP: an adaptive fuzzy network for multi-channel classification. In *Proceedings of the Third International Conference on Industrial Fuzzy Control and Intelligent Systems*, pages 155–160, Piscataway, NJ, 1993. IEEE Press.

[3] J. Borenstein, H. R. Everett, and L. Feng. *Navigating Mobile Robots: Systems and Techniques*. A. K. Peters, Wellesley, Massachusetts, 1996.

[4] G. A. Carpenter, S. Grossberg, N. Markuzon, and J. Reynolds. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3:698–713, 1992.

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

[6] G. A. Carpenter, S. Grossberg, N. Markuzon, J. Reynolds, and D. B. Rosen. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3(5):698–712, 1992.

[7] C. Chang and P. Gaudiano. Application of biological learning theories to mobile robot avoidance and approach behaviors. *Journal of Complex Systems*, 1:79–114, 1998.

[8] B. Dai, J.-P. Liu, and W.-S. Chang. Real-time obstacle detection via subsequent 2d images. In *Proceedings of SPIE*, volume 2463, pages 255–263. The International Society for Optical Engineering, 1995.

[9] D.V. Dasarathy. *Decision Fusion*. IEEE Computer Society Press, Los Alamitos, CA, 1994.

[10] I. E. Dror, M. Zagaeski, D. Rios, and C. F. Moss. Neural network sonar as a perceptual modality for robotics. In Omid Omidvar and Patrick van der Smagt, editors, *Neural Systems for Robotics*, pages 1–16. Academic Press, San Diego, 1993.

[11] S. Grossberg and D.S. Levine. Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, interstimulus interval, and secondary reinforcement. *Applied Optics*, 26:5015–5030, 1987.

[12] I. Horswill. Collision avoidance by segmentation. In *Proceedings of the 1994 International Conference on Intelligent Robots and Systems*, 1994.

[13] K. Joarder and D. Raviv. Autonomous obstacle avoidance using visual fixation and looming. In *Proceedings of SPIE*, volume 1825, pages 733–744, 1992.

[14] S. Martens, G. A. Carpenter, and P. Gaudiano. Neural network sensor fusion for spatial visualization. In *SPIE Conference on Sensor Fusion and Decentralized Control in Robotic Systems*, pages 100–111, 1998.

[15] S. Martens, P. Gaudiano, and G. A. Carpenter. Mobile robot sensor integration with fuzzyARTMAP. In *Proceedings of the International Symposium on Computational Intelligence in Robotics and Automation*, pages 307–312, 1998.

[16] P. W. B. Moore, H L Roitblat, R. H. Penner, and P. E. Nachtigall. Recognizing succcessive dolphin echoes with an integrator gateway network. *Neural Networks*, 4:701–709, 1991.

[17] R. R. Murphy. Sensor fusion. In *The Handbook of Brain Theory and Neural Networks*, pages 857–860. MIT Press, Cambridge, MA, 1994.

[18] R. R. Murphy. Biological and cognitive foundations of intelligent sensor fusion. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1):42–51, 1996.

[19] D. Raviv. Visual looming. In *Proceedings of SPIE*, volume 1825, pages 703–713, 1992.

[20] E. Şahin and P. Gaudiano. Mobile robot range sensing through visual looming. In *Proceedings of the International Symposium on Computational Intelligence in Robotics and Automation*, pages 370–375, 1998.

[21] E. Şahin and P. Gaudiano. Visual looming as a range sensor for mobile robots. In *From animals to animats 5*, pages 114–119, 1998.

[22] W. W. Streilein, P. Gaudiano, and G. A. Carpenter. A neural network for object recognition through sonar on amobile robot. In *Proceedings of the International Symposium on Computational Intelligence in Robotics and Automation*, pages 271–276, 1998.