

An outdoor 3-D visual tracking system for the study of spatial navigation and memory in rhesus monkeys

ZIA KHAN

Georgia Institute of Technology, Atlanta, Georgia

REBECCA A. HERMAN and KIM WALLEN

Emory University, Atlanta, Georgia

and

TUCKER BALCH

Georgia Institute of Technology, Atlanta, Georgia

Previous studies of the navigational abilities of nonhuman primates have largely been limited to what could be described by a human observer with a pen and paper. Consequently, we have developed a system that uses a pair of cameras to automatically obtain the three-dimensional trajectory of rhesus monkeys performing an outdoor spatial navigation and memory task. The system provides trajectories, path length, speed, and other variables that would be impossible for an unaided observer to note. From trajectory data, we computed and validated a path-length measurement. We use this measurement to compare the navigation abilities of several animals. In addition, we provide quantitative data on the accuracy of a method for automatic behavior detection. Currently, the system is being used to examine the sex differences in spatial navigation of rhesus monkeys. We expect that measures derived from the trajectory data will reveal strategies used by animals to solve spatial problems.

We describe a system that uses a pair of calibrated digital cameras to obtain the three-dimensional (3-D) trajectory of rhesus monkeys (*Macaca mulatta*) performing a spatial navigation and memory task in an outdoor setting at Yerkes National Primate Research Center Field Station in Lawrenceville, GA (see Figure 1). The monkeys search a $4.74 \times 4.74 \times 2.11$ m area in once-daily trials. Twelve goalboxes are attached in stable locations to three walls of the area, and a subset of these goalboxes are baited with food items on each trial. By manipulating the consistency of the baited locations and the availability of landmark cues in the area, we can note the ability of the animals to remember locations and their tendency to rely upon spe-

cific strategies to solve spatial problems. The addition of 3-D tracking in this spatial navigation and memory task will provide trajectories, path length, speed, and other variables impossible for an unaided human observer to note.

The 3-D tracking system addresses the challenges of outdoor tracking by leveraging an algorithm that adapts to lighting changes, whereas automated tracking in water mazes has been limited to 2-D tracking in a laboratory setting (Hval, Barrett, Wilcox, & Bailey, 2004; Mukhina, Bachurin, Lermontova, & Zefirov, 2001; Noldus, Spink, & Tegelenbosch, 2001). The outdoor setting presents substantial challenges due to natural variations in lighting, reflections, and shadows.

Previous studies of navigation in nonhuman primates have largely been limited to what could be described by a human observer with a pen and paper. The tendency for animals to follow an efficient pathway when visiting numerous target locations (the "traveling salesman" problem) has been found in a number of nonhuman primate species, including chimpanzees (Menzel, 1973), vervet monkeys (Cramer & Gallistel, 1997), yellow-nosed monkeys (MacDonald & Wilkie, 1990), and marmosets (MacDonald, Pang, & Gibeault, 1994). Recent progress in computerized testing of nonhuman primates has allowed for measurement of navigational abilities in virtual mazes (Sandstrom, Kaufman, & Huettel, 1998; Washburn

This research was supported by the Center for Behavioral Neuroscience, a Science and Technology Center Program of the National Science Foundation under Agreement IBN-9876754; by NIH Grants R01-MH50268 (to K.W.) and K02-MH01062 (to K.W.); and by NCRR Grant RR-00165 to the Yerkes National Primate Research Center, which is fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International. We thank Frank Dellaert and Michael Kaess for their help with camera calibration. Correspondence should be addressed to Z. Khan, Georgia Institute of Technology, College of Computing, 801 Atlantic Dr., Atlanta, GA 30332-0280 (e-mail: zkhan@cc.gatech.edu).

*Note—This article was accepted by the previous editor,
Jonathan Vaughan.*

A



B

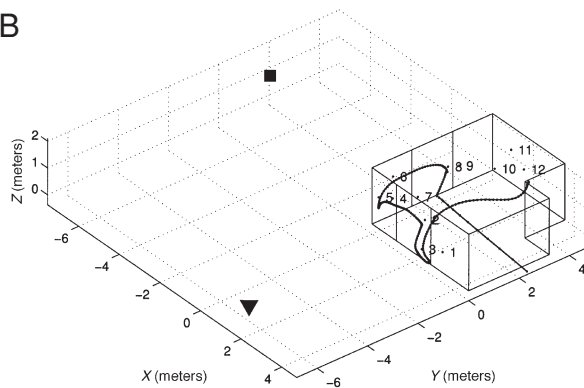


Figure 1. (A) A monkey conducting a spatial navigation task in which it visits goalboxes containing food is tracked with two digital video cameras. **(B)** On the basis of the 2-D measurements obtained from each camera, the 3-D location of the animal is estimated at a rate of 5 measurements per second. From the 3-D points and corresponding time stamp, the 3-D trajectory of the animal can be reconstructed and analyzed. The positions of Cameras 1 and 2 are indicated by the triangle and square, respectively. The numbers 1–12 indicate the locations of the goalboxes.

& Astur, 2003). However, the relationship between virtual navigation and real navigation is unknown, and the virtual system requires intensive training, which limits its utility.

Multicamera and single video camera 3-D tracking methods have been used primarily in indoor studies with controlled-lighting situations. Single-camera tracking methods that use a mirror have been used to compute the 3-D position of a single fish in an aquarium (Pereira & Oliveira, 1994) and have been automated in subsequent studies to obtain complete 3-D trajectories (Derry & Elliott, 1997). A two-camera system for 3-D visual tracking of multiple fish in an aquarium has been used in a detailed study of schooling behavior (Viscido, Parrish, & Grünbaum, 2004). Unlike the 3-D tracking system described

here, that system uses an observer in a postprocessing step, instead of during tracking, to correct trajectories.

Other technologies exist for 3-D sensing (reviewed in Parrish & Hamner, 1997). Their utility varies on the basis of the nature of the experiment, and few such technologies have been applied to obtaining 3-D trajectories. Sonar has been used in marine studies in the field to study fish aggregations. Radio receiver and transmitting collars have been successfully applied in field studies where position measurements are made over large distances (Richard-Hansen, Vié, & de Thoisy, 2000). Radar technology has been used primarily to collect 2-D trajectories of animals with bearing and range measurements (Capaldi et al., 2000). Many of these technologies require attaching some kind of device to the animals, which is either invasive, uncomfortable, and/or easily destroyed by the animal, its peers, or the environment.

Outdoor multicamera visual tracking remains an active area of research in computer science (see Collins, Lipton, Fujiiyoshi, & Kanade, 2001, for a review of recent advances). Perfect error-free visual tracking systems do not exist. In this article, we show how some of the advances in this field can be utilized to collect data in a behavioral study conducted in a challenging outdoor environment. We allow an observer to interact with the tracking system to correct tracking failures when they occur, thus allowing the collection of complete, accurate trajectories.

3-D VISUAL TRACKING

Reconstruction of the 3-D trajectory of an animal from a pair of cameras would present a substantial challenge for a human observer. At each time step, the observer would have to track the position of the animal from two vantage points and attempt to triangulate the actual 3-D location of the animal in the facility. Markers at known positions may assist in the processes of trajectory construction, but they can be prone to error. The visual tracking system described here automates both the process of tracking and triangulation, thus enabling a researcher to automatically collect complete 3-D trajectories.

Video Capture and Calibration

Digital video is captured on MiniDV tapes with two Sony VX-2000 cameras in progressive scan mode positioned so that accurate 2-D measurements can be obtained for a pair of axes (see Figure 1). The video data are transferred from tape to a hard disk with Linux-based DVCAM capture software. Every 6th frame from a 30-frames/sec sequence is recorded to disk, providing 5 frames every second of the trial. Since camera clocks are difficult to synchronize at frame resolution, the digital videos are synchronized manually on the basis of the frame number in the video when the animal enters the facility.

Next, the cameras are calibrated. Because we use a simple pinhole model for the cameras, the quality of the calibration is related to the quality of the camera. Camera imperfections such as radial distortion must be minimized to obtain good estimates of camera calibration parameters,

which justified our selection of high-quality Sony VX-2000 cameras. An observer uses a calibration program that loads a sample image from each camera, then clicks in 2-D the location $\hat{x}_{c,i}$ of $i = 1 \dots N$ 3-D corners of the facility X_i where they appear in each image from camera $c = \{1, 2\}$. The 2-D and 3-D points are represented by homogeneous coordinates where an additional coordinate is appended to the point. Hence, 2-D points are 3-D vectors, and 3-D points are 4-D vectors. The principal point (p_x, p_y) , or origin of the camera-imaging plane, was set to the center of the image (e.g., $p_x = 360$, and $p_y = 240$ for a 720×480 pixel image). The orientation of the camera is captured by a pan angle p_c , a tilt angle t_c , a roll angle r_c , and 3-D position $X_c = (x_c, y_c, z_c)^T$. Internally, the camera is parameterized by the focal lengths $f_{c,x}$ and $f_{c,y}$, which account for nonsquare pixels. Hence, the projection matrix of the c th camera can be obtained by a matrix multiplication

$$P_c = \begin{bmatrix} f_{c,x} & 0 & p_x \\ 0 & f_{c,y} & p_y \\ 0 & 0 & 1 \end{bmatrix}, [R(p_c, t_c, r_c) | X_c]$$

which results in a 3×4 projection matrix where $[\cdot | \cdot]$ denotes that the matrix and vector are appended. The $R(p, t, r)$ function returns a pan-tilt-roll rotation matrix with the convention that x coordinate of the camera points right, y points down, and z points out

$$R(p, t, r) = \begin{bmatrix} c_p c_r - s_p s_r s_t & -c_p s_r - c_r s_p s_t & -c_t s_r \\ c_r s_p + c_p s_r s_t & -s_p s_r + c_p c_r s_t & -c_r c_t \\ -c_t s_r & -c_r c_t & s_t \end{bmatrix},$$

where $c_p = \cos p$, $s_p = \sin p$, $c_t = \cos t$, $s_t = \sin t$, $c_r = \cos r$, and $s_r = \sin r$.

We used the Levenberg–Marquardt algorithm, a nonlinear numerical minimization algorithm, to correct the camera position and parameter estimates to minimize the errors between the clicked points and 2-D projections of the 3-D points onto the cameras (Hartley & Zisserman, 2000). Mathematically, the error function we minimize can be written as

$$\text{Error} = \sum_{c=\{1,2\}} \sum_{i=1}^N \|P_c X_i - \hat{x}_{c,i}\|^2.$$

The term $P_c X_i$ projects the i th known 3-D point onto the 2-D image plane of the c th camera and then computes the distance of that projected point with the point $\hat{x}_{c,i}$ that the user clicked. The algorithm was initialized with a rough estimate of the 3-D position of each camera and focal lengths. The calibration projection matrices P_1 and P_2 for Camera 1 and Camera 2 are subsequently used to triangulate the position of the animal.

Tracking and Triangulation

The tracker finds regions of movement by detecting pixels that are significantly different from the stationary background. First, the 720×480 pixel images obtained from the digital video are incrementally rescaled to a

90×60 pixel image. At each step, the width and height of the image is reduced by half by removing pixels, and the resulting image is blurred with a small Gaussian kernel. The small images are subsequently converted to the LUV color space where perceptually different colors are specified by dissimilar 3-D values. The L denotes the luminance coordinate, and U and V are coordinates that define the chromaticity of a pixel. The combination of down-sampling and blurring removes high-frequency camera noise, minor variations in lighting, or insignificant movement such as rain. Although some accuracy is lost in determining the exact center of the tracked subject, the procedure greatly reduces noise in the background subtraction algorithm. Specifically, we used the adaptive background subtraction technique of Stauffer and Grimson (1999) to locate regions of movement near the animal's expected location in these down-sampled images. The background subtraction technique models the color of a pixel as a mixture of Gaussian distributions that indicate which colors are likely when a pixel is part of the stationary background and account for noise in the camera images. In addition, the procedure gradually adjusts the distributions to account for slow changes in color that might be caused by passing clouds or changing weather conditions. We also use the cylinder model extension of Magee (2004) in which we consider pixels as background when pixel variation occurs mostly in brightness and not color. We found that the LUV color space was best suited for this background subtraction algorithm. The final result of the background subtraction procedure is a binary image that indicates regions of significant movement (Figures 2A and 2B).

To locate the animal in 3-D, the tracker finds the 2-D centers of regions of movement in both of the cameras and uses the measurements to triangulate an estimated 3-D position. The 2-D positions of the centers of foreground regions are obtained with the connected-components procedure described in Bruce, Balch, and Veloso (2000; see Figures 2C and 2D). Next, the centers of the largest movement region from each frame that are within a reasonable distance of the previously detected region of movement are used to obtain the estimated 3-D position of the animal with the inhomogeneous linear method described in Hartley and Zisserman (2000). For the convenience of the reader, we provide the detailed steps of the algorithm below: (1) Obtain the 2-D center (x, y) of the foreground region in the image obtained by Camera 1 and the 2-D center (x', y') from Camera 2. (2) Compute the 4×4 matrix \mathbf{A} by stacking the rows of the projection matrices of each camera as follows:

$$\mathbf{A} = \begin{bmatrix} x\mathbf{p}^3 - \mathbf{p}^1 \\ y\mathbf{p}^3 - \mathbf{p}^2 \\ x'\mathbf{p}'^3 - \mathbf{p}'^1 \\ y'\mathbf{p}'^3 - \mathbf{p}'^2 \end{bmatrix}.$$

\mathbf{p}^i denotes the i th row of the projection matrix of Camera 1, and \mathbf{p}'^i is the i th row of the projection matrix of

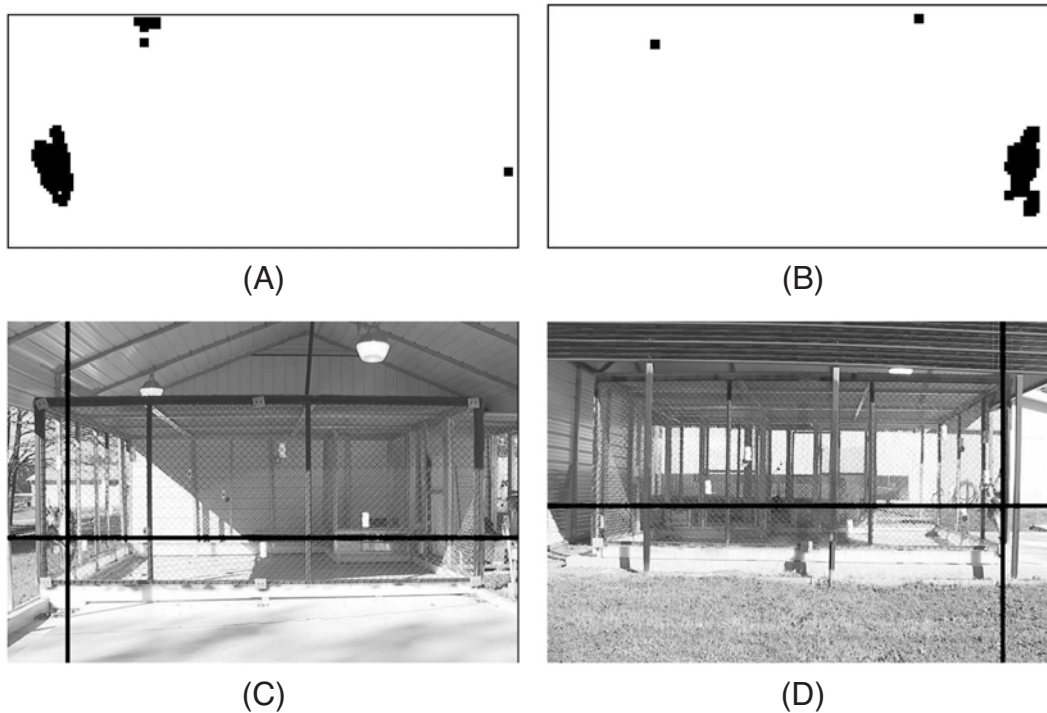


Figure 2. Background subtraction is used to locate regions of movement in Cameras 1 and 2 (A and B, respectively). Next, a connected-components procedure finds the center of the largest region within a reasonable distance from the previous 2-D position of the animal in each camera (C and D). Each 2-D measurement is used to triangulate the 3-D position of the animal.

Camera 2. (3) Extract a 4×3 submatrix from the four rows and first three columns of \mathbf{A} and store in \mathbf{C} . (4) Store the fourth column of \mathbf{A} in b . (5) Compute the pseudoinverse of \mathbf{C} , a 3×4 matrix, and store in \mathbf{C}^+ (using, for example, the *pinv* function in MATLAB). (6) Compute the triangulated point $X = \mathbf{C}^+b$. In summary, the method finds the least squares solution to a system of linear equations and estimates the 3-D point nearest to two rays emanating from both of the cameras.

Pereira and Oliveira (1994) detailed a triangulation algorithm that relies on a mirror, which enabled them to mathematically include a second *virtual camera* in their calculations. Consequently, they could triangulate an animal by measuring angles in the field of view and the reflection with a single camera. Since a large mirror is impractical outdoors, the virtual camera used by Pereira and Oliveira can be replaced by a second camera in an outdoor application. Even though the same calculations can be used to triangulate an animal, one would be forced to assume that the positions are precisely determined for both cameras. Moreover, the cameras themselves would need to be angled such that they are level with the ground, positioned at the same elevation with respect to each other, and parallel to the facility, since the method does not account for camera pan, tilt, or roll. This presents a problem because precise measurement and positioning of cameras is difficult outdoors and would substantially reduce the amount of behavioral data that could be col-

lected with such a system. In addition, this method does not account for nonsquare pixels of a CCD camera. The calibration and triangulation method described here accounts for these variables, allowing efficient outdoor data collection.

The tracking program runs interactively, thereby enabling an observer to correct failures as they occur, to provide uninterrupted trajectories. Failures are primarily caused by the following factors: changes in overall lighting due to passing clouds; movements of the shadows of trees near the facility; long periods of time during which the animal remains perfectly stationary (in these cases, the animal is considered background, and the tracker becomes prone to considering noise as movement, leading to tracking failure); passing vehicles or people in the far background of the facility; and movements of the observer. Consequently, an observer assists the tracking procedure by occasionally providing the correct position of the animal in the two camera images when a tracking failure occurs, as is shown in Figure 3. During interactive tracking, videos collected at 5 Hz can be played back at 30 Hz, so a 5-min trial can be observed in about 50 sec. The observer can rapidly track an animal, intervening as necessary, to obtain a complete trajectory. A single 5-min trial can be processed in 1–2 min. In a postprocessing step, the 3-D trajectories are smoothed with the robust local regression smoothing procedures provided in the MATLAB *curve-toolbox*.

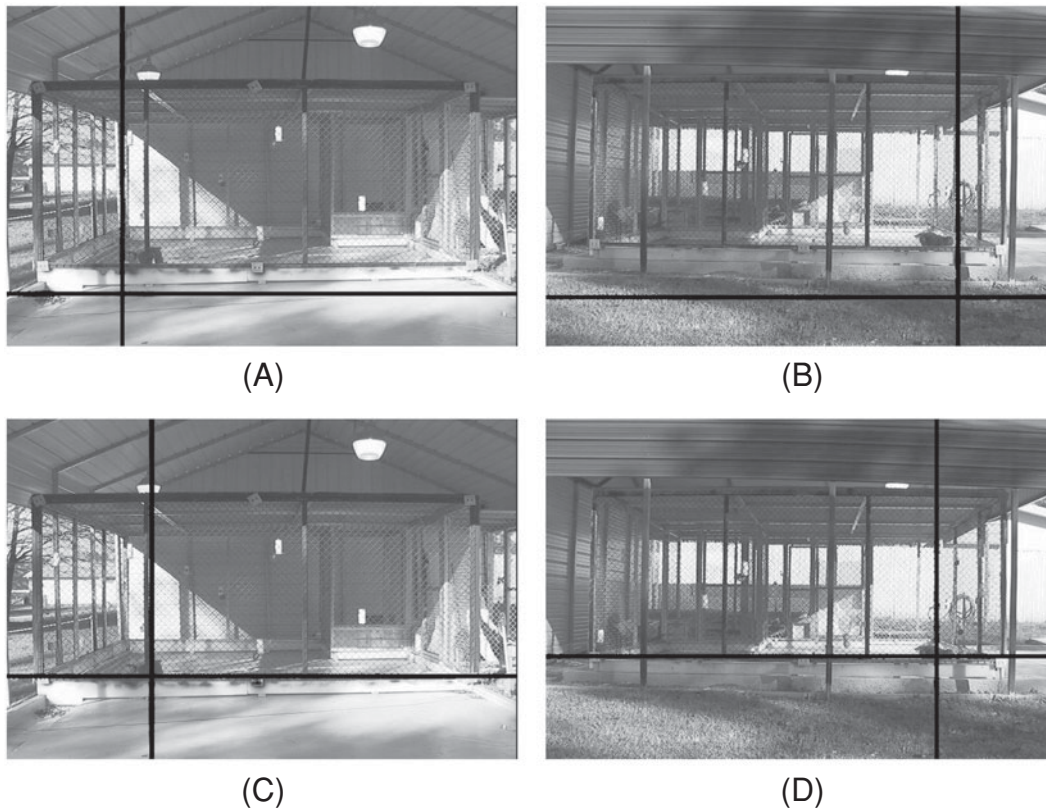


Figure 3. The tracking procedure is interactive. An observer can intervene when the tracker fails because of adverse lighting conditions such as moving shadows. In panels A and B, the tracker has wrongly detected a moving shadow of trees waving in the wind behind the cameras as the current location of the animal. The observer correctly locates the position of the animal, as is shown in panels C and D, and resumes tracking.

The sequence of operations performed by the tracking system is summarized as follows. (1) Capture: Record digital video data to a computer hard disk. (2) Calibration: Enter measurements and obtain estimates of camera calibration parameters. (3) Synchronization: Enter the frame number for each camera in which the animal enters the facility. (4) Load frame: Get a frame of the video, rescale the image, and convert to the LUV color space. (5) Background subtraction: Find regions of movement in both of the cameras and adapt the background model. (6) Triangulation: Estimate the 3-D position of the animal, correcting the position as necessary. (7) Go to step 4 until the end of the video. (8) Smoothing: Smooth 3-D trajectory and record to a data file. When tracking is complete, the final trajectory data consist of a sequence of time stamps and 3-D measurements.

TRAJECTORY ANALYSIS

A 3-D plot of the monkey's trajectory can be used to evaluate qualitative differences between paths selected by the animal in the spatial navigation and memory task. The plot provides only limited information. However, more substantial benefits arise from quantitative measures that are derived from the trajectory. Global features such as

path length allow assessment of the overall navigation abilities of the animal. A time series of trajectory features derived from a small window in time allows a researcher to evaluate the dynamics of the animal's movement. Moreover, the same features enable the automatic detection of behavior, such as a goalbox visit.

Path Length

Path length is a global feature of the monkey's trajectory that measures its navigation abilities. Animals that remember which goalboxes hold food are expected to take a shorter, more efficient path to find and visit correct goalboxes. From the trajectory data, path length was computed by summing the distance traveled between each pair of successive 3-D measurements. In Figure 4, we compare three trajectories of increasing efficiency and decreasing length. The animal in panel A made four incorrect visits to goalboxes that did not contain food and utilized a path that explored most of the facility. In panel B, the animal made three incorrect visits, but utilized a shorter, more efficient path between goalboxes. In panel C, the animal made one incorrect visit and utilized the shortest, most efficient path of the animals compared.

To determine the accuracy of the path-length estimation, we used the system to track a human walking a tra-

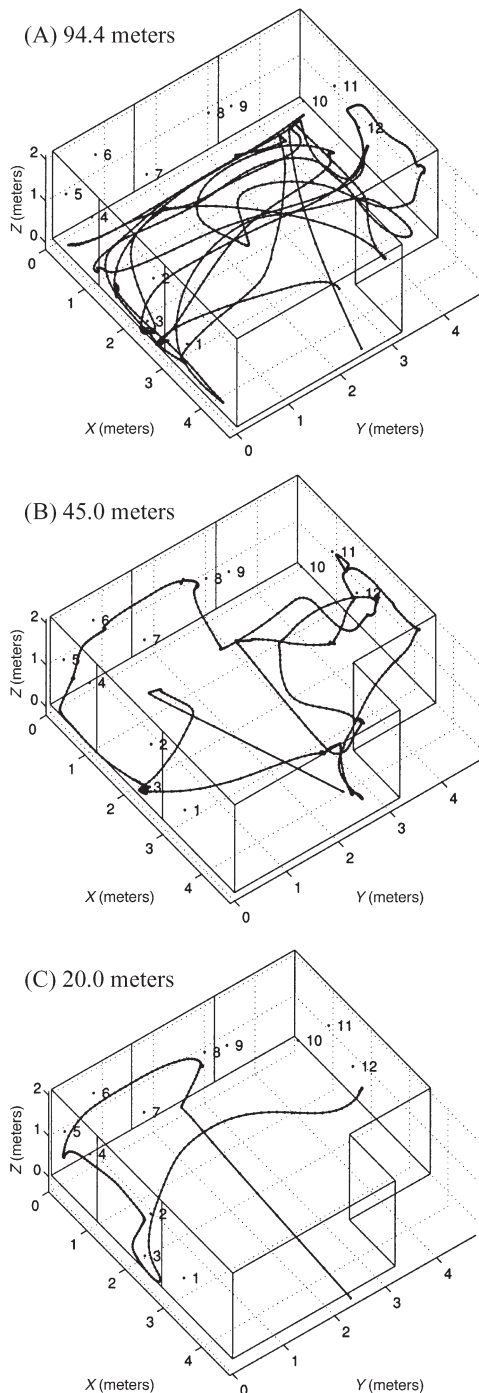


Figure 4. Trajectory length is a simple measure of how efficiently the animal searched for and found goalboxes containing food. Panels A, B, and C show three trajectories of increasing efficiency and decreasing length. In panel C, the animal made only one incorrect goalbox visit.

jectory of known length. Eight points at known locations were marked on the floor of the facility. Next, a trajectory was created by randomly picking neighboring points until a trajectory was compiled where the human walked

along 8 segments of known length starting from a fixed initial location. Figure 5 shows one such trajectory and the path computed by the tracking system. In total, 5 trajectories were compiled, providing 40 short trajectory segments of known length. The average absolute difference between the real and estimated segment lengths was $2.03 \pm 7.14\%$.

Deviations from the designated trajectory primarily occurred in situations where the human was asked to turn. The dynamics of the movement shifted the tracker from the individual's center to the individual's side. Yet, in all cases, the 3-D position reported by the tracker remained within the tracked individual's body.

Features

Time series of local trajectory features derived from a small window in time provide useful information on the dynamics of the animal's movement. Such features are intrinsic or extrinsic. The intrinsic features are characteristics of the animal's path, whereas the extrinsic features are computed relative to objects in known 3-D locations. In Figure 6B, we compute the instantaneous speed of the animal in meters per second between successive time steps and plot the time of goalbox visits. The speed increases as the animal maneuvers toward the goalbox and drops to zero as it pauses, visits the goalbox, and consumes a food item. In addition, we observed a smaller peak in speed prior to the incorrect goalbox visit, which suggests that the maximum instantaneous speed prior to a visit could be a measure of the animal's confidence in its memory (Hampton, 2001). In Figure 6C, we also plot the distance to the nearest goalbox over time, an extrinsic feature of the trajectory. As was expected, the distance to the goalbox dropped just before a visit. Before several of the visits, a sharp peak was observed, which occurred because the nearest goalbox changed as the animal moved toward the destination goalbox.

Goalbox Visits

Local trajectory features can enable a researcher to define criteria for identifying behaviors that are easy to replicate. In the spatial navigation task, a human observer records visits when an animal reaches into a goalbox or clearly looks into the opening of a goalbox. However, the observer may encounter ambiguous situations where it is unclear whether the animal is examining the goalbox or another object in its field of view. An observer must attempt to judge when the animal is visiting the box, a process that may introduce bias into the collected data. Consequently, we examined how effectively both distance to goalbox and instantaneous speed can be used to automatically identify goalbox visits from the trajectory.

As a first step, we utilized a single feature—distance to goalbox—to identify visits. We assigned a 3-D measurement to a goalbox number when the animal was found to be within a specified distance, measured in meters, to the goalbox (see Figure 7). Initially, we used the time the animal entered the goalbox region as the time of the visit.

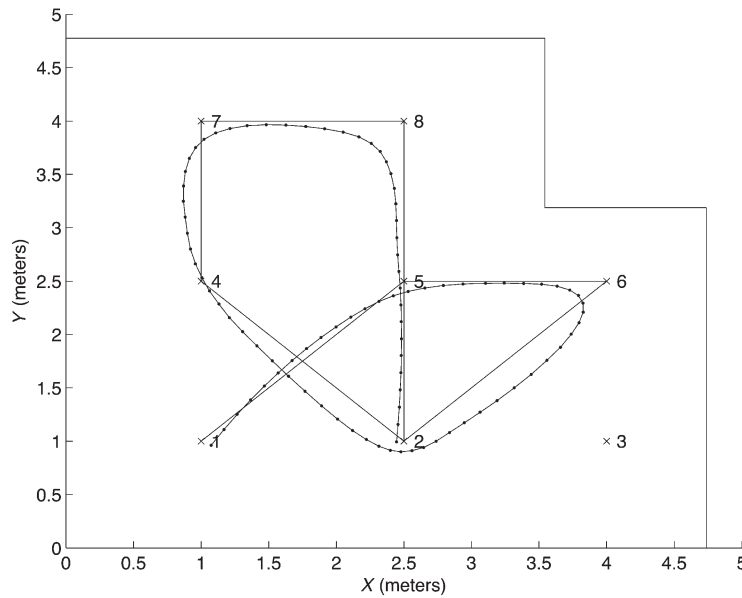


Figure 5. We measured the accuracy of the trajectory length calculation by comparing the length of a randomly selected trajectory taken by a human walking between known points. The trajectory was recorded by the cameras and then evaluated by the visual tracking system. The solid line shows the randomly selected trajectory that sequentially visits Points 1, 5, 6, 2, 4, 7, 8, 5, and 2, each designated by \times . The dotted solid line shows the trajectory recorded by the visual tracking system.

However, this approach recorded visits well before the human-recorded visits. By examining the trajectories, we noticed that the animal had a tendency to maneuver within the detection region before and after visiting a goalbox. Consequently, we computed visits by selecting the midpoint between start and end times of successive measurements that were labeled with the same goalbox number. Unfortunately, this strategy fails when the animal remains stationary near a goalbox for an extended period of time. We found that this was a relatively rare occurrence, however, and the midpoint provided a better estimate of visit time. After visit times were determined, visits to the same goalbox within 15 sec of each other were combined into a single visit by averaging all of the visit times. This compensated for situations where the animal may have maneuvered near a goalbox, briefly exiting the detection region and reentering prior to the actual visitation.

We compared the automatically identified visits with those recorded by a trained observer. In our analysis, we assume that the visits recorded by the observer provide *ground-truth*. In reality, trained observers can make errors, which we assume to be extremely rare. Since camera clocks and the observer’s clock were not perfectly synchronized, we considered a visit to be correctly identified by the computer when the goalbox number was correctly assigned and the visit was within 15 sec of the time recorded by the observer. In Table 1, we show visit times determined for the trajectory in Figure 7. All visits except that to Goalbox 10 were within 15 sec of the observer-recorded visits. After the visit to Goalbox 10, the animal

remained stationary near the goalbox for several seconds. The error recorded both a missed detection and a false positive.

We used this approach to analyze thirty-one 5-min trials where the observer found 205 goalbox visits. By adjusting the size of the detection region around the goalbox from 0.0 to 0.7 m, we examined the trade-off between detection rate, percent of correctly identified visits, and false positive rate, detections not recorded by the observer per second of video. We plotted the data in a receiver operating characteristic (ROC) curve in Figure 8 and found that the best trade-off between false positives and true positives was achieved when the detection region was set to 0.325 m. The system correctly found 131 of the observer-identified visits. However, it reported 65 false positives, visits that were not recorded by the observer.

By introducing a second trajectory feature—instantaneous speed—we were able to reduce the number of false positives. Instead of assigning a 3-D measurement to a goalbox number when it was found to be within a specified distance from a box, we considered the speed of the animal. A goalbox number was assigned a 3-D measurement only when the instantaneous speed was found to be under a specified threshold. In Figure 8, we plot the ROC curves for a threshold of 0.3 m/sec and 0.05 m/sec. At detection rates below 73% and above 10%, using both features to identify goalbox visits generated fewer false positives than using the distance to goalbox alone. To correctly identify 131 visits, which is the best detection rate for the single feature approach, the false positives were reduced

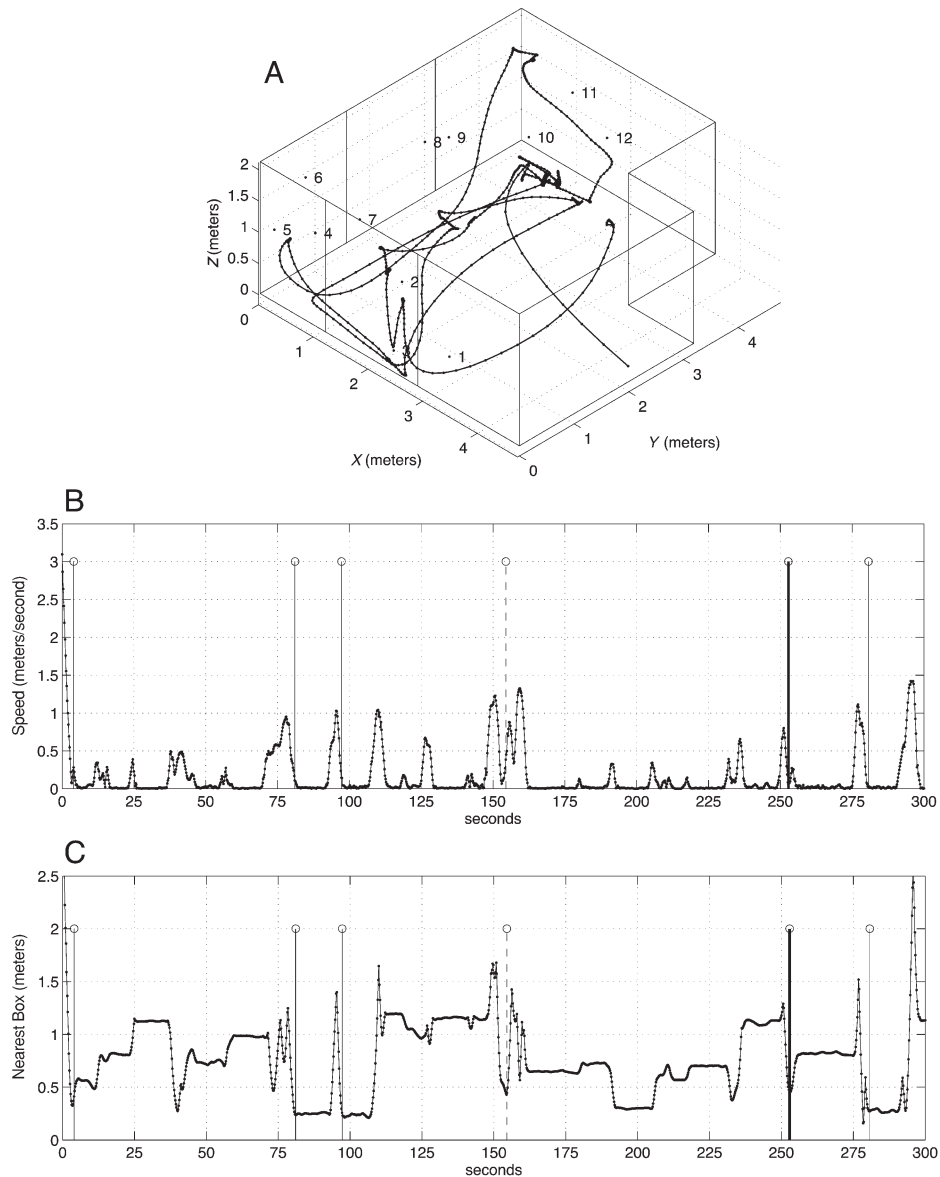


Figure 6. In panel B, we show the speed, an intrinsic feature of the trajectory in panel A, in meters per second. Panel C shows an extrinsic feature, the distance in meters to the nearest goalbox. The solid ticks show a correct goalbox visit. The bold ticks show an incorrect visit to a goalbox without a food item, and the dashed ticks show a return visit to a previously visited goalbox.

to 59 visits when the detection region was set to 0.345 m and the speed threshold to 0.3 m/sec. We could further reduce the false positives to 46 visits by setting the detection region to 0.4 m and the speed threshold to 0.05 m/sec. These results suggest that approaches that utilize several trajectory features will detect behavior more accurately than approaches that use only a single trajectory feature.

Goalbox visit detection remains an area for future development. The specific criteria we selected do not account for dynamics of the animal's movement before and after a goalbox visit, which could be used to better

identify visits. In addition, features of the environment such as the configuration of the walls of the facility near a goalbox could contribute to more accurate detections. Consequently, we are currently investigating pattern recognition and machine learning techniques that combine multiple features of the trajectory and environment to robustly identify behavior (Duda, Hart, & Stork, 2000).

DISCUSSION

Multicamera visual tracking allows collection of movement data in a spatial navigation and memory task in

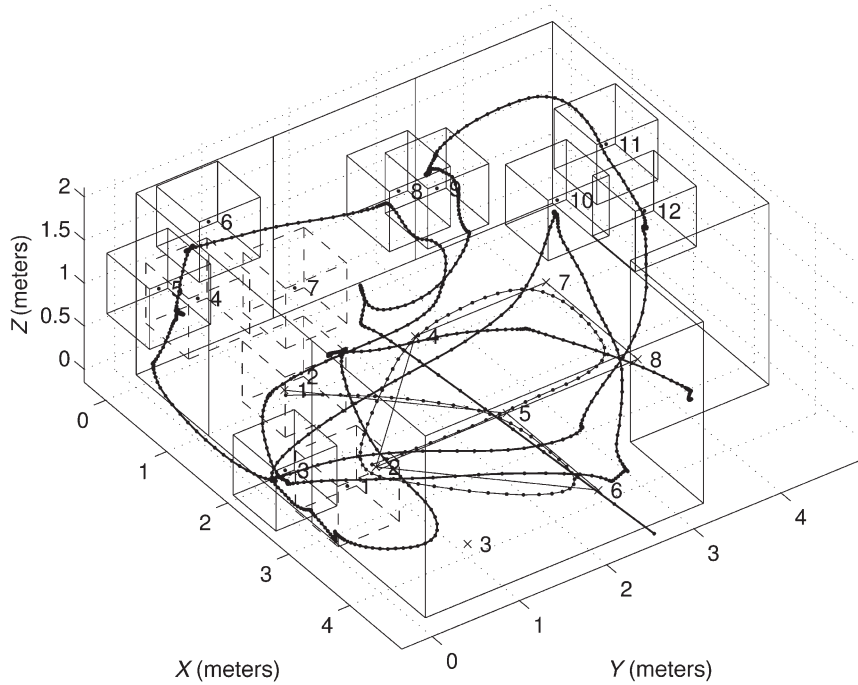


Figure 7. Goalbox visits can be automatically identified by measuring the spatial relationship of the trajectory to the known 3-D positions of the goalboxes. By finding the time the animal enters and leaves a region around a goalbox, visits can be automatically identified. The regions indicated by solid lines highlight visits to Goalboxes 3, 5, 6, 8, 9, 10, 11, and 12. The dashed-line boxes indicate goalboxes that were not visited by the animal.

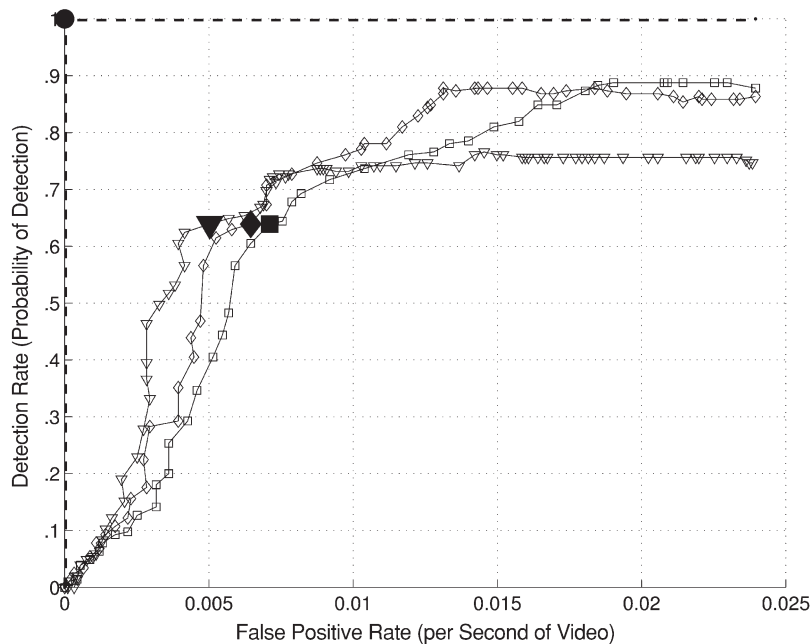


Figure 8. By varying the size of the goalbox visit detection region, we obtain a receiver operating characteristic (ROC) curve that shows the trade-off between detection rate and false positive rate. Better goalbox visit detection approaches generate ROC curves that approach the optimal curve, shown as a thin dashed line. Perfect detection with zero false positives is indicated by the circle on the dashed line. Using open squares, we plot the ROC curve for detection approach that uses only distance to goalbox to identify visits. Using \diamond and ∇ , we plot detection approaches that limit an additional trajectory feature—instantaneous speed—at 0.3 m/sec and 0.05 m/sec, respectively, to reduce false positives. The bold square shows the best trade-off between detection rate and false positive rate for the single-feature detection approach. Using two features, we can achieve this same detection rate with fewer false positives, as shown by the bold \diamond and ∇ .

Table 1
The Goalbox Visit Times as Recorded by a Human Observer and Estimated by the Computer (in Seconds) Since the Start of the Trial for the Trajectory Shown in Figure 7

Goalbox Number	Human	Computer	Difference
9	14.0	19.3	5.3
8	24.3	29.9	5.6
11	48.5	50.8	2.3
12	49.9	57.6	7.7
3	70.3	84.1	13.8
5	102.6	113.6	11.0
6	120.6	127.0	6.4
8	131.0	135.7	4.7
10	179.3	195.4	16.1

Note—All visits except that to Goalbox 10 were within 15 sec of the observer-recorded visits.

which an animal navigates in three dimensions. Visual tracking is an area of active research in computer science. Although a perfect outdoor visual tracking system does not exist, we can obtain practical data for a behavioral experiment in a challenging outdoor environment by using adaptive algorithms and designing the system so that observers can correct errors when they occur. The system we describe here enables an observer to collect accurate trajectory data in substantially less time than what was required to collect the video.

A 3-D plot of an animal's trajectory allows an observer to easily see qualitative differences between paths and movement, but the plot provides only limited information. Quantitative measures derived from the trajectory, which will vary on the basis of the design of the experiment, are essential for analyzing the trajectory data set. Unlike behavioral observations, a trajectory can be analyzed using different quantitative measures without requiring a researcher to conduct the experimental trial a second time. For instance, in this task we selected path length, a global feature of a monkey's trajectory, to assess its navigation abilities. But, path length is limited in that it does not measure motivation. An animal may have a shorter trajectory simply because it is not motivated to visit goalboxes. To address this limitation, we intend to examine quantitative measures of motivation, such as average change in speed, in addition to path length. The trajectory data allows us to examine this second component without significant change to the experiment.

A human observer can be easily trained to collect experimental measures that require gestalt judgments such as highly accurate recognition of a behavior. Precise quantitative determinations, such as the animal's trajectory and path length, are difficult for a human observer to collect, whereas computers are exactly the opposite. Thus, the goalbox detection results we have presented illustrate the current limitations of using a computer. Tasks that require modeling the effect of context on perception, like automated behavior recognition, rapidly lead to complex inference and learning problems. Solutions to these problems are an active area of research in computer sci-

ence (see Forsyth & Ponce, 2002, and Russell & Norvig, 2003, for reviews).

The 3-D visual tracking system described here is currently being used to examine the sex differences in spatial navigation of rhesus monkeys. In this study, we have derived quantitative measures such as path length, by using a computer to obtain trajectory data, and we had a human observer to record visits to goalboxes and additional behavior of the animals. Our approach combines the strengths of both computer and human. We expect that measures derived from the trajectory data, combined with human observations, will reveal strategies used by the animals to solve spatial problems.

REFERENCES

- BRUCE, J., BALCH, T., & VELOSO, M. (2000). Fast and inexpensive color image segmentation for interactive robots. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems* (Vol. 3, pp. 2061-2066). Piscataway, NJ: IEEE.
- CAPALDI, E. A., SMITH, A. D., OSBORNE, J. L., FAHRBACH, S. E., FARRIS, S. M., REYNOLDS, D. R., ET AL. (2000). Ontogeny of orientation flight in the honeybee revealed by harmonic radar. *Nature*, **403**, 537-540.
- COLLINS, R., LIPTON, A., FUJIYOSHI, H., & KANADE, T. (2001). Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE*, **89**, 1456-1477.
- CRAMER, A. E., & GALLISTEL, C. R. (1997). Vervet monkeys as traveling salesmen. *Nature*, **387**, 464.
- DERRY, J. F., & ELLIOTT, C. J. H. (1997). Automated 3-D tracking of video-captured movement using the example of an aquatic mollusk. *Behavior Research Methods, Instruments, & Computers*, **29**, 353-357.
- DUDA, R. O., HART, P. E., & STORK, D. G. (2000). *Pattern classification* (2nd ed.). New York: Wiley Interscience.
- FORSYTH, D. A., & PONCE, J. (2002). *Computer vision: A modern approach*. Upper Saddle River, NJ: Prentice-Hall.
- HAMPTON, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, **98**, 5359-5362.
- HARTLEY, R., & ZISSERMAN, A. (2000). *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press.
- HVAL, E. S., BARRETT, S. F., WILCOX, M., & BAILEY, T. A. (2004). An improved Morris water maze tracking algorithm for psychophysical studies. *Annals of Biomedical Engineering*, **32**, 1141-1152.
- MACDONALD, S. E., PANG, J. C., & GIBEAULT, S. (1994). Marmoset (*Callithrix jacchus jacchus*) spatial memory in a foraging task: Win-stay versus win-shift strategies. *Journal of Comparative Psychology*, **108**, 328-334.
- MACDONALD, S. E., & WILKIE, D. M. (1990). Yellow-nosed monkeys' (*Cercopithecus ascanius whitesidei*) spatial memory in a simulated foraging environment. *Journal of Comparative Psychology*, **104**, 382-387.
- MAGEE, D. (2004). Tracking multiple vehicles using foreground, background and motion models. *Image & Vision Computing*, **22**, 143-155.
- MENZEL, E. W. (1973). Chimpanzee spatial memory organization. *Science*, **182**, 943-945.
- MUKHINA, T. V., BACHURIN, S. O., LERMONTOVA, N. N., & ZEFIROV, N. S. (2001). Versatile computerized system for tracking and analysis of water maze tests. *Behavior Research Methods, Instruments, & Computers*, **33**, 371-380.
- NOLDUS, L. P. J. J., SPINK, A. J., & TEGELBOSCH, R. A. J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, **33**, 398-414.
- PARRISH, J. K., & HAMNER, W. M. (Eds.) (1997). *Animal groups in three dimensions*. Cambridge: Cambridge University Press.
- PEREIRA, P., & OLIVEIRA, R. F. (1994). A simple method using a single video camera to determine the three-dimensional position of a fish.

- Behavior Research Methods, Instruments, & Computers*, **26**, 443-446.
- RICHARD-HANSEN, C., VIÉ, J. C., & DE THOISY, B. (2000). Translocation of red howler monkeys (*Alouatta seniculus*) in French Guiana. *Biological Conservation*, **93**, 247-253.
- RUSSELL, S. J., & NORVIG, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- SANDSTROM, N. J., KAUFMAN, J., & HUETTEL, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Cognitive Brain Research*, **6**, 351-360.
- STAUFFER, C., & GRIMSON, W. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 246-252). Los Alamitos, CA: IEEE.
- VISCIDO, S. V., PARRISH, J. K., & GRÜNBAUM, D. (2004). Individual behavior and emergent properties of fish schools: A comparison of observation and theory. *Marine Ecology Progress Series*, **273**, 239-249.
- WASHBURN, D. A., & ASTUR, R. S. (2003). Exploration of virtual mazes by rhesus monkeys (*Macaca mulatta*). *Animal Cognition*, **6**, 161-168.

(Manuscript received March 16, 2004;
revision accepted for publication August 14, 2004.)