AI Magazine Volume 18 Number 1 (1997) (© AAAI)

Dynamic Object Capture Using Fast Vision Tracking

Randy Sargent, Bill Bailey, Carl Witty, and Anne Wright

■ This article discusses the use of fast (60 frames per second) object tracking using the COGNACHROME VISION SYSTEM, produced by Newton Research Labs. The authors embedded the vision system in a small robot base to tie for first place in the Clean Up the Tennis Court event at the 1996 Annual AAAI Mobile Robot Competition and Exhibition, held as part of the Thirteenth National Conference on Artificial Intelligence. Of particular interest is that the authors' entry was the only robot capable of using a gripper to capture and pick up the motorized, randomly moving squiggle ball. Other examples of robotic systems using fast vision tracking are also presented, such as a robot arm capable of catching thrown objects and the soccer-playing robot team that won the 1996 Micro Robot World Cup Soccer Tournament in Taejon, Korea.

The place was the 1996 Annual AAAI Mobile Robot Competition and Exhibition, held as part of the Thirteenth National Conference on Artificial Intelligence (AAAI-96) in Portland, Oregon. The goal was to demonstrate a robot that autonomously collects 15 tennis balls and 2 quickly and randomly moving, self-powered squiggle balls and delivers them to a holding pen within the allotted time.

Our robot covers less than a square foot (.09 m^2) of floor space, has a gripper slightly larger than a single ball, and has a high-performance vision system. It collected all the balls and received a perfect score for the event (figure 1). We attribute most of our success to the COGNACHROME VISION SYSTEM, a portable, high-performance system capable of very fast (60 frames per second) tracking of many objects that are distinguished by color (see http://www.newtonlabs.com/cognachrome/). Our perfect score tied us for first place with another team, led by Sebastian Thrun of Carnegie Mellon University, whose robot also used the COGNACHROME VISION SYSTEM.

The Robot Hardware

The prototype robot we used for this contest is named M1 (figure 2). M1's basic frame is constructed from stock aluminum extrusion to form an open cage 6½" by 8" by 2½" high. Connected to this frame are two driven wheels (forming a simple, differential drive); a caster wheel; eight infrared proximity sensors; eight contact sensors; a gripper; batteries; a small video camera; and the vision system, which also serves as the robot's controller.

Sensors

Sensors on the robot fall into three categories: (1) vision, (2) infrared (IR) obstacle detection, and (3) contact.

Fast Vision Tracking with the COGNACHROME VISION SYSTEM The robot's primary sensor is the Newton Research Labs COGNACHROME VISION SYSTEM (figure 3). This system allows very fast (60 frames per second), accurate tracking of many objects that are distinguished by color. Tracking by color is a natural for this contest: The tennis balls are bright yellow, and the squiggle ball is red. We mark our goal area with a blue square. For our robot, fast position data are instrumental for quickly and accurately servoing to follow and capture the moving squiggle ball with a gripper that is only marginally bigger than the ball itself. M1 uses a small camera with a wide-angle lens mounted on a single stepper motor to permit camera tilt. Camera pan was provided by pivoting the robot itself.

Infrared Obstacle Detection To assist in object or wall avoidance, an array of narrow-beam IR light-emitting diodes (LEDs) are driven one at a time with a modulation of 40 kilohertz (kHz) (figure 4). The reflected IR light is detected with a pair of standard IR remote-control detection modules (Sharp GP1U52X or

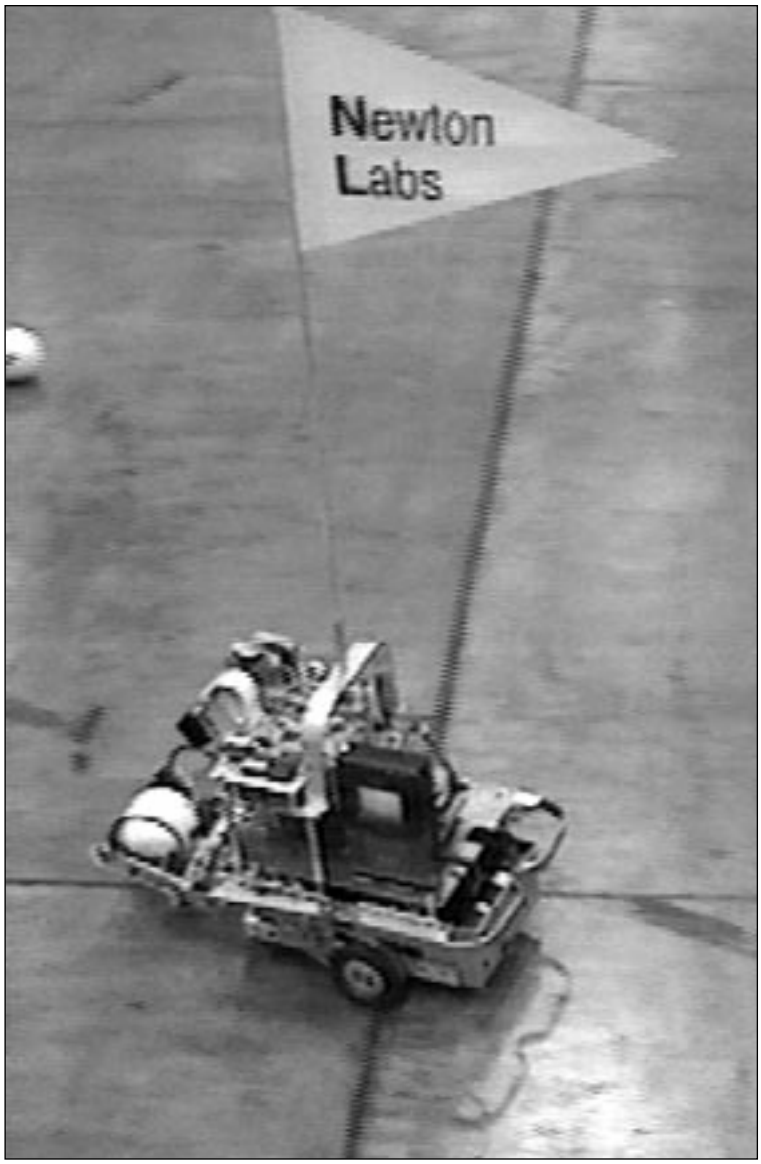

Figure 1. M1 Carries a Tennis Ball during Its Winning Run at the 1996 Annual AAAI Mobile Robot Competition and Exhibition.

equivalent). The directions of the eight LEDs are distributed on a horizontal plane over the forward 180 degrees, with the two IR detectors facing the two forward quadrants. Each LED is fired in turn, and the resulting detector data are latched, providing eight bits in parallel (one bit each direction) to the controlling processor. This system provides reliable obstacle detection in the 8- to 12-inch (20–30 centimeter [cm]) range. Although the system provides only yes-no information about obstacles in the eight directions around the front half of the robot, in fact, crude distance measurements can be made between the robot and large objects, such as walls, by seeing how many of the directions appear to have obstacles. The more directions that appear to have an obstacle, the closer the obstacle probably is. A fairly robust wall-following behavior was constructed using only these sensors.

Contact Sensors Contact sensors around the periphery of the basic frame detect forward, side, and reverse contact (six bits). In addition, contact sensors are placed on the gripper (two bits) because the gripper is the forwardmost component of the robot. The robot also uses the gripper contact sensors to detect and align with the gate before dropping a ball in the goal area.

Actuators and Power

M1 uses a two-wheel differential drive, consisting of a pair of NEMA 23 frame-stepper motors rated at 6.0 volt (V), 1.0 amp, connected independently to the drive wheels with a toothed belt and sprocket combination (figure 5). A third, unpowered caster wheel completes the basic chassis.

An SGS-Thomson L297/L298 stepper motor bipolar chopper drive powers the NEMA 23 motors, with the current limit set to 300 milliampere (ma). Even with this low–current-limit setting, steep accelerations and decelerations are possible. The battery system supplies 30V with a storage capacity of 600 mA-hour to the chopper drive, which results in an upper-limit step rate in excess of 6000 half steps a second. Using stepper motors allows accurate drive control, and this particular implementation appears to result in good performance and low power consumption at both low and high speeds.

A multiple-output switcher-based power supply provides 5V and 12V for the electronic subsystems. An additional 5V linear regulator is connected to the 12V switcher to provide power to more ripple-sensitive, but lower-power demand, electronics. M1 uses a switching power supply because its efficiency helps to lower power consumption and increase battery life.

A small gripper is mounted on the front of the robot (figure 6). To capture and keep the self-propelled squiggle ball, a gripper needs to be fast and keep a firm grip (otherwise, the squiggle ball wiggles free). Grasping and holding a tennis ball is comparatively easy. To simplify both construction and operation, the gripper is built with a single activating motor, a standard model aircraft servo motor. The single motor actuates both the grasp and lift actions in sequence—the lift only happens once the gripper has closed on the

object, regardless of the size of the object (the tennis balls and squiggle balls are different sizes). This dual actuation is accomplished by attaching the motor's pull point such that the grasp action is favored over the lift action. Once the grasp tightens on the ball, continued motor action lifts the ball (figure 7).

The Robot Software

In many applications, the COGNACHROME VISION SYSTEM outputs its tracking data to another central processing unit. However, for this robot, we decided to interface the robot sensors and actuators to spare input-output on the vision board and write our control software on the vision board itself. We wrote a fairly simple, reactive controller for our robot.

Reactive High-Level Control

With the hard part (the vision tracking) already taken care of by the vision system's built-in functions, we spent several weeks (including a few days at the last minute at AAAI-96) writing and testing a simple, reactive control system.

The control system has four basic states: (1) find and approach ball, (2) lift ball, (3) find and approach goal, and (4) drop ball. Each state has several substates, as shown in figure 8.

Substates of Interest Two substates of interest are approach ball and follow wall.

Approach ball: Approach ball is active when a ball is seen in the Find and Approach Ball state, and the ball isn't already within gripper grasping range. If more than one ball is seen, the closest ball is generally chosen, with some hysteresis to prevent oscillation between two balls of similar distance.

M1 must approach a ball in such a way that it enters the gripper area from the front. If the ball is directly to the left or right of the gripper, M1 backs up until the ball clears the gripper's side. Otherwise, M1 approaches the ball with a simple feedback loop: Set M1's rotational velocity to be proportional to the angle required to bring the ball directly in front of M1. If the ball is close enough to being directly in front of M1, move forward with a velocity inversely proportional to some function of the angle error.

Follow Wall (in Find and Approach Ball): While the follow-wall substate in the Find and Approach Ball state is active, M1 will stop and pivot back and forth at a certain period. (During the first half of the contest, M1 pivots every 12 seconds, and during the second half, M1 pivots every 6 seconds.) The purpose of the wall follow is to help guarantee that the entire region is searched. However, in larger rooms, following the wall isn't adequate enough to search the entire room. The pivot behavior forces M1 to look toward the center of the room every so often, extending the distance from the wall at which balls can be seen.

Although it isn't shown in the diagram, each of the four states has special time-outs to try to detect if the robot isn't making progress. In this case, the robot might stop and then start again (in case a stepper motor

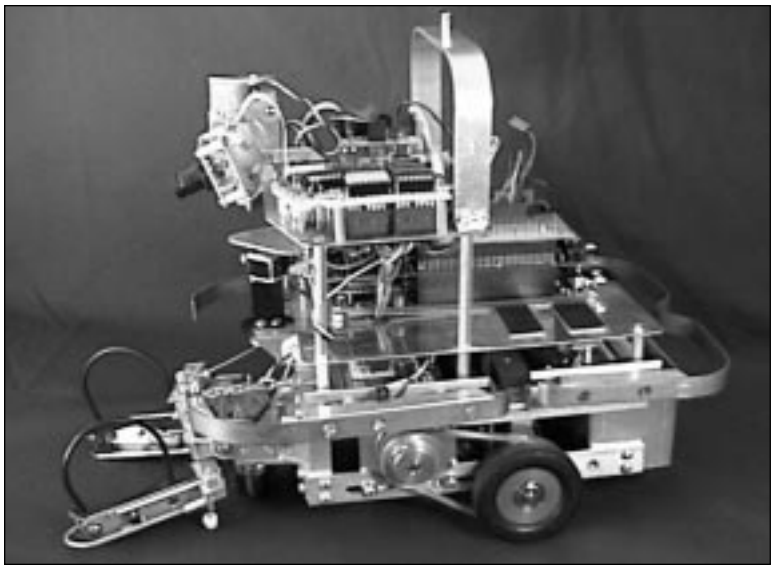

Figure 2. Our Robot, M1.

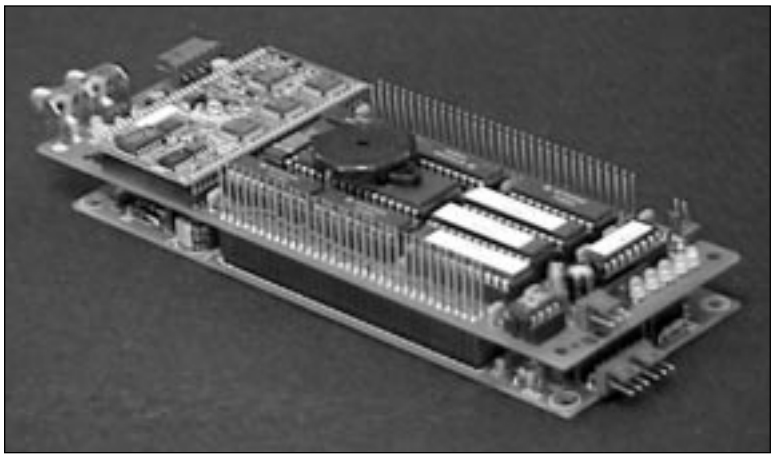

Figure 3. The COGNACHROME VISION SYSTEM.

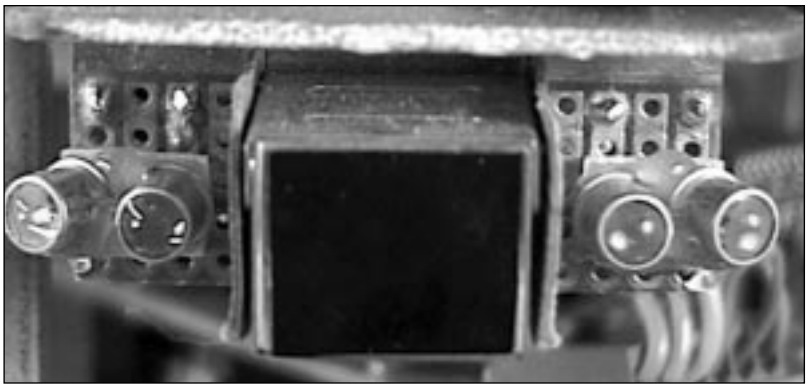

Figure 4. M1's Right Half Infrared Sensor Array.

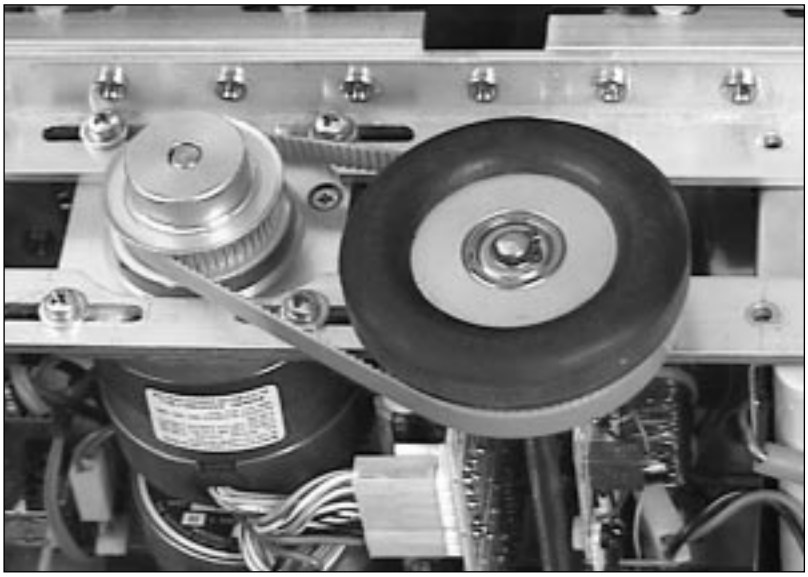

Figure 5. M1's Drive Train.

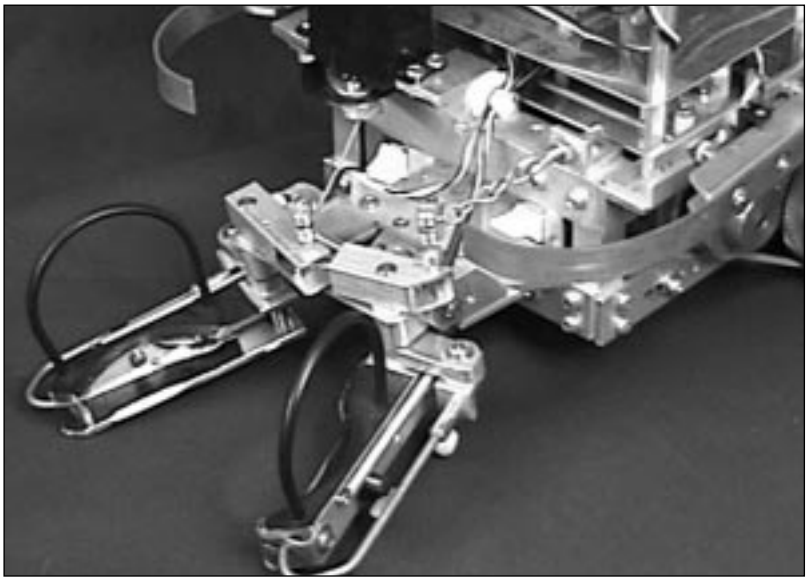

Figure 6. M1's Gripper.

had stalled) or might back up (in case the robot had somehow gotten itself into a tangle of some sort).

Camera Calibration

M1 uses the vision system to detect balls on the floor and the blue marking on the gate. Given the location of a ball in the field of view and the assumption that the ball is on the floor, M1 can compute the position of the ball relative to the robot.

M1 uses a fairly wide-angle lens (about 90 degrees). Such a lens results in a pronounced fish-eye effect. Typically, we make the simplifying assumption that the (x,y) coordinates returned from the vision system map linearly onto a virtual plane that is perpendicular to the axis of the camera. However, for this application, we decided we needed more accuracy. (Given that we find ball positions by computing the intersection of the floor with a line from the camera and that the camera is fairly close to the floor, small angle errors can lead to large position errors.)

Therefore, we needed to calibrate the camera. That is, we wanted a function that takes (x,y) coordinates from the vision system and returns coordinates in a physical coordinate system we could deal with. (We actually use spherical coordinates rather than planar.) We started to deduce the proper mathematical form of the function describing this mapping for the particular wide-angle lens we used but found it was easier (and more accurate) to just use least squares fits to create two bivariate quadratics (where the variables are the x and y coordinates), one for the horizontal angle and one for the vertical angle.

To generate the calibration data for the least squares fit, we set up a vision target a distance away from the robot. We then had the robot pivot from side to side and rotate the camera up and down in a predefined grid pattern, recording the (x,y) coordinates of the target given by the vision system at each step. (The target was far enough away to allow the simplifying assumption that the camera did not change position, only orientation, despite the fact that the camera was not on the robot's pivoting axis.)

This method of gathering the data worked well because of our precise control of the robot's position and camera angle (made possible in part by using stepper motors).

Low-Level Motor Control

M1 uses stepper motors to drive its wheels. One problem with stepper motors is that if you try to run them past their limits (run

them too fast or accelerate or decelerate too quickly), they stall. M1 has no stall-detection sensors. M1 does have stall-recovery behaviors in place (that is, if the control software decides that no progress has been made for long enough, it will slow to a stop, which recovers from the stall), but it is much better to avoid stalls in the first place.

For this reason, there is a layer of software between the high-level control and the motors. Whenever the high-level control software commands a speed, the low level smoothly accelerates or decelerates to this speed, within the safety parameters of the motors.

Other Applications for Fast Vision Tracking

Making a winning entry for the Clean Up the Tennis Court event was made much easier by having a vision system that was capable of quickly tracking targets of interest. We believe that fast vision tracking has the potential to help many other applications as well. Included here are an assortment of projects for which the COGNACHROME VISION SYSTEM is currently used.

Playing Robot Soccer

We entered (and won) the first International Micro Robot World Cup Soccer Tournament (MIROSOT) held by the Korea Advanced Institute of Science and Technology in Taejon, Korea, in November 1996 (figure 9) (Sargent et al. 1996). We used the COGNACHROME VISION SYSTEM to track our three robots (position and orientation), the soccer ball, and the three opposing robots. The 60-Hz update rate from the vision system was instrumental in our success; other teams obtained robot and ball position data in the 2- to 10-Hz range. Our robots could literally run circles around their opponents.

Because of the small size of the robots (each fit into a cube 7.5 cm on a side), we opted for a single vision system connected to a camera facing down on the field instead of a vision system in each robot. (In fact, the rules of the contest required markings on the top of the robot that encouraged off-board vision; all but one of the teams likewise used a single camera above the playing field. The odd team out decided not to use vision at all, which severely limited its capability.)

Please see http://www.newtonlabs.com/soccer for video footage, stills, and technical information about our entry. The MIROSOT organizers' site, http://www.mirosot.org, describes the contest.

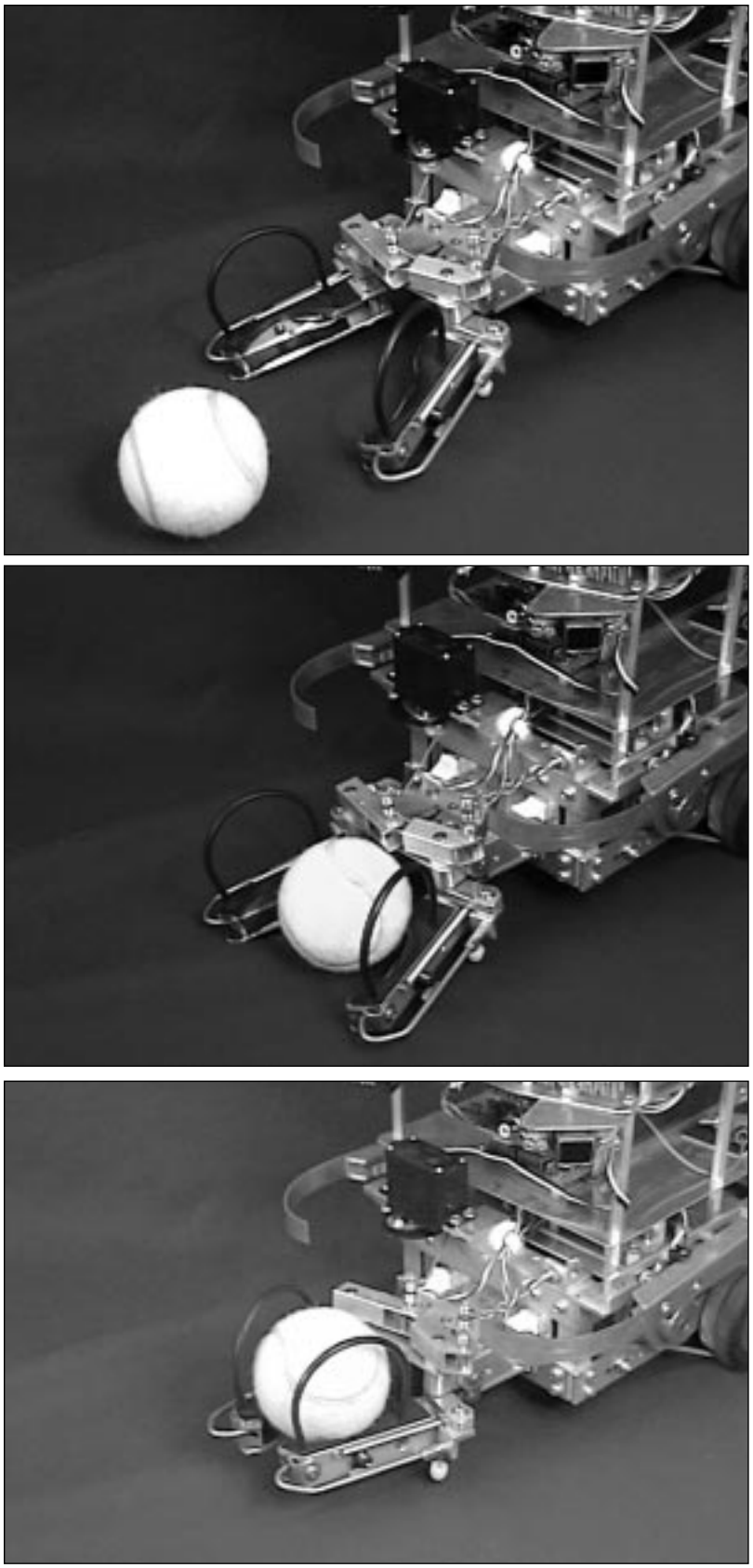

Figure 7. Sequence of M1 Picking Up a Tennis Ball.

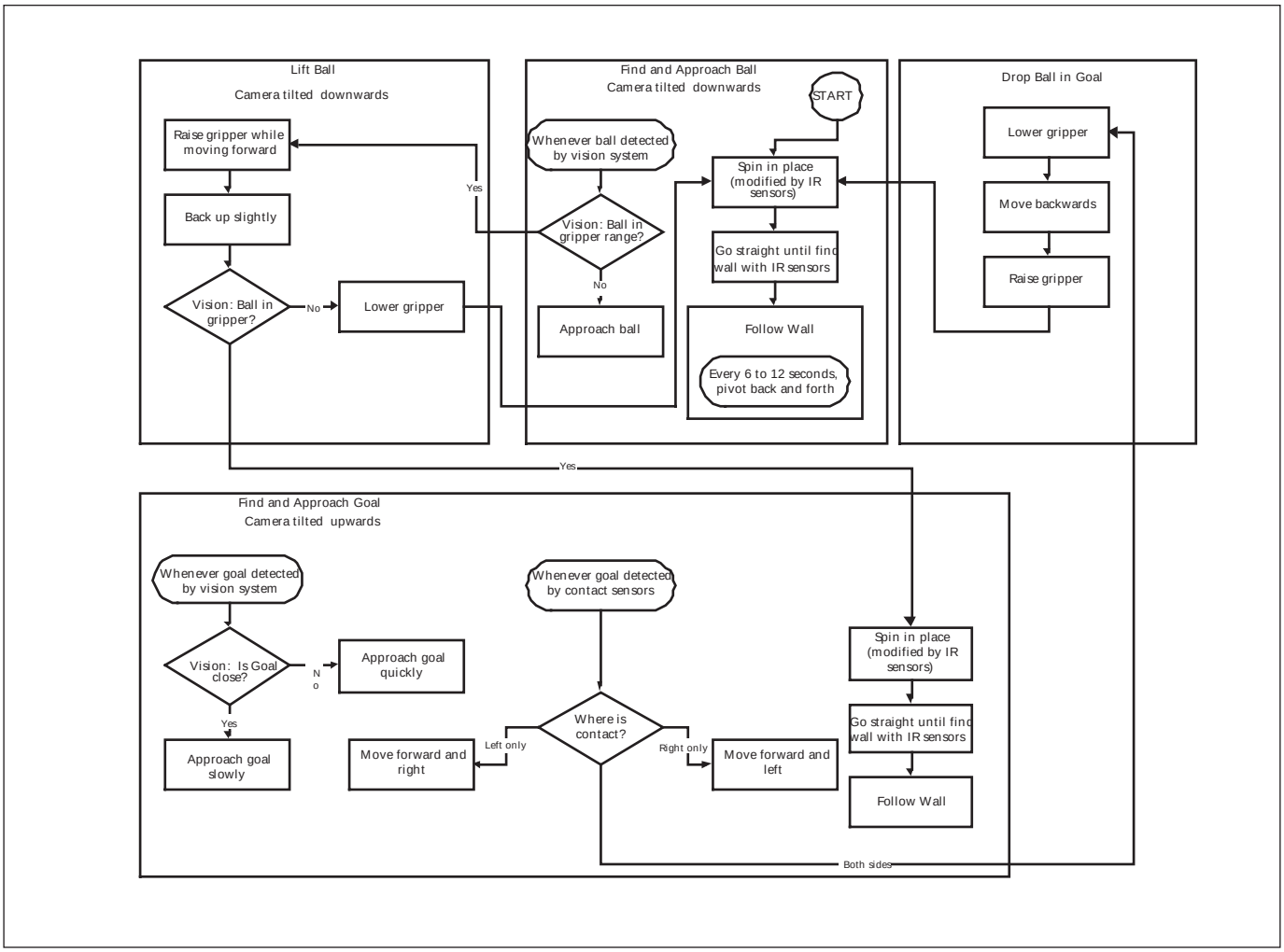

Figure 8. Detailed State Diagram for High-Level Control of M1.

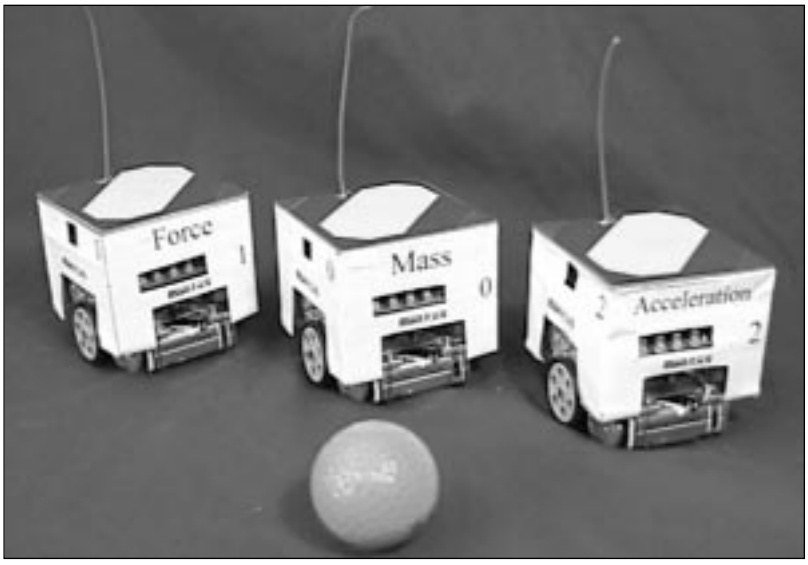

Figure 9. FORCE, MASS (*the Goalie), and* ACCELERATION *are Three Members of the Newton Labs World Champion Robot Soccer Team.*

In the foreground is the soccer ball (actually an orange golf ball).

Catching Balls and Paper Airplanes

Two COGNACHROME VISION SYSTEMS (Hong 1995; Hong and Slotine 1995) were integrated in the new version of the Adaptive Robot-Catching Project led by Jean-Jacques Slotine of the Massachusetts Institute of Technology (MIT). The project uses an advanced manipulator and fast-eye gimbals developed under Kenneth Salisbury of the MIT AI Lab.

Using two-dimensional stereo data from a pair of COGNACHROME VISION SYSTEMS, they predict the three-dimensional trajectory of an object in flight and control their fast robot arm (the whole-arm manipulator [WAM]) to intercept and grasp the object. (Please see http://www.ai.mit.edu/projects/wam/index.html#S2.2 for more information and animations of the arm catching various objects).

Performance Robotics

Performance artist and roboticist Barry Werger creates performance robotics pieces using

Pioneer mobile robots equipped with the COGNACHROME VISION SYSTEM (Please see www.activmedia.com/RealWorld/ for more information. RWI resells the COGNACHROME VISION SYSTEM as the FAST TRACK vision system). By providing the robot and human players with appropriately colored tags, the robots can interact with each other, and humans, at a distance in a theatrically interesting way. Please see www.cs.brandeis.edu/~barry/performance.html for more information (including information about future performances.)

Group Behavior and ocial Interaction of Robots

Maja Mataric, Werger, Dani Goldberg, and Francois Michaud at the Volen Center for Complex Systems at Brandeis University study group behavior and social interaction of robots (please see http://www.cs.brandeis.edu/~agents/projects.html and http://www.cs.brandeis.edu/~barry/research.html for more information). Along with other robots, they use Pioneer mobile robots outfitted with COGNACHROME VISION SYSTEMS.

In conjunction with shorter-range, or less specific, sensors, such as sonar, the Pioneers use color-based tracking to help recognize other robots, obstacles, and goals. Werger says:

> I have combined these two [vision-based long-range obstacle avoidance and vision-based following of intermittently blocked objects] to address some of the problems we have in our mixed robot environment. ...that is, the Pioneers are faster and bigger than our other, more fragile robots; the long range avoidance allows them to keep a safe distance from other robots, even in fairly dynamic environments, when following a dynamic target. The vision allows us to make these distinctions very easily, which the sonar does not [private communication, 22 November 1996, with R. Sargent].

Autonomous Docking of Spacecraft

The University of Maryland (UMD) Space Systems Laboratory and the KISS Institute for Practical Robotics have simulated autonomous spacecraft docking in a neutral buoyancy tank for inclusion on the UMD *Ranger* space vehicle (figure 10). Using a composite target of three brightly colored objects designed by David P. Miller (Miller and Wright 1995), the spacecraft knows its distance and orientation and can servo to arbitrary positions around the target (figure 11). (See http://www.kipr.org/robots/scamp.html for more information and pictures.)

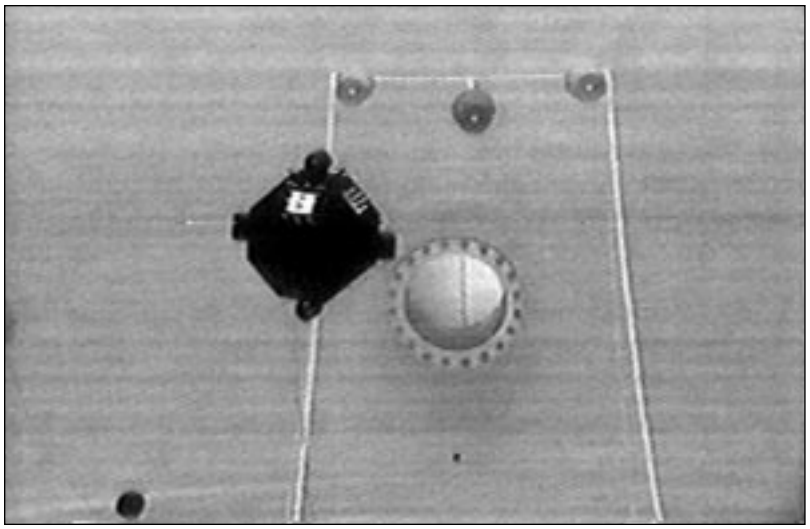

Figure 10. The SCAMP Underwater Vehicle Positions Itself Relative to the Target.

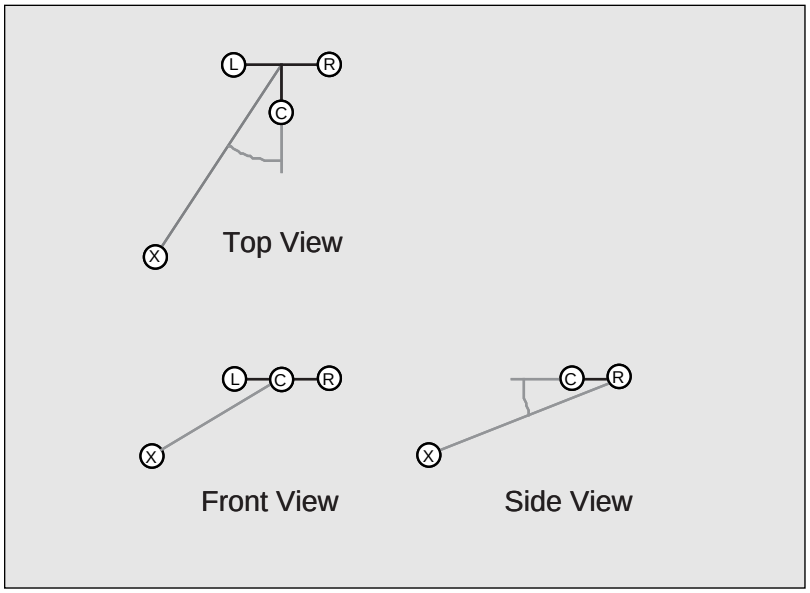

Figure 11. The Target Is Composed of Three Parts: L, C, and R (left, center, and right).

X is the position of the camera. Note that θ is measured in the plane of LRX, not LRC.

$$\theta = \tan^{-1}\left[\frac{a}{b}\left(\frac{2x_C - x_L - x_R}{x_L - x_R}\right)\right] \qquad \phi = \sin^{-1}\left[\frac{a}{b}\left(\frac{2y_C - y_L - y_R}{x_R - x_L}\right)\right] \qquad r = \frac{K\cos\theta}{x_R - x_L}$$

Relationships yielding the 3-dimensional position of the robot relative to the target, given the 2-dimensional positions (in camera space) of the three target elements as viewed from the robot. x_l, y_l, x_c, y_c, x_r, y_r are the positions, in camera space, of the left, center, and right targets, respectively, as viewed from the robot.

Conclusions

We have found through the AAAI competition, as well as many other applications, that a fast vision tracking system can be a useful sensor for robotic systems. For this particular contest, fast vision tracking worked especially well. The data from the vision system were appropriate for the problem at hand and allowed us to use a simple reactive system for control. The vision system's fast update rate was crucial in being able to follow and catch the squiggle ball. We look forward to future opportunities to apply fast vision tracking to other problems.

References

Hong, W. J. 1995. Robotic Catching and Manipulation Using Active Vision. M. S. thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology.

Hong, W. J., and Slotine, J. J. E. 1995. Experiments in Hand-Eye Coordination Using Active Vision. In Proceedings of the Fourth International Symposium on Experimental Robotics, ISER'95, 30 June–2 July, Stanford, California.

Miller, D. P., and Wright, A. 1995. Autonomous Spacecraft Docking Using Multi-Color Targets. In Proceedings of the Sixth Topical Meeting on Robotics, February, Monterey, California.

Sargent, R.; Bailey, B.; Witty, C.; and Wright, A. 1996. Use of Fast Vision Tracking for Cooperating Robots in the MIROSOT Micro-Robot World Cup Soccer Tournament. In Proceedings of the Micro-Robot World Cup Soccer Tournament, MIROSOT '96, 9–12 November, Taejon, Korea.

Randy Sargent holds the title of design engineer at Newton Research Labs in Bellevue, Washington. He received a B.S. in computer science at the Massachusetts Institute of Technology (MIT) and an M.S. in media arts and science from the MIT Media Laboratory. Formerly holding titles of research scientist and lecturer at MIT, he is one of the founders of the MIT Lego Robot Contest (AKA 6.270), now in its seventh year.

Carl Witty holds the title of research scientist at Newton Research Labs in Bellevue, Washington. He received a B.S. in computer science from Stanford University and an M.S. in electrical engineering and computer science from the Massachusetts Institute of Technology. A member of the winning team in the 1991 Association of Computing Machinery Programming Contest, his interests include robots, science fiction and fantasy, mathematics, and formal methods for software engineering.

Bill Bailey holds the title of design engineer at Newton Research Labs in Bellevue, Washington. He specializes in analog and digital electronics, software, and mechanical design. When not working on robotics, he also works at a small software firm in Redmond, Washington.

Anne Wright is president of Newton Research Labs in Bellevue, Washington. The original developer of the COGNACHROME VISION SYSTEM, she received a B.S. and an M.Eng. in computer science from the Massachusetts Institute of Technology (MIT). She also helped lead and develop technology for the MIT Lego Robot Contest from 1992 to 1994.