# Autonomous Visual Control of a Mobile Robot

**Michael R. Blackburn and Hoa G. Nguyen**
Naval Command, Control and Ocean Surveillance Center
Research, Development, Test & Evaluation Division
San Diego, CA 92152-7383

## Abstract

An autonomous mobile robot with a vision based target acquisition system must be able to find and maintain fixation on a moving target while the system itself is in motion. This capability is achieved by most animate systems, in addition to man, but has proven to be difficult for artificial systems. We propose that efficient and extensible solutions to the target acquisition and maintenance problem may be found when the machine sensor-effector control algorithms emulate the mechanisms employed by biological systems. In nature, motion provides the foundation for visual target detection, acquisition, tracking and trailing, or pursuit. We present in this paper a summary of some simple and robust visual motion based mechanisms we have developed to solve these problems, and describe their implementation in an autonomous visually controlled mobile robot.

## 1 Introduction

The objective of this research is to develop an autonomous mobile robot capable of visual target detection, tracking, trailing, and obstacle avoidance. Specifically, the robot is tasked with following a human walking through an office complex. For a demonstration of autonomy, all sensor-effector loops must be completed on the robot, without external assistance in the form of target designation or environmental modeling. The robot must accomplish this task without the aid of any explicit a priori knowledge of the floor plan, or the aid of any special codings or markings in the environment, including any special treatment of the target. Vision will be the only means by which the robot will be permitted to gain information about the external environment. Further, only visual motion information will be used.

## 2 Algorithms

We fitted a mobile robot with video camera, pan and tilt mechanism, on-board computer and biologically based visual-motor control algorithms. The basic information that we made available to the robot controllers through the vision system was motion, contained in the sequence of video frames. Using this information the robot could be able to detect targets while either stationary or in transit. The motion analysis algorithms, developed in earlier work [Blackburn et al., 1987], were enhanced to allow separation of unique target motion from the collateral optic flow accompanying the movement of the robot through a visually complex environment. The modifications included the use of center-surround receptive fields to minimize the optic flow created by the transiting robot and enhance the unique target motion.

### 2.1 Functional Description

Figure 1 diagrams the various visual-motor functions which perform our tracking, trailing and obstacle avoidance tasks. The behavior of the animate target determines the behavior of the robot. Unique motion in the periphery causes a visual reorienting reflex (saccade) which either moves a processing window within the available visual space (small saccades) or the entire camera pan and tilt unit (large saccades), placing the center of the visual field (fovea) on the center of mass of a moving target. A large saccade is also performed when the processing window reaches the limit of the image frame. A large saccade generates a window recentering command.

A smooth pursuit reflex, which takes input from motion in the foveal region, keeps the fovea

Proc. 1994 ARPA Image Understanding Workshop, Monterey, CA; Vol. II, pp. 1143-1150

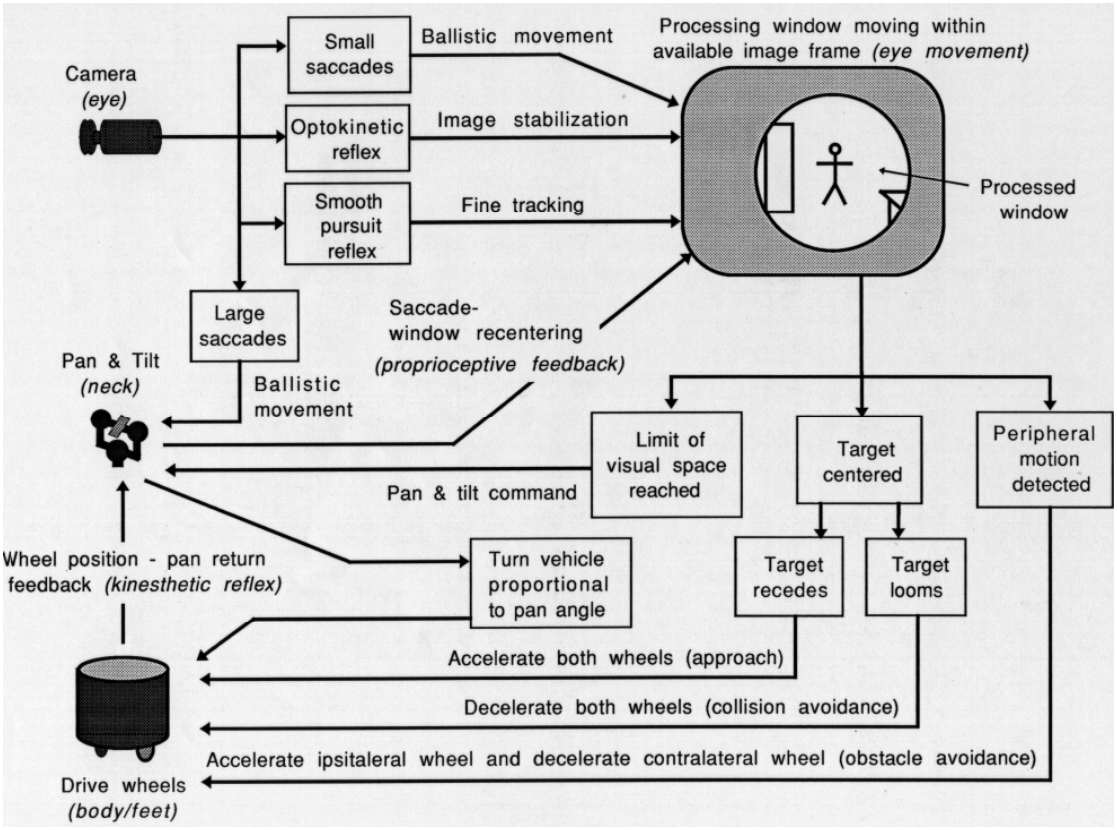


Figure 1. Visual-motor functions and relationships

centered on the acquired moving target. The opto-kinetic reflex, which responds to full field motion, stabilizes the eye when the body is in motion.

Reorientation of the robot to trail an acquired target is accomplished by basing commands to the robot drive motors on the camera pan angle, requiring the robot to drive in the direction of the gaze. This process is analogous to the targeting motion of the eyes, head and body in biological systems.

Trailing is accomplished by triggering forward thrust of the robot when the predominant motion of a centered target is toward the center of the visual field (contracting motion field). Collision is avoided by decreasing forward thrust when the target motion is away from the center (expanding). Obstacle avoidance is achieved by decreasing thrust on the side of the robot opposite to the peripheral motion away from the center of the visual field.

The obstacle avoidance reflex, which is transitory, assumes precedence over the pursuit reflex, allowing the robot to skirt around obstacles in pursuit of a target.

For an in-depth discussion of the biological visual processes from which we derived our algorithms, see Blackburn et al. [1993].

## 2.2 Receptive Fields And Log Polar Mapping

As a basis for motion analysis, sequential frame subtraction is performed. The differences are taken of the current frame (R) and the previous frame (H), resulting in both "on" (B1) and "off" (B0) elements.

$$B0 = \max(0, R - H)$$
$$B1 = \max(0, H - R). \quad [1]$$

The "on" elements indicate light intensity increasing in a localized region while the "off" elements

indicate decreasing intensity. The output matrix is organized into local receptive fields and submitted to a log-polar transformation [Blackburn, 1993a] where the receptive field centers are placed proportionally further apart with their distance from the receptor matrix center, and the receptive field radii are also increased proportionally with the distance.

The log-polar transform is accomplished by:

$$G_{1i,j} = (1/p) * \Sigma_{a,b}(B_{1a,b}); \text{ s.t.} \|(a,b)-(x,y)\| <= RFr \quad [2]$$

where i and j are the coordinates in the transformed map, a and b are coordinates of elements located within the local receptive fields, x and y are locations of the local receptive field centers in the receptor matrix, and p is the variable number of elements in the local receptive fields. RFr is the radius of the local receptor fields, defined by:

$$RFr = \gamma * E, \quad [3]$$

where $\gamma$ is a constant computed as $(2 * (1-\cos(2 * \pi/m)))^{1/2}$ to insure that for m number of local receptive fields for any given eccentricity, the radius of each local receptive field reaches the center of the next local receptive field on the circumference.

The eccentricity (E) of a local receptive field, defined as the location of the field center relative to the center of the receptor matrix, varies exponentially with the serial position from the center along the radius of the receptor matrix (with the constraint of a finite packing density of elements near the center forcing each radius to be at least one element diameter greater than the previous).

$$E = \max (i, \exp (\zeta * (i/n))), \quad [4]$$

where i is the serial distance on a radius from the receptor matrix center (from 1 to n), n defines the number of local receptive fields to be located on a radius from the receptor matrix center, and $\zeta$ = log(N/2) with N/2 representing the number of receptors (or pixel elements) available along the receptor matrix radius.

The x,y locations of the receptive local field centers on the receptor matrix are determined by

$$x = (N/2) - E_{x,y} * \sin \Theta \quad [5]$$
$$y = (N/2) + E_{x,y} * \cos \Theta, \quad [6]$$

where $\Theta$ is incremented from $\pi/2$ to $5\pi/2$ by $2\pi/m$. The locations of receptive field centers from one eccentricity to the next is staggered by $\pi/m$ so that a slightly asymmetric hexagonal matrix of receptive field centers results.

The averaging of pixels in receptive fields emphasizes large-magnitude effects. This is a desirable feature in building reliable artificial vision systems and may have been part of the reason for its adoption by nature.

## 2.3 Motion Analysis

We have found that the log polar transformation, which is also found in biological visual systems, greatly simplifies motion analysis. On the computational surface that has undergone a log-polar transformation, a centered target will cause an optic flow that moves in parallel in one direction for the receding condition, and in parallel in the opposite direction for the expanding (looming) condition.

Peripheral receptive fields are large and set far apart compared to the central receptive fields. Thus, the center of the receptor surface is more sensitive to slow motion, while the peripheral region is more sensitive to fast motion. The direction of motion on the log-polar plane can be assessed using a simple compare-to-threshold approach combined with feed forward facilitation or feedback inhibition relative to the preferred direction of the motion analyzer [Blackburn et al., 1987].

The direction of motion is determined by dynamic filtering. The filter elements (MAI, MAII, and MAIII) are defined by:

$$MAI = C1 * MAI(t-1) + \Sigma_{i,j} G_{1i,j}, \quad [7]$$
$$MAII_{i,j} = C2 * MAII_{i,j}(t-1) + G_{1i,j} \quad [8]$$
$$MAIII_{ui,j} = C2 * MAIII_{ui,j}(t-1) + \Sigma_k (1/k) * G_{1i+k,j} \quad [9]$$
$$MAIII_{di,j} = C2 * MAIII_{di,j}(t-1) + \Sigma_k (1/k) * G_{1i-k,j} \quad [10]$$

where u indicates a filter element supporting the detection of upward motion on the transformed map, d indicates downward motion, i and j index the location of elements, k indexes the offset of

input elements in the +/- vertical directions, and C1 and C2 are constants of persistence (1.0 > C1 > C2 > 0). Upward motion on the transformed map results from motion toward the center of the receptor surface, while downward motion on the transformed map results from motion away from the center. (On- and off-center pathways are processed in parallel until the final output, when their products are combined. These equations are shown only for the on-center activity.)

The input to the motion analysis subnetwork on the subsequent increment of time is then passed through to the direction of motion detectors ($MAIV_{ui,j}$ and $MAIV_{di,j}$) based upon the filter activity,

$$MAIV_{ui,j} = \max (0, C3 * MAI - MAIII_{ui,j}) * G_{1i,j} + \Sigma_k \max(0, MAII_{i+k,j} - C3*MAI) * (1/k) * G_{1i,j}, \quad [11]$$

$$MAIV_{di,j} = \max(0, C3 * MAI - MAIII_{di,j}) * G_{1i,j} + \Sigma_k \max(0, MAII_{i-k,j} - C3*MAI) * (1/k) * G_{1i,j}, \quad [12]$$

where C3 is a gain constant (1.0 > C3 > 0). Equations [7] through [10] are duplicated for the off-center activity, and added to $MAIV_{ui,j}$ and $MAIV_{di,j}$ as in equations [11] and [12].

One output of the motion analysis subnetwork ($MAVI_{ui,j}$ and $MAVI_{di,j}$) is the net positive difference of the opposite direction of motion detectors,

$$MAVI_{ui,j} = \max (0, MAIV_{ui,j} - MAIV_{di,j}) \quad [13]$$
$$MAVI_{di,j} = \max (0, MAIV_{di,j} - MAIV_{ui,j}). \quad [14]$$

Another output of the motion-analysis subnetwork ($MAV_{ui,j}$ and $MAV_{di,j}$) is a measure of the motion contrast between the center and the surround of a local region. The sums of the local motion detectors in a neighborhood are taken for the opposing directions and compared. The largest represents the net or most likely direction of motion due to self movement through the environment. If the direction of motion of the center of the neighborhood is consistent with this net motion, then the center can be ignored, otherwise it likely signals unique target motion.

$$MAV_{ui,j} = MAVI_{ui,j}, \text{ if } \Sigma MAVI_{ui,j} < \Sigma MAVI_{di,j}$$
$$= 0 \text{ else} \quad [15]$$
$$MAV_{di,j} = MAVI_{di,j}, \text{ if } \Sigma MAVI_{di,j} < \Sigma MAVI_{ui,j}$$
$$= 0 \text{ else} \quad [16]$$

The outputs of the MAV elements are sent to the target acquisition subnetwork while the output of the MAVI elements are sent to the approach and avoidance subnetworks (described below).

## 2.4 Target Acquisition, The Saccade Reflex

Targets are detected by a model of the vertebrate optic tectum, using a biased cooperative mechanism between hemifields. The optic tectum determines the center of mass of potential targets and directs motors controlling sensor positioning to bring that center of mass of the potential target to the center of the receptor field. The mechanism employed in the present application differs somewhat from mechanisms previously reported by this group that contribute to the generation of scan paths [Blackburn, 1993b]. Instead of selecting a defined region of the visual space that exceeded all other regions on an activity criterion, the unique motion potentials (from the MAV elements of each receptive field) were weighted by the distance of the receptive field centers from the center of the receptor surface, and integrated separately in each hemifield. This modification brings the model closer to mechanisms implicated by the behavior of amphibians, and somewhat further from mechanisms implicated by the behavior of mammals. The final target location was the vector average computed from the sum of the weighted activity of one or both of the hemifields if that sum exceeded a running global threshold. The advantage of the amphibian model is that it allowed the machine target acquisition subnetwork to select the center of mass of most targets that either occupied space in parts of a hemifield or in parts of both hemifields.

The input to the target detection and centering subnetwork comes from the unique motion detectors ($MAV_{di,j}$ and $MAV_{ui,j}$). These are weighted by the distances of their locations from the center of the receptor matrix and normalized by the sum of their potentials to find the location of the center of activity for target localization. A bias that is proportional to eccentricity is applied to the input to favor peripheral over central targets.

The input is retinotopically distributed and integrated over time, allowing excitation to build up in a local area,

$$OT\_in_{i,j}(t) = C4 * OT\_in_{i,j}(t-1) + W_i * (MAV_{di,j} + MAV_{ui,j}), \quad [17]$$

where C4 is a constant of persistence ($1.0 > C4 > 0$), and $W_i$ is a bias factor that increases with eccentricity (i).

The required X and Y change in receptor matrix orientation (accomplished by camera pan and tilt commands) to center the matrix on a new target are

$$dX = \Sigma_{i,j}(x\_distance_{i,j} * OT\_in_{i,j}) / \Sigma_{i,j} OT\_in_{i,j}$$
$$dY = \Sigma_{i,j}(y\_distance_{i,j} * OT\_in_{i,j}) / \Sigma_{i,j} OT\_in_{i,j} \quad [18]$$

Noise is filtered from the subnetwork by disallowing contributions to dX and dY from one hemisphere if the sum of inputs in that hemisphere ($\Sigma OT\_in_{i,j}$) is less than a dynamic threshold ($\Theta$). The threshold is increased whenever it is exceeded by the sum of inputs. Otherwise it dissipates like all other potentials with persistence in the network.

$$\Theta = C6 * \Theta + C7 * \Sigma_{i,j} OT\_in_{i,j}, \text{ if } \Theta < \Sigma_{i,j} OT\_in_{i,j}$$
$$= C6 * \Theta, \text{ else} \quad [19]$$

where C6 is the threshold persistence and C7 is a gain factor ($1.0 > C7 > C6 > 0$).

## 2.5 Target Tracking By The Smooth Pursuit Reflex

Once acquired, a target must be kept on the center of the receptive field where the resolution is the greatest. The higher pixel density in the center of the receptive field permits the early assessment of the direction of a target that is moving slowly.

The smooth pursuit mechanism receives its input from the motion analysis subnetwork. Due to errors inherent in the mechanical pan and tilt unit, slow pursuit is performed by adjusting the processing window within the available video frame. The rate of change of the video window (dU, dV) is computed by:

$$dU = C8*(dU + \Sigma_{i,j}(x*RFr_{i,j}*MAVI_{di,j})/\Sigma_{i,j} MAVI_{di,j}) \quad [20]$$

$$dV = C8*(dV + \Sigma_{i,j}(y*RFr_{i,j}*MAVI_{di,j})/\Sigma_{i,j} MAVI_{di,j}) \quad [21]$$

where x and y define the quadrant of the location of activity (+/- 1), and C8 is a constant of persistence ($1.0 > C8 > 0$).

## 2.6 Approach/Avoidance Responses

While the target is centered in the window, the forward velocity of the robot can be controlled by the advance or retreat of the target. This motion on the optical Z axis is assessed by the opposite directions of motions on the computational plane in the central region. Any motion toward the center of the receptor plane can be considered as a possible retreat of the target and worthy of an approach response, while bilateral motions away from the center indicate a target whose image is growing larger, probably due to its advance upon the platform, and demand a reduction in forward thrust. These reductions are proportional to the location of the motion on the computational surface, such that peripheral locations generate the largest reductions, contributing to collision avoidance.

While the platform is moving through the environment, unilateral image flows away from the center of the receptor surface in the peripheral region indicate the presence of potential obstacles. The required response is to reduce the thrust on the contralateral drive motor, and increase the thrust on the ipsilateral drive motor. When traveling down a corridor with sufficient pattern contrast on the two walls, such a reflex would tend to keep the platform as nearly in the center of the corridor as possible.

The output of the motion analysis subnetwork ($MAVI_{ui,j}$ and $MAVI_{di,j}$) is also used to control the robot drive motors according to simple rules. Motor commands accumulate and dissipate according to

$$motor_{L,R} = C5 * input(t-1) + input, \quad [22]$$

where C5 is the persistence of the input ($1.0 > C5 > 0$). The input comes from the two hemi visual fields and causes an increase or decrease in thrust in both drive motors.

When either hemi visual field detects motion toward the center (indicating a receding target), thrust is increased to both motors inversely proportional to the absolute value of the distance from the center to the location of the motion on the receptor surface

$$input=+\Sigma_{i,j}(max\_dist - abs(x\_dist_{i,j})*MAVI_{ui,j}), \quad [23]$$

where max_dist is the greatest lateral extent of the receptor matrix. The sign of $x\_dist_{i,j}$ indicates the location of the motion on the left (-) or the right (+) of center.

When both hemi visual fields detect motion away from the center, thrust is decreased to both motors directly proportional to the absolute value of the distance from the center to the location of the motion on the receptor surface

$$input = - \Sigma_{i,j}(abs(x\_dist_{i,j}) * (MAVI_{di,j}). \quad [24]$$

Potential obstacles that are detected by asymmetric optic flow away from the center of the receptor matrix cause increased thrust on the same side (g) and decreased thrust on the side opposite (f) to the optic flow. These changes in thrust are transitory and non-zero only under the conditions of asymmetric optic flow, and during an active forward drive command. The degree of change, resulting in a turn away from the obstacle, is proportional to the net forward thrust.

$$motor_g(t) = motor_g(t\text{-}1) + (motor_g(t\text{-}1)/max\_thrust) * \Sigma_{i,j}(max\_dist - abs(x\_dist_{i,j})) *MAVI_{di,j}, \quad [25]$$

$$motor_f(t) = motor_f(t\text{-}1) - (motor_f(t\text{-}1)/max\_thrust) * \Sigma_{i,j}(max\_dist - abs(x\_dist_{i,j})) *MAVI_{di,j}. \quad [26]$$

## 2.7 Orienting Reflex

The robot will turn toward a translating target based on the disparity between the axis of the camera and the axis of the robot body. The pan disparity is sensed by counters on the pan axle. It is either negative, indicating a target location on the left of the robot axis, zero, indicating a target location in front of the robot, or positive, indicating a target location on the right of the robot axis. This turning reflex is inhibited by the obstacle avoidance reflex if the required turn is in the direction of the obstacle.

The turn command is transient and inversely proportional to the net forward thrust:

$$motor_L = motor_L(t\text{-}1) + pan\_disp * (1.0\text{-}motor_L/max\_thrust), \quad [27]$$

$$motor_R = motor_R(t\text{-}1) - pan\_disp * (1.0\text{-}motor_R/max\_thrust). \quad [28]$$

## 2.8 Arbitration Of Target Orientation And Obstacle Avoidance

Without a mechanism to prioritize the reflexes, the robot could be forced into an obstacle by the pursuit reflex, or loose track of its target by deflection from an obstacle. Since collision with obstacles must be avoided in most cases, the turning reflex to reduce the pan-axis disparity should be inhibited as long as there is an obstacle in that direction. Yet, in order to maintain a fix on the target, the camera pursuit reflex should be allowed to increase the axis disparity. As long as the window and saccade mechanisms can keep the target in the center of the receptor surface the platform will move forward on its own body axis. The design of the system insures that the peripheral vision available to the robot, when its camera has panned to an extreme (as in the case of a target moving behind an obstacle), allows the detection of new obstacles yet in the forward direction of the platform. Thus, the platform always moves in a direction that it can see. When the original obstacle has been passed, the orienting reflex will be released and the extreme disparity of camera and body axes will cause the platform to turn sharply in the original direction of the target.

# 3. Hardware

We use a Transitions Research Corporation (TRC) Labmate Mobile Robot Base. A single CCD video camera with a 90 degree field of view, mounted on a pan and tilt mechanism built in-house, provides monocular input to the vision processing hardware. Camera position is taken from shaft encoders located on the pan and tilt axles. Wheel motion information is obtained from encoders located on both left and right drive motor axles. Vision processing hardware includes an Imaging

Technologies OFG Frame Grabber coupled to a Hyperspeed Technology coprocessor board with two i860 microprocessors. The vision processing hardware cards are hosted on an 80486 PC computer located in the robot housing. The PC provides I/O to the Labmate and pan and tilt controllers. The Hyperspeed board receives video data directly from the OFG board at frame rate over an ITI vision bus. One i860 processor is dedicated to subsampling the input frame and making decisions about the required motor responses, while the other i860 processor integrates the visual input into receptive fields and performs motion analysis. Pan, tilt and drive motor commands are sent to the 80486 for integration and execution.

## 4 Results

### 4.1 Frame Rate

Actual processing rate with the algorithms described herein is approximately 8 frames per second.

### 4.2 Resolution

The pixel matrix provided to the robot vision system was 128 by 128 distributed evenly over a 68 degree square visual field. This resulted from sampling every third pixel in a 384 pixel square portion from the original 512 by 480 input frame. The 128 by 128 window was selected from within the available data based on smooth pursuit commands. Due to the log-polar mapping the resolution at the center was roughly 2 sampled pixels per degree visual angle, while at the periphery the resolution decreased to 0.14 sampled pixels per degree. However, all of the available pixels from the 128 by 128 sample that fell in a receptive field were included in the field average.

### 4.3 Motion Sensitivity

Moving objects can be detected anywhere in the visual field if they cross any of the 128 by 128 sampled pixels. Slow moving targets or targets that changed velocity frequently are more likely to evoke responses from the central fields. Conversely, rapidly moving objects are more likely to evoke responses from peripheral fields. The optimal target velocity is a function of field size and frame rate. At a frame rate of 8 frames per second, the optimal velocity of a target translating across the horizontal near the center of the visual space is 4 degrees of visual angle per second. At a distance of ten feet from the camera, this is a speed of about 0.7 foot per second. The optimal velocity for a peripheral location under the same conditions is about seven times greater, or about 5 feet per second.

### 4.4 Behavioral Capabilities

Testing was performed in a large partitioned room with an open work area of 32 by 18 feet. Three walls of this work area contained windows, doors and office furniture. An example of target acquisition and pursuit is shown in the photographs of Figure 2. From a resting position the robot turned and moved forward in pursuit of a human walking into its visual space. Obstacle avoidance was disabled during this demonstration run to allow the robot to approach the cluttered desk. With obstacle avoidance in place the robot tended to approach the target only slowly, until the position of the target allowed the robot a clear run down the center of the floor.

## 5 Discussion

We have been able to demonstrate target acquisition, tracking and trailing with some obstacle avoidance using biologically based algorithms. Several difficulties with functional integration remain. For example, if a target is able to escape the smooth pursuit mechanism and moves out of the central region of the robot's visual field, the obstacle avoidance response will interpret the target as an obstacle and cause the robot to turn away. While the target acquisition mechanism may reacquire the target, the robot can become disoriented. The optokinetic reflex also tends to drive the robot into obstacles. While the arbitration procedure is designed to avoid this, forward motion is restricted when the pan disparity is great (to avoid driving into a blind region) which can eliminate the image flow that clues the robot to the presence of the obstacle. The present system can acquire new targets while on the move, but the target motion required for this is often unrealistic. That is, the signal to noise ratio for segmenting unique target motion from induced motion in the background

Figure 2. The autonomous visually guided robot trailing a walking human in a cluttered environment.

is still unreasonably high. The paradox is that successful pursuit of the moving target minimizes relative motion of the target on the receptor surface while the pursuit motions of the robot increase induced motion of the background. The acquisition of additional targets is presently inhibited by an increased threshold during pursuit and by competition between central and peripheral regions of the retina. One of the critical problems here for which we have only a partial solution (equations [15] and [16]) is that of separating unique target motion from the motion of the background during robot transits or camera pans. Biological systems probably have a more flexible mechanism of attention control, permitting frequent and repeated sampling of potential targets, with some additional criteria for target discrimination.

## Acknowledgments

Supported by the Advanced Research Projects Agency and the Office of Naval Research under contract number N0001493WX2D002. Additional support for this project was provided by the Naval Command, Control and Ocean Surveillance Center, Independent Exploratory Development Program. The cooperation and assistance of CDR Bart Everett, Steve Timmer and Theresa Tran for the loan of the robot base, pan and tilt mechanism, and in hardware engineering is greatly appreciated.

## References

Blackburn, M.R. [1993a]. "A simple computational model of center-surround receptive fields in the retina." NRaD TD 2454, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, San Diego, California.

Blackburn, M.R. [1993b]. "Machine visual targeting modeled on biological reflexes." NRaD TD 2455, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, San Diego, California.

Blackburn, M.R., H.G. Nguyen and T.T. Tran [1993]. "Autonomous Mobile Robot Vision: Target Tracking, Trailing, and Obstacle Avoidance," in *IR-IED '93 Annual Report*, NRaD TD 2604, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, San Diego, CA, pp.63-85.

Blackburn, M.R., H.G. Nguyen, and P.K. Kaomea. [1987]. "Machine visual motion detection modeled on vertebrate retina." *SPIE Proceedings*, vol. 980, pp. 90-98.