Journal of Experimental Psychology:
Human Perception and Performance
2005, Vol. 31, No. 2, 235–247

Copyright 2005 by the American Psychological Association
0096-1523/05/$12.00 DOI: 10.1037/0096-1523.31.2.235

Multiple-Object Tracking Is Based on Scene, Not Retinal, Coordinates

Geniva Liu, Erin L. Austen, Kellogg S. Booth, Brian D. Fisher, Ritchie Argue, Mark I. Rempel, and James T. Enns
University of British Columbia

This study tested whether multiple-object tracking—the ability to visually index objects on the basis of their spatiotemporal history—is scene based or image based. Initial experiments showed equivalent tracking accuracy for objects in 2-D and 3-D motion. Subsequent experiments manipulated the speeds of objects independent of the speed of the scene as a whole. Results showed that tracking accuracy was influenced by object speed but not by scene speed. This held true whether the scene underwent translation, zoom, rotation, or even combinations of all 3 motions. A final series of experiments interfered with observers' ability to see a coherent scene by moving objects at different speeds from one another and by distorting the perception of 3-D space. These reductions in scene coherence led to reduced tracking accuracy, confirming that tracking is accomplished using a scene-based, or *allocentric*, frame of reference.

An important task of the visual system is to keep track of objects as they move through space. Whether the observer is an air traffic controller tracking airplanes on a radar screen or an athlete tracking team members and opposing players on a field, there is a need to maintain a visual index for objects that are changing in their spatial location over time. It has been shown that human observers can track up to four or five randomly moving objects with fairly good accuracy (Pylyshyn & Storm, 1988; Scholl, 2001; Yantis, 1992). Tracking performance is high even when the tracked objects are identical to untracked objects in all respects other than their motion paths, pointing to a tracking ability that is based solely on the spatiotemporal history of the objects.

The ability to track multiple objects has been used to investigate the role of perceptual organization in tracking (Yantis, 1992), the deployment of attention in depth (Viswanathan & Mingolla, 1998), and the nature of visual object representations (Scholl & Pylyshyn, 1999). However, there is a fundamental question concerning tracking that is not yet well understood. In what frame of reference are objects being tracked? Are the visual indexes or "pointers" that observers use to track objects pointing to locations in a *retinotopic* map (a coordinate system with respect to the retina), or are they pointing to locations in an *allocentric* map (a coordinate system with respect to the scene)? Although there is little direct research on this question, there are good reasons to suspect that either one of these options may be correct.

One reason to suspect that tracking is accomplished using a retinal frame of reference is that the entire human visual system is organized at the physiological level in a retinotopic fashion. When neurons in one visual area of the brain (e.g., V1) communicate with neurons in other areas (e.g., V5 or temporal lobe), they tend to maintain a strict spatial correspondence. Thus, when neurons in different visual areas are responding to the same object, they are automatically linked by virtue of their common reference to the same visual-field location (Lennie, 1998; Van Essen et al., 2001). A mechanism that was designed to keep a "finger" on an object as it moved over time would simply have to track the changing neural activity in one of these retinotopically organized visual areas.

Yet there are equally compelling reasons to suspect that tracking is accomplished using a reference frame tied to locations in the world rather than in the eye. One such reason comes from an examination of eye movements. Saccades that are made from one location to another are referenced to stationary environmental landmarks rather than to specific retinal coordinates. This is evident when small changes are made to the locations of saccadic targets while the eye is en route to the target; the eye automatically corrects for these changes in location even when observers are unaware that the target has moved (Deubel, Bridgeman, & Schneider, 1998). Smooth pursuit movements are also linked to environmental rather than to retinal locations, as can be seen when one tracks a moving object while simultaneously rocking one's head back and forth (Raymond, Shapiro, & Rose, 1984). The above-cited and many other psychophysical studies suggest that visual perception is geared toward registering the position of objects in the environment rather than registering objects with respect to their retinal location (Fecteau, Chua, Franks, & Enns, 2001; Li & Warren, 2000; Liu, Healey, & Enns, 2003).

The goal of the present study was to determine whether multiple-object tracking is based on retinal coordinates or scene coordinates. Our approach began with the longstanding observation that tracking accuracy varies systematically with object speed: Objects moving at a slower speed are generally tracked more

Geniva Liu, Erin L. Austen, Kellogg S. Booth, Brian D. Fisher, Ritchie Argue, Mark I. Rempel, and James T. Enns, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

This study was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grants to Kellogg S. Booth, John C. Dill, James T. Enns, Brian D. Fisher, Karon E. MacLean, and Ronald A. Rensink and by NSERC and Michael Smith Foundation for Health Research graduate fellowships to Geniva Liu and Erin L. Austen. We thank Mark Liljefors and Alexander Stevenson for their computer programming skills.

Correspondence concerning this article should be addressed to James T. Enns, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4, Canada. E-mail: jenns@psych.ubc.ca

accurately than objects moving at a faster speed (Pylyshyn & Storm, 1988; Yantis, 1992). However, retinal motion and scene motion have typically been confounded in previous studies. In the present study, we varied the speed of object motion relative to the center of the scene (allocentric speed) separately from the speed of scene motion relative to the viewing frame (retinal speed). If tracking is based on a retinal frame of reference, then accuracy should vary directly with the speed at which objects transit the eye, regardless of their relative speed of movement within the scene. However, if tracking is based on an allocentric frame of reference, then accuracy should vary most directly with the speed of objects within the scene, and retinal speed should not matter.

Overview of Experiments

In Experiment 1, tracking accuracy for objects moving within the confines of a 2-D rectangle was compared with tracking accuracy for objects moving within a depicted 3-D box. In both conditions, object speed was varied. The results showed that objects could be tracked equally well in both situations, with a small tendency for tracking to be even more accurate in the 3-D display. Most critically for the remaining experiments, tracking accuracy declined systematically with increases in object speed.

In Experiment 2, tracking accuracy for objects within the 3-D box was measured while the box as a whole underwent a "wild ride," consisting of dynamic and simultaneous translations in the picture plane, rotations in depth around the vertical axis, and dilations and contractions in depth. That is, in addition to varying the relative speed of objects within the 3-D box, we varied the motion of the whole box in a complex way. Yet the results showed clearly that tracking accuracy was unaffected by these global variations in scene motion. Only the motion of the objects relative to the scene as a whole influenced tracking accuracy.

In Experiment 3, we removed most of the pictorial support for the 3-D box to see how perception of a stable scene depended on the wire frame and grid floor that had been used to convey the layout of the scene. The results showed that tracking accuracy was unaffected by the removal of these cues to the third dimension. This suggested that the movement of the objects themselves, within the confines of the depicted 3-D box, were sufficient to provide the structure from motion necessary to perceive the layout of the 3-D scene.

In Experiments 4–6, we tested the allocentric-tracking hypothesis by attempting to reduce the perceived coherence of the 3-D structure. In Experiment 4, the objects to be tracked moved at two different speeds within the same scene, thereby sharply reducing both the coherence of the 3-D scene and tracking accuracy. In Experiment 5, we reduced scene coherence by projecting the image of the scene onto the junction of two dihedral surfaces. Even though the retinal projection for the observer was identical to the conditions in which tracking accuracy had been high (Experiments 2 and 3), tracking accuracy was reduced along with the coherence of the scene. In Experiment 6, scene and retinal coherence were teased apart further. In one condition, we reduced retinal coherence but maintained scene coherence by projecting the image obliquely onto one surface that observers viewed from an oblique angle; in a second condition, we reduced both retinal and scene coherence by projecting the image obliquely onto two dihedral surfaces while observers viewed from an oblique angle. Consistent with the allocentric-tracking hypothesis, tracking accuracy was reduced in the second condition, in which scene coherence was disrupted. Taken together, these results provide strong support for the view that multiple-object tracking is accomplished using an allocentric frame of reference. (Online demonstrations of many of these experiments can be viewed at http://www.interchange.ubc.ca/vsearch/tracking/.)

Experiment 1: Object Speed Reduces Tracking Accuracy

The purpose of Experiment 1 was to establish several important baseline measurements for the experiments that followed. First, because the displays in all of the subsequent experiments depicted objects moving in a 3-D scene, we sought to compare tracking accuracy in 2-D and 3-D displays as directly as possible. Previous studies have reported that multiple-object tracking is not impaired in accuracy when objects disappear briefly as they pass behind occluding surfaces (Scholl & Pylyshyn, 1999). Studies using the additional cue of binocular disparity have reported improved tracking accuracy relative to control displays that lacked this cue (Viswanathan & Mingolla, 2002). Tracking accuracy is also improved when the moving objects are distributed across two planes in depth rather than moving in a single plane only (Viswanathan & Mingolla, 2002). Our goal in Experiment 1 was therefore to provide as rich a 3-D environment as possible, using only pictorial and motion cues for depth, and to compare tracking under these conditions with the "standard" case of tracking on a 2-D screen.

To manipulate motion relative to an allocentric frame in the present study, we depicted the objects moving within a 3-D box defined by a wire frame and a grid floor, as shown in Figure 1. To help reinforce the perception of 3-D motion, we added the depth cue of dynamic changes in relative size. When objects were closest to the viewer, they subtended 0.8° of visual angle, and when they were farthest away, they subtended 0.5°. At intermediate depth locations, object size varied linearly between these extreme values.

Our second goal was to establish tracking accuracy when the objects in the scene were moving at various rates of speed. In Experiment 1, all of the objects moving in a display moved at the same speed, but the rate of speed on any given trial was 1°, 2°, or 6° per second. This was a sufficient variation in speed to have a large influence on tracking accuracy.

Our third goal was to measure the decline in tracking accuracy as the number of objects was increased, thereby allowing us to obtain a stable measure of tracking capacity for every condition that was tested (Pashler, 1988). In Experiment 1, a total of 16 moving objects were present in each display, but the number designated as targets varied randomly among 2, 4, 6, and 8. This turned out to be a large enough range to observe tracking accuracy that was near perfect in some cases and near chance in others.

Method

Participants. Twenty-five undergraduate students (17 female, 8 male; mean age = 19.4 years) participated in an hour-long session in exchange for course credit. All reported normal or corrected-to-normal vision. Twelve participated in the 2-D condition, and 13 participated in the 3-D condition.

Apparatus. A Dell Pentium III 533-MHz computer running custom software written in C++ using OpenGL for 3-D graphics controlled displays and data collection for all experiments. Observers were seated at

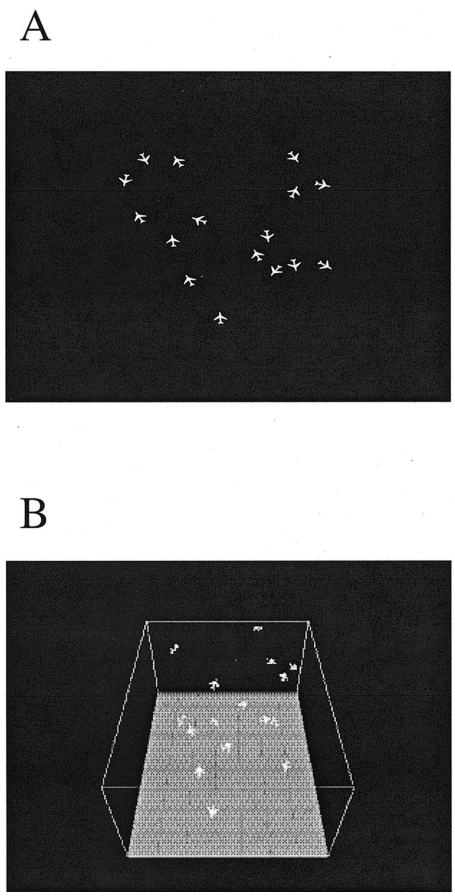

Figure 1. Experiment 1 compared tracking accuracy for objects moving in either a depicted 2-D (A) or a depicted 3-D (B) environment.

a viewing distance of 57 cm from a 19-in. (48.26-cm) Sony Trinitron monitor (resolution: 1,024 × 768 pixels) that had a 35° (wide) × 26° (high) viewable area.

Stimuli and procedure. Moving objects consisted of 16 small white airplanes seen within a viewing frame, as shown in Figure 1. In the 2-D condition, the frame was a 2-D rectangle subtending 20° (width) × 16° (height) of visual angle. In the 3-D condition, the frame was a depicted 3-D rectangular wire-frame box, drawn in white on a black background, with an aspect ratio of 285 pixels horizontally (x), 360 pixels in depth (y), and 285 pixels vertically (z). This corresponded to approximately 16° (x) × 20° (y) × 16° (z) of visual angle when each dimension was viewed from an orthogonal vantage point. The floor of the rectangular box was light gray and overlaid with the outline of a square black grid. The airplanes subtended 0.65° of visual angle in the 2-D condition and 0.50°–0.80° of visual angle in the 3-D condition. In the 3-D condition, the objects and the frame were drawn to create a camera angle of 45° to the x–y plane of the frame.

It is important to note that although observers tracked small airplane shapes in this experiment, the shape of the objects to be tracked had no influence on performance. We confirmed this with our own tests comparing airplanes with spheres of similar size, and others have found the same result (T. Horowitz, personal communication, May 5, 2003).

At the beginning of each trial, 16 stationary objects were randomly positioned onscreen. After 1 s, each object in a subset of 2, 4, 6, or 8 objects was surrounded by green marking circles. The marking circles flashed off and on four times at 200-ms intervals and then remained onscreen for another 2 s. This designated the target set that observers were to track through a period of motion. The marking circles then disappeared, and all objects began to move at a constant speed of 1°/s, 2°/s, or 6°/s for a duration of 10 s.

Objects moved in a straight line in a randomly chosen direction until they reached the edge of the frame. Upon meeting the frame edge, an object's trajectory was changed so that it appeared to bounce off the edge (2-D condition) or the wall (3-D condition) and continue on a trajectory consistent with the physics of a billiard ball moving at a constant speed. Object boundaries were allowed to intersect from the perspective of the observer for both the 2-D and 3-D conditions. However, in the 3-D condition, objects were not allowed to occupy the same region of 3-D space. At the end of the 10-s period of continuous motion, all of the objects stopped moving and a green circle surrounded 1 object. This circle flashed briefly four times and then remained onscreen until the observer responded. On half of the trials, this probe surrounded a target object (one of the objects to be tracked); on the other half, it surrounded a nontarget object.

The observer's task was to indicate whether the probed object was part of the original target set. Observers pressed the *Z* key for target objects and the *slash* key for nontarget objects. Correct responses were followed by a centrally presented *plus* symbol, incorrect responses were followed by a *minus* symbol, and no response was followed by a *0*. At the end of each block of trials, a message displayed the percentage of errors for the most recent block of trials, and observers were prompted to initiate the next block when ready. Observers were instructed to respond accurately but to guess when uncertain.

Observers performed 20 practice trials in each condition before formal testing began. The 2-D and the 3-D condition each consisted of 144 trials (3 object speeds × 4 sets of objects to be tracked × 2 probe types × 6 repetitions). The order of conditions was randomized throughout the experiment. The testing session was divided into three blocks of 48 trials, with a self-paced break between each block. Observers were instructed to maintain fixation at the center of the display throughout the trials, but eye movements were not monitored; eye movements have been shown not to affect tracking accuracy (Scholl & Pylyshyn, 1999).

Results

Performance was evaluated using two different dependent measures. First, tracking accuracy was examined with an analysis of variance (ANOVA) on the between-observers factor of display (2-D, 3-D) and the within-observer factors of object speed (1°/s, 2°/s, 6°/s) and target number (2, 4, 6). Trials involving 8 targets were excluded from the analysis because accuracy was near the chance level of 50% in all conditions (less than 60% correct). This analysis revealed the broad effects of the experimental factors on tracking accuracy.

Second, tracking accuracy was assessed using a measure of capacity (K) adapted from Pashler (1988). This measure was originally developed as a quantitative estimate of the number of items held in memory during a change-detection task. The formula is $K = [NT * (p\text{Hits} - p\text{FA})]/(1\ p\text{FA})$, where K is capacity, NT is the number of items to be tracked, pHits is the proportion of hits, and pFA is the proportion of false alarms. The upper limit of K is bounded by the number of objects that the observer is asked to track; as a result, to avoid artificial deflation of K, we excluded trials on which the target number was 2.

Accuracy. Mean proportions of correct responses are shown in Figure 2. Tracking accuracy decreased as object speed increased from 1°/s to 6°/s, $F(2, 44) = 70.50$, $p < .001$. Accuracy also decreased as the number of targets increased from 2 to 6, $F(2, 44) = 142.47$, $p < .001$. Finally, there was a Display × Target Number interaction, $F(2, 44) = 3.08$, $p < .06$, indicating that as

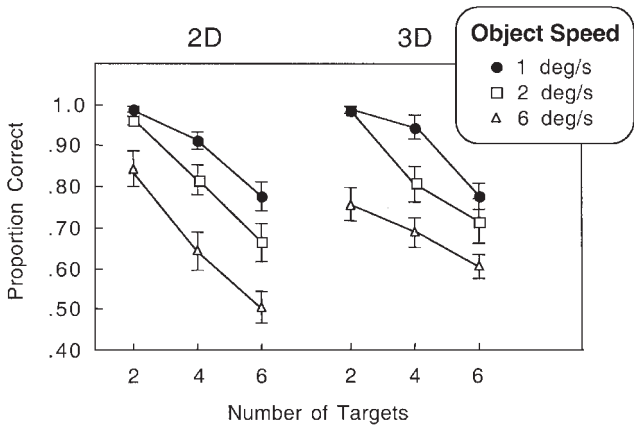

Figure 2. Mean proportions of correct responses (reflecting tracking accuracy) in Experiment 1. Error bars represent plus or minus 1 standard error of the mean. deg = degree.

number of targets increased, accuracy remained higher in the 3-D displays than in the 2-D displays.

Capacity. Capacity decreased as object speed increased from 1°/s ($K = 3.66$) to 2°/s ($K = 2.53$) to 6°/s ($K = 1.60$), $F(2, 46) = 25.52$, $p < .001$. No other effects were significant.

Discussion

Experiment 1 replicated previous findings that tracking accuracy is impaired by increases in both object speed and number of objects tracked (Pylyshyn & Storm, 1988; Yantis, 1992). As measured by capacity (K), the number of items that observers could successfully track ranged from 3 or 4 items when objects were moving slowly to only 1 or 2 items when objects were moving rapidly.

These results also established that the addition of several pictorial depth cues (wire frame, grid floor) and dynamic changes in relative size did not negatively affect tracking accuracy. If anything, there was a trend for improved tracking with larger target numbers in the 3-D condition. Whereas previous studies have reported that occlusion cues and binocular disparity can enhance tracking performance (Viswanathan & Mingolla, 2002), this benefit was only weakly present for the pictorial 3-D cues used here.

Experiment 2: Tracking During a "Wild Ride"

The next step was to manipulate the speed of the objects in the 3-D scene independent of their speed on the retina. We manipulated the speed of the 3-D scene by applying three motion transformations to the scene as a whole. These were *translation* (back and forth movement of the box horizontally across the screen), *rotation* (a swiveling of the box about its central vertical axis), and *zoom* (movement of the box both toward and away from the viewer). Application of all three of these motion transformations to the box of moving objects had the effect of making the box swing, swoop, and rotate (i.e., undergo a "wild ride") while the observer attempted to track the target objects inside the box. Two example video frames from these motion sequences are shown in Figure 3. (Online demonstrations of this experiment can be viewed at http://www.interchange.ubc.ca/vsearch/tracking/.)

The speed of the objects relative to the box boundaries was varied in a way similar to that in Experiment 1. That is, while the box itself was undergoing a complex path of motion, the objects inside it were all moving at 1°/s or 6°/s relative to an imaginary observer who was viewing the box from a distance of 57 cm and keeping a constant viewing angle on the box.

We must also note that prior to collecting data in this condition, we tested tracking accuracy in each of the scene-motion conditions (translation, rotation, zoom) individually. Twenty participants contributed data to each of these conditions, but tracking accuracy did not vary significantly with condition, and none of the conditions was significantly different from the 3-D condition in Experiment 1 (mean tracking accuracy for each condition is summarized in Table 1). For this reason, only the condition in which all three transformations of scene motion were applied simultaneously is presented here in detail.

Method

Participants. Seventeen undergraduate students (13 female, 4 male; mean age = 21.0 years) participated in an hour-long session in exchange for course credit. All reported normal or corrected-to-normal vision.

Stimuli, design, and procedure. Displays consisted of 16 objects moving within a depicted 3-D wire-frame box and were identical to those in the 3-D condition in Experiment 1, with the following modifications. In this and in all subsequent experiments, the moving objects were 3-D spheres that appeared as disks from any given vantage point. Observers were asked

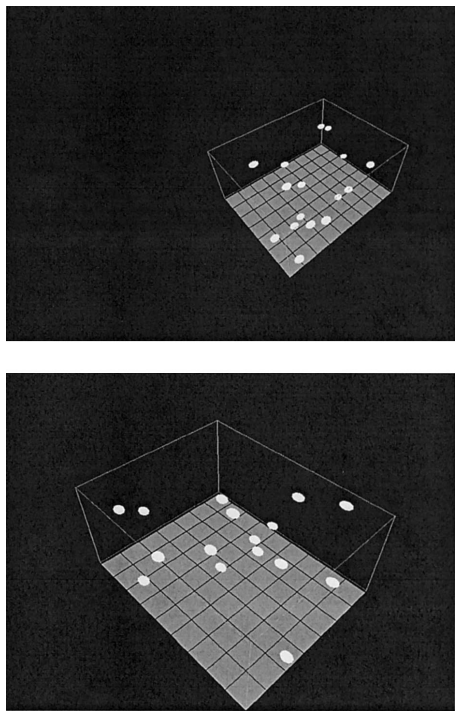

Figure 3. Example video frames of the displays in Experiment 2, consisting of a depicted 3-D wire frame box with a textured floor.

Table 1
Mean Percentages of Accurate Responses (With Standard Errors in Parentheses) as a Function of Scene Speed, Object Speed (Degrees per Second), Number of Objects, and Individual 3-D Transformation in Experiment 2

	1° per second			6° per second		
Scene speed	2 obj.	4 obj.	6 obj.	2 obj.	4 obj.	6 obj.
			Translation			
None	99 (1)	93 (3)	83 (4)	79 (4)	65 (5)	58 (5)
Slow	99 (1)	95 (2)	76 (4)	80 (4)	60 (5)	62 (5)
Fast	98 (1)	98 (1)	77 (3)	78 (4)	58 (5)	65 (4)
			Rotation			
None	98 (1)	89 (3)	73 (4)	83 (4)	64 (5)	62 (5)
Slow	98 (2)	88 (3)	84 (4)	88 (4)	67 (4)	58 (5)
Fast	94 (3)	91 (3)	87 (4)	80 (4)	73 (4)	60 (5)
			Zoom			
None	98 (1)	90 (3)	76 (4)	78 (4)	63 (5)	57 (6)
Slow	98 (1)	87 (4)	73 (4)	78 (5)	57 (4)	53 (5)
Fast	98 (1)	90 (3)	72 (3)	76 (4)	55 (5)	50 (5)

Note. obj. = object.

to track only 2, 4, or 6 objects because Experiment 1 had shown that accuracy for 8 objects was very poor. Finally, because the greatest differences in tracking accuracy were observed with the most extreme object speeds in Experiment 1, only speeds of 1°/s and 6°/s were tested in the current experiment.

Scene-motion transformations. The wire-frame box containing the moving objects underwent motion involving simultaneous changes in translation, rotation, and zoom. Considered singly, each transformation was as follows: Translation involved moving the box horizontally across the screen (*x*-axis motion). The speed of translation was measured by taking the left–right distance in degrees of visual angle traversed across the screen over time. Rotation involved moving the box around its central vertical or *z*-axis. On half of the trials, the box rotated clockwise for the duration of the tracking episode; on the remaining half, it rotated counterclockwise. Rotation speed was measured as the polar angle of rotation around the *z*-axis over time. Zoom involved proportionate expansion and contraction of the box and its contents, which is retinally equivalent to moving the box closer to and further from the observer. Zoom speed was measured as the change in size (in degrees of visual angle) over time. For all types of transformations, when the box changed directions from left to right, clockwise to counterclockwise, or near to far, the speed of the box changed smoothly, following a sinusoidal function.

Scene motions were classified as *no motion*, *slow*, and *fast*. For no motion, the box remained static while objects moved within the confines of the box as they had in Experiment 1 (translation: 0°/s; rotation: 0°/s; zoom: 0°/s). For slow motion, the speeds were 2.40°/s (translation), 2.00°/s (rotation), and 1.60°/s (zoom). For fast motion, the speeds were 3.40°/s (translation), 4.00°/s (rotation), and 3.25°/s (zoom). These three scene-motion conditions were presented randomly and equally often within each block of trials. There were a total of 144 trials (3 scene speeds × 2 object speeds × 3 target numbers × 2 probe types × 4 repetitions).

Results

Accuracy. Mean proportions of correct responses are shown in Figure 4. An ANOVA indicated that tracking accuracy decreased as object speed increased from 1°/s to 6°/s, $F(1, 16) = 93.95$, $p < .001$. Accuracy also decreased as target number increased from 2 to 6, $F(2, 32) = 46.52$, $p < .001$. No other effects, including that of scene speed ($F < 1$), were significant.

Capacity. Capacity decreased as object speed increased from 1°/s ($K = 3.64$) to 6°/s ($K = 2.03$), $F(1, 20) = 32.45$, $p < .0001$. No other effects, including that of scene speed ($F < 1$), were significant.

Discussion

Experiment 2 revealed two main results. First, regardless of whether the 3-D box remained stationary or was moving, tracking was influenced by increases in object speed and number of objects tracked. Faster object motion and a larger number of target objects both reduced tracking accuracy. Second, adding slow or fast motion to the entire scene caused no additional impairment in tracking accuracy. This is inconsistent with tracking being accomplished with a retinotopic frame of reference, because adding scene motion to the object motions resulted in both (a) a marked increase in the retinal motions of many of the objects being tracked for large portions of the tracking episode and (b) a marked increase in the variability of the retinal motions of the various objects in the box. Given the sensitivity of tracking accuracy to variations in object speed shown in both Experiment 1 and the current experiment, it is surprising that these additional variations in speed caused by scene motion had no measurable influence. We can only conclude that object tracking was not based on retinal coordinates but, rather, was based on the speed of objects relative to the boundaries of the box, whether the objects were stationary or moving.

We also note that the excellent tracking accuracy for objects inside the moving box cannot be attributed to observers smoothly pursuing the center of the box with their gaze such that the motion of the objects on the retina would be roughly equal in the stationary and moving conditions. This possibility can be ruled out because only one of the three motion transformations (translation) lends itself to the possibility of maintaining equal retinal motion through smooth pursuit. It is certainly possible to track the center

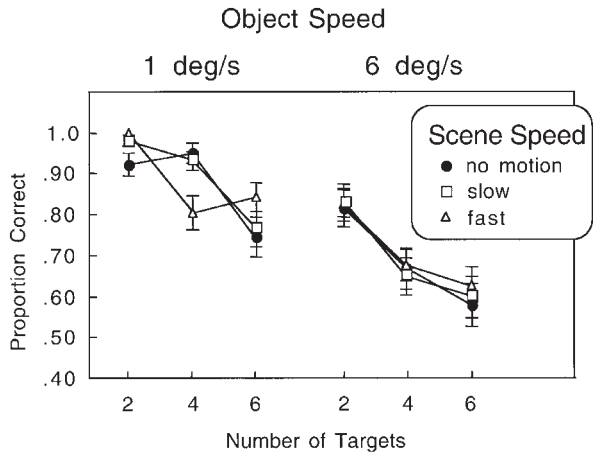

Figure 4. Mean proportions of correct responses (reflecting tracking accuracy) in Experiment 2. Error bars represent plus or minus 1 standard error of the mean. deg = degree.

of the box when it is simply moving back and forth in a predictable pattern across the screen. However, the situation is very different for rotation, in which fixating the center of the box would result in objects generally speeding up their retinal motion as they passed by the center of gaze and generally slowing down their retinal motion as they moved to the outside edges of the box. For zoom, the same strategy of central fixation would result in retinal motions that were slower than usual as the box moved away and faster than usual as the box moved toward the observer.

Our testing of each of these conditions separately (see Table 1) indicated that all three conditions resulted in patterns of tracking accuracy that were not significantly different from the pattern reported for the "wild ride" (see Figure 3), in which all three motions were applied simultaneously. What makes this even more remarkable is the observation that the maximum retinal motions possible in the wild ride were markedly faster even than any of the motions considered separately. Yet this also had no influence on tracking accuracy.

Experiment 3: 3-D Structure of the "Wild Ride"

Tracking accuracy in Experiment 2 was strongly affected by the speed of individual objects, but it was unaffected by the speed of the box in which the objects were moving. This points to an allocentric tracking mechanism, one that tracks objects with respect to their position in the environment rather than with respect to the position and viewpoint of the observer.

If tracking is allocentric, then one way to interfere with it might be to reduce the visual cues supporting the perception of the 3-D space in which objects are moving. In Experiment 2, observers saw objects moving inside the confines of a wire-frame box that included a textured floor. These two features may have provided important pictorial cues regarding the 3-D structure of the box as well as important cues to the nature of the motion path undertaken by the box as a whole. However, we note that at the same time, there were additional 3-D cues intrinsic to the moving objects themselves. For example, objects expanded and contracted slightly as they moved to support the appearance that they were either nearer to (larger) or farther from (smaller) the observer. The moving objects also changed direction every time they encountered the invisible walls of the box. Finally, when the box moved, the moving objects also moved coherently on the screen such that regardless of the individual path being taken by each object within the box, its motion was also consistent with the moving box as a whole.

In Experiment 3, we stripped away the wire frame and the textured floor from the displays to test whether tracking accuracy was dependent on these supports for the perception of a stable 3-D environment. If tracking is impaired by the removal of these features, it would suggest that the visible features of the box are essential for establishing a 3-D environment in which the objects can be tracked. If tracking is unaffected, then perhaps the perception of a stable 3-D environment for the moving objects is less important than we suspect.

Method

Participants. Twenty undergraduate students (14 female, 6 male; mean age = 21.0 years) participated in an hour-long session in exchange for course credit. All reported normal or corrected-to-normal vision.

Stimuli, design, and procedure. All methodological details were identical to those in Experiment 2, with the exception that the wire frame and textured floor of the box were not displayed.

Results

Accuracy. Mean proportions of correct responses are shown in Figure 5. An ANOVA indicated that tracking accuracy decreased as object speed increased from 1°/s to 6°/s, $F(1, 19) = 121.24$, $p < .001$. Accuracy also decreased as target number increased from 2 to 6, $F(2, 38) = 45.93$, $p < .001$. There was also an Object Speed × Target Number interaction, $F(2, 38) = 4.60$, $p < .02$, reflecting that the decrease in accuracy with number of targets was greater for fast than for slow object motion. No other factors—including scene speed, $F(2, 38) = 2.98$— were significant.

Capacity. Capacity decreased as object speed increased from 1°/s ($K = 3.64$) to 6°/s ($K = 2.03$), $F(1, 19) = 87.03$, $p < .001$. No other effects—including that of scene speed, $F(2, 38) = 1.35$—were significant.

We also compared accuracy and capacity measures with those in Experiment 2 with mixed-design ANOVAs to determine whether the presence of the wire frame and textured floor had any effect on tracking performance. No factors involving experiment were significant in either measure (all Fs < 1).

Discussion

As in previous experiments, tracking accuracy was strongly impaired by increases in object speed and in the number of objects to be tracked. But also, as in Experiment 2, tracking accuracy was unaffected by increases in the speed of the box in which the objects moved. This means that tracking accuracy was unaffected by the presence versus absence of the wire frame and the textured floor of the box in which the objects moved. On the face of it, this result is not consistent with an allocentric tracking mechanism.

However, it is also possible the 3-D cues that remained associated with the moving objects themselves were sufficient to support the perception of a coherent 3-D scene. Previous research on the perception of 3-D structure from coherent motion indicates that

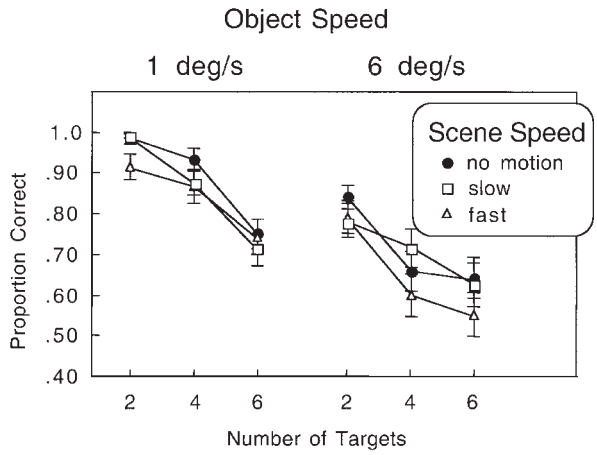

Figure 5. Mean proportions of correct responses (reflecting tracking accuracy) in Experiment 3. Error bars represent plus or minus 1 standard error of the mean. deg = degree.

human observers are adept at interpreting volume from the motion of only three or four points, provided that the points are fixed to the surfaces of a rigid object (Siegel & Andersen, 1988; Treue, Husain, & Andersen, 1991; Ullman, 1979). The differences for the present displays include the following: (a) the moving objects were not rigidly fixed to any part of the box, (b) there were a total of 16 objects moving independently, and (c) the rigid boundaries of the box were indicated by the dynamic changes in direction that occurred for the moving objects when they encountered these "walls." Thus, the perception of a coherent 3-D environment may have still been possible even after the wire frame and the textured floor had been removed from the displays. In short, observers may have been able to recover the structure of the environment solely from the relative motions of the objects. The remaining experiments were therefore designed to disrupt the perceived structure of the scene in even stronger ways.

Experiment 4: Simultaneously Tracking Objects Moving at Two Speeds

The results of Experiment 3 showed that visible cues to the boundaries of the 3-D box were not critical to accurate tracking, because accuracy was still high when the visible cues to the boundaries of the box were eliminated. The next factor that we considered was the increase in the variability of object-motion speeds that occurs when the box is placed in motion. We noted in the *Discussion* section of Experiment 2 that one of the remarkable features of allocentric tracking, from the perspective of the retinal motions involved, is that adding scene motion to the displays resulted in a marked increase in the variability of the retinal motions of the various objects in the box. This did not negatively affect allocentric tracking, likely because it did not increase the variability of the motions relative to the box as a whole. However, we suspected that adding variability to the speeds of individual objects relative to the box might impair tracking accuracy, especially when the entire box of objects was also in motion. If so, it would confirm that one of the factors assisting in the successful tracking of objects in Experiment 3, in which the boundaries of the box were not even visible, was the constant speed of motion of the objects inside the box.

In Experiment 4, half of the objects in a scene moved at one speed while the other half moved at another speed (1°/s or 6°/s). The target objects to be tracked in a scene were also equally divided between slow- and fast-moving objects. The motion of the box as a whole was varied in the same way as in Experiments 2 and 3. If multiple speeds of motion generally impair tracking accuracy, then the presence of two speeds of motion should impair tracking even when the box is stationary. However, if multiple speeds impair tracking accuracy only when the box itself is in motion, this would indicate that the constant speed of motion in previous experiments was a contributing factor to the establishment of a coherent 3-D scene in which objects were moving.

Previous studies of multiple-object tracking have had objects within a scene moving either at different speeds (Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999) or at the same speed (Viswanathan & Mingolla, 2002; Yantis, 1992), but none have directly compared these two conditions. Previous studies have also not distinguished between increased variability in allocentric and in retinal speed of motion.

Method

Participants. Twenty undergraduate students (10 female, 10 male; mean age = 21.7 years) participated in an hour-long session in exchange for course credit. All reported normal or corrected-to-normal vision.

Stimuli, design, and procedure. Displays and other details of the method were identical to those in Experiment 2, except that all objects, including targets, now moved at one of two speeds inside the box (1°/s or 6°/s). The target probe at the end of the tracking episode therefore also varied randomly between being a slow or a fast target.

Results

Accuracy. Mean proportions of correct responses are shown in Figure 6. An ANOVA indicated that tracking accuracy did not vary with object speed in this experiment ($F < 1$). This is likely because observers were tracking objects of both speeds on every trial and had to be prepared to respond to either a slow or a fast probe. However, as in the previous experiments, accuracy decreased as target number increased from 2 to 6, $F(2, 38) = 65.72$, $p < .001$.

Most important, accuracy decreased as the motion of the scene increased from no motion to slow to fast, $F(2, 38) = 8.92$, $p < .001$. This was the first time that this effect was observed in this study, suggesting that having multiple speeds of allocentric motion impairs the perceived structure of the 3-D scene. This interpretation is strengthened by a Scene Speed × Object Speed interaction, $F(2, 38) = 3.45$, $p < .05$, which reflects a larger impairment of object speed when the box was in fast motion than when the box was stationary. This interpretation is also strengthened by a Scene Speed × Number Tracked interaction, $F(2, 38) = 8.92$, $p < .001$, which reflects an exaggerated decrease in accuracy for larger numbers of targets when the box was in motion. No other effects were significant.

Capacity. Capacity did not vary with object speed ($F < 1$), but capacity did decrease with increases in scene speed: no motion ($K = 3.36$), slow motion ($K = 2.59$), fast motion ($K = 2.46$), $F(2, 38) = 5.48$, $p < .03$. A direct comparison of these capacity estimates with those in Experiment 2 revealed a significant Ex-

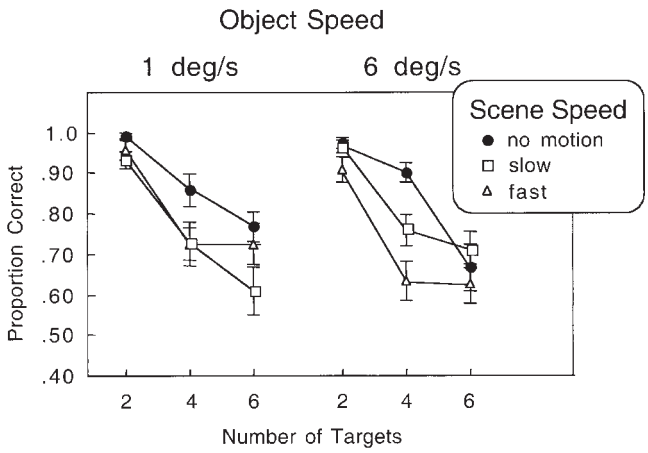

Figure 6. Mean proportions of correct responses (reflecting tracking accuracy) in Experiment 4. Error bars represent plus or minus 1 standard error of the mean. deg = degree.

periment × Scene Motion interaction, $F(2, 78) = 4.13$, $p < .02$. Whereas capacity was unaffected by scene motion in Experiment 2 (mean $K = 2.85$), it decreased significantly with increases in scene motion in the present experiment.

Discussion

These results indicate that an important factor contributing to allocentric object tracking is the perception of a coherent 3-D scene. When the perception of this coherent scene was impaired in the present experiment by increased variability in object speeds relative to the box as a whole, tracking accuracy was reduced. It is notable that the mere existence of two speeds of motion was in itself not detrimental to tracking accuracy. When the box was stationary, tracking accuracy for targets moving at two speeds was comparable to that observed in previous experiments, in which only one speed of motion was present in any given display. Yet the simultaneous presence of these two speeds of object motion was detrimental to tracking accuracy when the box of moving objects was itself moving. This indicates that the constant speed of object motion within a scene is an important cue to the perceived structure of the 3-D space in which objects are moving.

Experiment 5: Tracking Objects in a Nonrigid 3-D Space

Experiment 4 suggested that allocentric object tracking depends on the perception of a coherent 3-D scene. Experiment 5 put this idea to a further test by examining tracking accuracy when the 3-D space in which objects move appears to be unstable. The main manipulation was inspired by the observation that shape and space constancy break down when a scene is viewed from more than one vantage point (Cavanagh, Peters, & von Grünau, 1988; Cavanagh & von Grünau, 1989).

The general failure of shape constancy when a scene is viewed from more than one vantage point can be easily illustrated by placing a vertical fold in a $20 bill, centered on the face of the individual on the bill. The fold can be either concave (open book) or convex (book spine) with respect to the viewer. While you are viewing the folded face, slowly rotate the bill around its horizontal axis. You will notice that the facial expression of the depicted person changes dramatically as the bill is rotated. Now smooth out the fold in the bill and view the face again while slowly rotating the bill both horizontally and vertically. There is no longer any change in the expression of the face. Taken together, these two conditions illustrate that shape constancy is readily achieved when an image is viewed from a wide range of vantage points (the smooth bill), even though these vantage points may distort the retinal image of the scene considerably. However, shape constancy breaks down as soon as the scene is viewed from two or more vantage points, or when the image of the scene is projected onto two or more surfaces, as occurs when the bill is folded.

In Experiment 5, we projected the image of the tracking displays onto the convex corner of a dihedral surface, as illustrated in Figure 7. It is important to note that we had observers perform the tracking task on these displays from the same vantage point as the projector to ensure that the retinal projection in this experiment was roughly equivalent to what it had been in previous experiments. Yet, despite this equivalence at a retinal level, projecting the image in this way had the effect of greatly distorting the perceived structure of the space in which the objects were moving. The distortion included a rubber-like bending of the 3-D box and large apparent changes in object speed when objects crossed the folded center of the projection zone. If tracking is based on retinal coordinates, then tracking accuracy here should be comparable to that in Experiments 3 and 4. However, if tracking is allocentric, then accuracy should be impaired by the reduction in scene coherence.

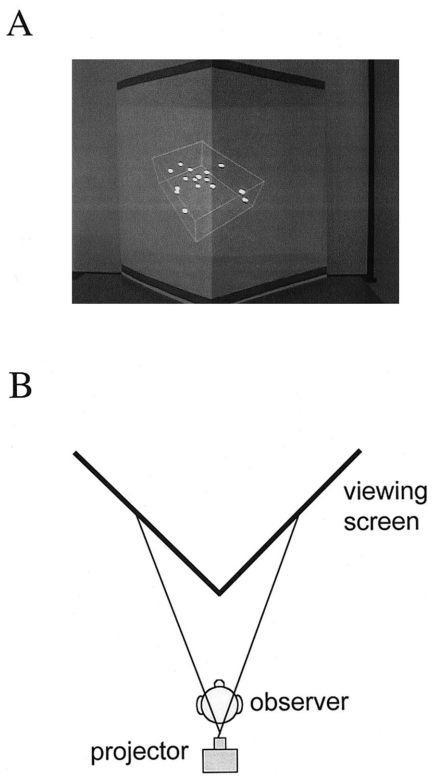

Figure 7. In Experiment 5, displays were projected onto a convex corner to test whether tracking accuracy would be influenced by the coherence of the 3-D scene. Panel A shows the display from the observer's perspective. Panel B depicts a bird's-eye view of the setup.

Method

Participants. Fourteen undergraduate students (7 female, 7 male; mean age = 23.8 years) participated in an hour-long session in exchange for course credit. All reported normal or corrected-to-normal vision.

Apparatus. Displays were projected onto a convex corner using an NEC MultiSync projector positioned 225 cm from the corner. The convex projection screen consisted of two white (3-ft × 5-ft) foam-core boards connected at a 90° angle. To help reinforce the spatial layout of the projection screen, a 1-in.-wide green ribbon outlined the screen and a 60-W lamp illuminated the screen from overhead on the right side, casting a noticeable shadow on the left side of the screen. Observers sat with their heads positioned directly underneath the projector, at a viewing distance of 171 cm, to ensure a projection that was nearly equivalent to that obtained when the displays were viewed on a computer screen. The image of the projector was focused for a depth of field that was 225 cm away.

Stimuli, design, and procedure. With the exception of the details of projection, the displays and other details of the method were identical to those in Experiment 2.

Results

Accuracy. Mean proportions of correct responses are shown in Figure 8. An ANOVA indicated that tracking accuracy decreased with increases in object speed, $F(1, 13) = 34.65$, $p < .001$, as had been found in Experiments 1–3. Also, as in previous experiments, accuracy decreased as target number increased, $F(2, 26) = 38.07$, $p < .001$. The critical result for this experiment, however, was that tracking accuracy decreased with increases in scene speed, $F(2, 26) = 7.28$, $p < .01$. This finding indicates that reducing the coherence of the 3-D scene impaired tracking accuracy even though the retinal speeds of motion were identical to those in other conditions in which scene speed was not a factor (Experiments 2 and 3).

This interpretation is strengthened by a Scene Speed × Object Speed interaction, $F(2, 26) = 3.13$, $p < .06$, which reflects a larger impairment of object speed when the box was in fast motion than when the box was stationary. It is also strengthened by a Scene Speed × Number Tracked interaction, $F(2, 26) = 2.97$, $p < .03$, which reflects an exaggerated decrease in accuracy for larger numbers of targets when the box was in motion. No other effects were significant.

Capacity. Capacity was impaired both by increases in object speed (1°/s, $K = 3.08$; 6°/s, $K = 1.65$), $F(1, 13) = 29.91$, $p < .001$, and increases in scene speed (no motion, $K = 2.92$; slow motion, $K = 2.27$; fast motion, $K = 1.90$), $F(2, 26) = 5.02$, $p < .02$. There was also a Scene Speed × Object Speed interaction, $F(2, 26) = 5.51$, $p < .01$, reflecting the fact that differences in scene speed were greater when the objects were moving slowly than when they were moving quickly. This is likely because accuracy in the fast-object condition was already so low (near the chance level of 50%) that the full effects of scene speed could no longer be measured.

A direct comparison of these tracking-capacity estimates with those in Experiment 2 revealed an Experiment × Scene Motion interaction, $F(2, 66) = 3.78$, $p < .03$. Whereas capacity was unaffected by scene motion in Experiment 2 (mean $K = 2.85$), it decreased significantly with increases in scene motion in the present experiment.

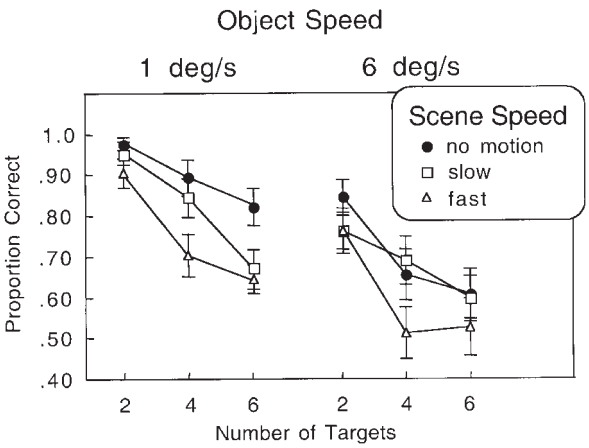

Figure 8. Mean proportions of correct responses (reflecting tracking accuracy) in Experiment 5. Error bars represent plus or minus 1 standard error of the mean. deg = degree.

Discussion

These results confirm that allocentric object tracking depends on the perception of a coherent 3-D scene. Scene perception was impaired in this experiment by projection of the image of the moving objects onto a convex corner, a manipulation that is known to sharply impair recovery of the 3-D structure of a scene. At the same time, the retinal relations among the moving objects were preserved by this manipulation. Notably, the projection had little if any influence on tracking objects when the box was stationary. However, it sharply reduced tracking accuracy when the box was in motion. These results are all consistent with the idea that allocentric object tracking depends on the perception of a coherent 3-D scene.

Experiment 6: Preserving Scene Versus Retinal Consistency

In contrast to the previous experiment, in which perception of the 3-D scene was distorted while the retinal relations among the moving objects were preserved, Experiment 6 tested two conditions in which the retinal relations among moving objects were distorted relative to those in previous experiments. Condition A involved the projection of the tracking displays onto a flat screen at an oblique angle, as illustrated in Figure 9A. Observers also viewed these displays from an oblique angle, one that was 30° away from the angle of projection. Because the image was projected onto a single plane, it was expected that participants would be able to maintain the perception of a stable scene, much as they could if they were viewing a $20 bill at an oblique angle (Cavanagh et al., 1988). They should have been able to do this even though the retinal relations among moving objects now differed from those in the previous experiments. Thus, Condition A examined tracking accuracy when the stability of the scene was preserved and the retinal relations were distorted.

Condition B tested a dihedral viewing condition similar to that in Experiment 5, with the exception that observers now viewed the display from an angle 30° different from the angle of projection, as shown in Figure 9B. This meant that the 3-D coherence of the scene was distorted, because of the dihedral projection planes, and that the retinal relations among objects were distorted, because of the oblique viewing angle relative to the projector. If tracking is allocentric, then accuracy should be impaired in Condition B relative to Condition A. Alternatively, if tracking is based on retinal coordinates, then accuracy should be comparably impaired in both conditions relative to Experiment 5.

Method

Participants. Twenty-eight undergraduate students (17 female, 11 male; mean age = 24.2 years) participated in an hour-long session in exchange for course credit. All reported normal or corrected-to-normal vision. Fourteen students participated in each of the two conditions, but the data of 4 of the students in Condition B were not analyzed because these participants failed to complete the testing session after complaining that the task was too difficult.

Stimuli, design, and procedure. With the exception of the details of projection, the displays and other details of the method were identical to those in Experiment 5. In Condition A, the projector was 171 cm from the center of projection on the viewing screen, positioned at a 45° angle with

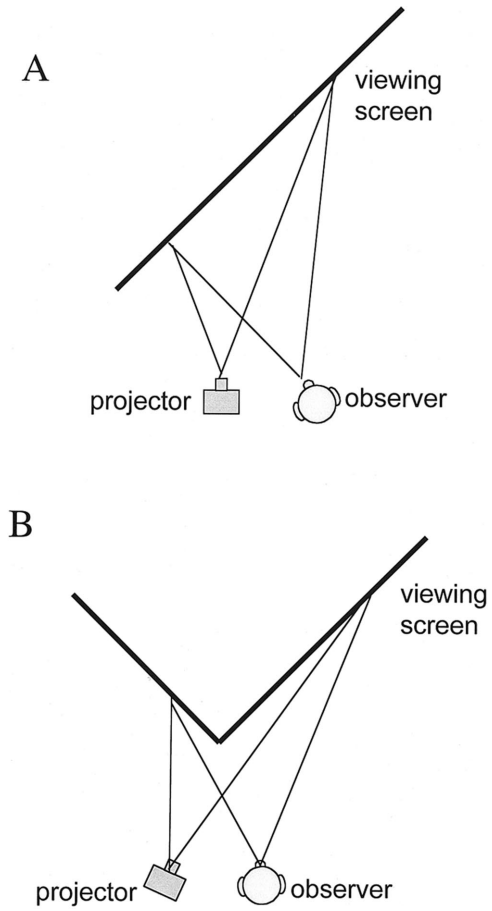

Figure 9. In Experiment 6, displays were projected either onto a single surface lying obliquely with respect to the projector (A) or onto a dihedral pair of surfaces (B). In both cases, the line of projection differed from the line of sight by 30°.

respect to the screen. Observers sat with their heads 225 cm from the center of projection, and their line of sight was 30° from the line of projection. In Condition B, the same viewing conditions were used as in Experiment 5, with the exception that the projector and the observer were both moved 15° away from the center of the screen, in opposite directions. Both the projector and the observer were oriented toward the nearest point on the projection screen.

Results

Accuracy. Mean proportions of correct responses are shown in Figure 10. An ANOVA indicated that tracking accuracy decreased with increased object speed, $F(1, 22) = 187.97$, $p < .001$; increased scene speed, $F(2, 244) = 6.43$, $p < .01$; and as target number increased, $F(2, 44) = 62.84$, $p < .001$, as reported in previous experiments. The critical result for this experiment, however, was that overall tracking accuracy was greater in Condition A ($M = 79\%$) than in Condition B ($M = 71\%$), $F(1, 22) = 8.96$, $p < .01$. There was also a significant Condition × Number Tracked interaction, $F(1, 22) = 4.54$, $p < .02$, because the accuracy differences between conditions grew larger along with the number of targets to be tracked. There was also a Condition × Scene Speed interaction, $F(2, 44) = 42.65$, $p < .08$, reflecting the greater influence of scene speed on Condition B than on Condition A. Simple effects tests indicated that whereas scene speed was not significant in Condition A ($F < 1$), it was highly significant in Condition B, $F(2, 18) = 7.03$, $p < .01$. Finally, direct comparisons indicated that Condition A of the present experiment resulted in improved tracking accuracy relative to that in Experiment 5 (in which retinal relations among objects were preserved but scene stability was undermined): The main effect of experiment was nonsignificant, $F(1, 26) = 3.48$, $p < .07$; the Experiment × Number Tracked interaction was significant, $F(2, 52) = 3.33$, $p < .05$; and the Experiment × Scene Speed interaction was nonsignificant, $F(2, 52) = 2.95$, $p < .06$. In contrast, similar comparisons indicated that accuracy in Experiment 5 was not significantly different from that in Condition B: main effect of experiment, $F(1, 22) = 1.47$; Experiment × Number Tracked and Experiment × Scene Speed interactions (Fs < 1).

Capacity. Analyses of capacity pointed to a similar picture. Mean tracking capacity was greater in Condition A ($K = 3.1$) than in Condition B ($K = 2.1$), $F(1, 22) = 9.99$, $p < .01$, and increases in object speed resulted in reduced tracking accuracy (1°/s, $K = 3.48$; 6°/s, $K = 1.74$), $F(1, 22) = 115.47$, $p < .001$. Direct comparisons with Experiment 5 indicated that tracking accuracy was significantly greater in Condition A, $F(1, 35) = 5.36$, $p < .03$, and was similar in Condition B, $F(1, 35) = 1.02$, $p < .30$.

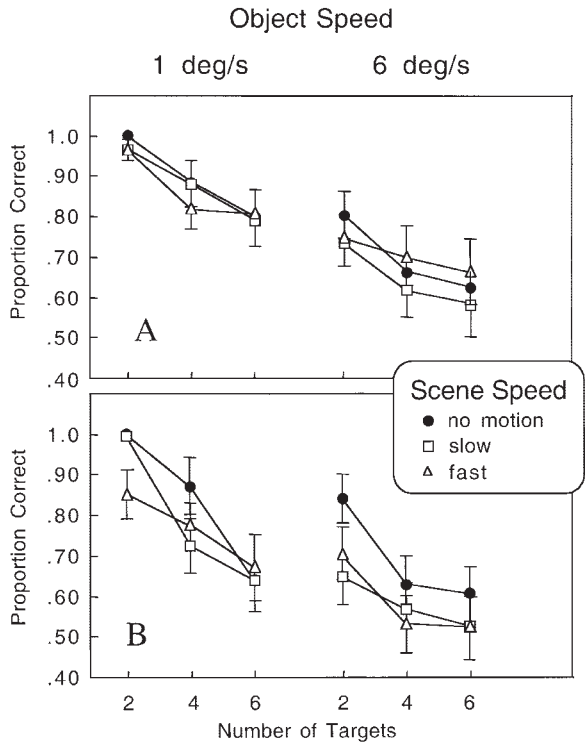

Figure 10. Mean proportions of correct responses (reflecting tracking accuracy) for images of moving objects projected onto a single surface lying obliquely with respect to the projector (A) and onto a dihedral pair of surfaces viewed obliquely (B) in Experiment 6. Error bars represent plus or minus 1 standard error of the mean. deg = degree.

Discussion

These results provide further confirmation of an allocentric tracking mechanism based on coherent scene perception. In Condition A, scene coherence was maintained through oblique projection onto a single surface, but retinal coherence was disrupted through oblique viewing; this resulted in improved tracking relative to that in Experiment 5, in which scene coherence was disrupted. This shows that just as adding greater variability to retinal motions through the "wild ride" does not impair tracking (Experiment 2), disrupting retinal coherence through oblique viewing also does not impair tracking. However, when scene coherence and retinal coherence are disrupted through oblique projection onto a convex corner and oblique viewing, tracking accuracy drops to levels comparable to those in Experiment 5, in which only scene coherence was disrupted. This suggests that scene coherence is indeed the critical factor in maintaining accurate tracking.

General Discussion

The question addressed by this study was whether multiple-object tracking is based on a retinotopic or allocentric spatial frame of reference. Although there has been much interest in the capacity of human vision to keep track of a small number of objects in a dynamically changing environment (Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999; Yantis, 1992), this question has not been a research focus to date.

Our approach to addressing this question was based on the fact that tracking accuracy generally declines as the speed of objects to be tracked is increased. However, in the studies used to establish this finding, the motion of objects on the retina (assuming a fixed gaze) and the motion of objects relative to a scene have not been examined separately. We decoupled these factors in the present study by designing an object-tracking environment in which the motion of objects relative to a scene could be varied independently of their motion on the retina. This was accomplished by moving the entire scene of objects in complex ways across the viewing screen. We were then able to ask whether the motion of the scene as a whole had an influence on tracking accuracy over and above the influence of the motions of the objects within the scene.

The clear answer that we obtained after measuring the influence of three different scene motions—translation, rotation in depth, and zoom (both alone and in combination)—was that motion of the scene as a whole had negligible effects on tracking accuracy. In short, the number of objects that could be tracked successfully was not different when the depicted 3-D box in which the objects moved was stationary or when it was moving in complex ways on the screen. This is an important result, both for understanding of the human visual system and for the application of current knowledge about human vision to the problems of human–computer interaction. In the discussion that follows, we address implications of this finding for both of these areas in turn.

Implications of Allocentric Tracking for Human Vision

A first implication of allocentric object tracking is that tracking must depend on continual visual interaction with a perceived environment. That is, if the spatial reference is not to be found in the eye or in some other feature linked to the viewer's ego center, then it must be tied to the environment. This immediately leads to numerous questions of interest, such as

> Which visual cues are used to perceive the layout or structure of this environment during tracking?
>
> How richly and for how long is that environment represented in the visual system?
>
> To what extent is visual interaction with the environment based on past experience versus being based only on the most recently updated information?

These and many other questions relevant to the finding of allocentric tracking will now have to be addressed.

In the present series of experiments, we have only begun to address these issues. The experiments reported here focused primarily on whether the stability of the perceived 3-D environment was important. We first looked at whether any cues from the environment itself needed to be explicitly presented. When we removed all external visual support for the scene in Experiment 3, which included the wire-frame box and the textured floor, we found that (a) tracking accuracy was still allocentric (not significantly affected by complex motion of the box as a whole) and (b) tracking accuracy was comparable to that obtained when these supporting features of the 3-D scene were visible. This means that the structural cues remaining in the moving objects themselves—including changes in object size consistent with relative distance and constant motion trajectories that changed only when the invisible walls of the box were encountered—were sufficient to allow observers to maintain the perception of a coherent 3-D environment.

Our second approach involved presenting all of the visible scene structure (wire frame and textured floor) but reducing the coherence of the object motions internal to the box by allowing objects to move at one of two different speeds (Experiment 4). This had no negative effect on tracking accuracy when the 3-D environment was stationary, as would be expected on the basis of previous studies (Pylyshyn & Storm, 1988). However, the presence of two different object motions led to a marked reduction in tracking accuracy when the box as a whole was in motion. The fact that the perceived coherence of the 3-D scene was also much reduced in this condition suggests that scene stability is obtained from the relative positions of multiple objects in the scene and that relative object position is therefore a critical factor in sustaining high levels of tracking accuracy.

This hypothesis was put to a direct test in Experiment 5, in which scene stability was reduced by projecting the scene's image onto the junction of two dihedral surfaces. Despite the fact that the retinal projection was still identical to that in previous conditions in which tracking accuracy was high (Experiments 2 and 3), tracking accuracy was again reduced when the scene was placed in motion under these conditions. The apparent plastic transformations of the scene structure that occurred when the box was in motion resulted in a significant decrease in tracking accuracy. Taken together, these results converge on the conclusion that multiple-object tracking is accomplished using allocentric spatial references rather than a retinal frame of reference. These spatial references can be fairly abstract—as seen in Experiment 3, in

which all explicit evidence regarding the box was removed—but they do need to be stable, as indicated by the results from Experiments 4–6.

With the fact that a stable scene is critical to accurate object tracking established, it is of interest to consider the perspective that this gives to some of the extant findings regarding object tracking in the literature. Consider first the finding that whether or not observers are permitted to move their eyes during a tracking episode has no effect on tracking accuracy (Pylyshyn & Storm, 1988). Maintaining high levels of tracking accuracy when eye movements are permitted would require an additional mental operation if tracking is fundamentally retinotopic. However, for an allocentric mechanism, eye movements are not an impediment to tracking unless the periods between eye fixations become so long that they exceed the capacity of the system to maintain an updated representation of the environment. In fact, eye movements may even improve tracking accuracy, especially if they benefit the perception of the stable environment in which the objects are moving (e.g., by providing greater spatial detail of the environment).

Consider also the finding that tracking accuracy remains high despite the presence of surfaces in a scene that temporally occlude vision of the objects to be tracked (Scholl & Pylyshyn, 1999). Again, an allocentric tracking mechanism would only be disturbed by occluded objects if they interfered with either the ability to maintain a representation of the object (if the occluded period exceeded the temporal capacity of the system) or the perception of a stable environment. Again, to the extent that visual occlusion is a rich cue to the 3-D structure of an environment, there is also the real possibility that visual occlusion could be used to enhance tracking accuracy through its effect in reinforcing the coherence and stability of the environment in which objects are moving. The finding that the addition of binocular disparity to a display can improve tracking accuracy (Viswanathan & Mingolla, 2002) is completely consistent with this point.

Finally, it is worth considering how the perspective of allocentric tracking may alter one's interpretation of the finding that tracking becomes very difficult when the entities to be tracked undergo plastic transformations or are deformable substances rather than rigid objects (vanMarle & Scholl, 2003). The earlier work by vanMarle and Scholl was originally presented to show that the visual tracking mechanism accepts as inputs only objects that are "rigid" and "cohesive." The present study raises the possibility that the critical stability may lie in the environment in which the objects are moving, not necessarily in the items themselves that are being tracked. In vanMarle and Scholl's experiments, the deformable substances were moving across a 2-D surface with no depicted 3-D scene structure. The present experiments indicate that maintaining a coherent scene structure is critical to tracking ability. This raises the possibility that if the deformable substances were moving across a recoverable, nonuniform 3-D surface (e.g., a surface of hills and valleys), and the objects underwent retinal deformations in keeping with this surface structure, then objects should be very easy to track. That is, one would expect tracking to be robust despite the fact that the retinal projections of these objects would undergo considerable plastic retinal transformation as they became more fully visible or occluded and as they changed in their apparent distance. Important control conditions that would also have to be tested include these same object deformations in the absence of an interpretable landscape. Experiments like this clearly need to be done to determine the nature of the representations used in object tracking. The main benefit of the present finding, that tracking is allocentric under at least some circumstances, is to clarify and focus the questions that need to be addressed next.

Implications of Allocentric Tracking for Human–Computer Interaction

The finding of robust tracking despite large changes in the observer's vantage point on the scene speaks favorably for the design of shared-user environments in which users must track objects while scene changes or viewpoint changes are also occurring. One example of where such an environment could be used is in the "free flight" protocol that the Federal Aviation Authority is currently implementing for airline pilots and air traffic controllers. One of the features of this proposal is that it will give air traffic controllers and airline pilots a shared visualization of airspace in which to make decisions (Radio Technical Commission for Aeronautics, 1995, 1997). The present results suggest that, provided the scene or viewpoint changes are smooth and predictable, multiple users will be able to accurately track objects despite these changes.

One important issue highlighted by the present results is that tracking accuracy varies directly with the speed of the moving objects in their perceived environments. This general point—that tracking accuracy increases with reduced speed—has been made previously in the context of 2-D displays (Pylyshyn & Storm, 1988; Yantis, 1992). What the present results make clear is that environmental speed, rather than retinal speed, is the critical variable. In the context of a shared visualization environment such as air traffic control, in which object speeds are typically quite slow, this means that tracking capacity may actually be quite large, likely exceeding the three- or four-object limit observed in the present study, in which the slowest speeds tested were between 1 and 2 polar degrees per second. Future experiments will be needed to determine how the tracking of these relatively slow objects is influenced by the presence of objects moving at different speeds (Experiment 4) and by the presence of motion paths that are both more and less predictable than those tested here.

Another important question for future research concerns the role of user control in scene changes. In user-shared environments, one user will sometimes need to take control of changes in the scene. One way to explore this question would be to compare tracking accuracy when changes in the scene occur because of a voluntary decision made by a user (active change) and when changes in the scene are unexpected (passive changes). It is already known that if object trajectories are unpredictable, tracking is unaffected (Pylyshyn & Storm, 1988; Yantis, 1992). But what is not known is how important the predictability of the scene is; the present results suggest that scene stability may matter much more than the predictability of object motion. A related question is whether smooth changes in viewpoint are critical to robust tracking ability. To the extent that observers are able to represent a scene as a stable environment, the present finding of allocentric tracking suggests that observers may still be able to track accurately if viewpoint changes are unpredictable or abrupt. Our finding that tracking accuracy was unaffected by whether the scene motion was slow or fast points to the possibility that accurate tracking will survive

even "wilder" viewpoint changes, provided the scene structure and the relative positions of the objects within the scene remain stable over time.

A final implication pertains to the rendering of 3-D environments for dynamic scene visualization. The fact that tracking was unimpaired when important cues to the 3-D scene (i.e., wire frame, textured floor) were removed suggests that additional cues to 3-D structure may not be necessary, provided that certain critical elements of the scene structure remain in place or that the relative positions of objects in the scene remain stable over short periods of time. It will be important for future research to determine which visual cues are essential for creating a sufficiently stable environment to support accurate tracking. The present finding that tracking accuracy was impaired by variable object speeds when the scene was in motion suggests that the maintenance of relative positions among objects may be one of these critical cues.

References

Cavanagh, P., Peters, S. C., & von Grünau, M. (1988). Rigidity failure and its effect on the Queen. *Perception, 17,* A27.

Cavanagh, P., & von Grünau, M. (1989). 3-D objects that appear nonrigid during rotation. *Investigative Ophthalmology and Visual Science, 30*(Suppl.), 263.

Deubel, H., Bridgeman, B., & Schneider, W. X. (1998). Immediate postsaccadic information mediates space constancy. *Vision Research, 38,* 3147–3159.

Fecteau, J. H., Chua, R., Franks, I., & Enns, J. T. (2001). Visual awareness and the on-line modification of action. *Canadian Journal of Experimental Psychology, 55,* 104–110.

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27,* 889–935.

Li, L., & Warren, W. H. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research, 40,* 3873–3894.

Liu, G., Healey, C. G., & Enns, J. T. (2003). Target detection and localization in visual search: A dual systems perspective. *Perception & Psychophysics, 65,* 678–694.

Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception & Psychophysics, 44,* 369–378.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197.

Radio Technical Commission for Aeronautics. (1995). *Final report of RTCA Task Force 3 free flight implementation.* Washington DC: Author.

Radio Technical Commission for Aeronautics. (1997). *Government/industry operational concept for the evolution of free flight.* Washington DC: Author.

Raymond, J. E., Shapiro, K. L., & Rose, D. J. (1984). Optokinetic backgrounds affect perceived velocity during ocular tracking. *Perception & Psychophysics, 36,* 221–224.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Scholl, B. J., & Pylyshyn, Z. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38,* 259–290.

Siegel, R. M., & Andersen, R. A. (1988, January 21). Perception of three-dimensional structure from motion in monkey and man. *Nature, 331,* 259–261.

Treue, S., Husain, M., & Andersen, R. (1991). Human perception of structure from motion. *Vision Research, 31,* 59–75.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London, Series B, 203,* 405–426.

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B. H., Bakircioglu, M., & Miller, M. I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Research, 41,* 1359–1378.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science, 14,* 498–504.

Viswanathan, L., & Mingolla, E. (1998). *Attention in depth: Disparity and occlusion cues facilitate multi-element visual tracking* (Tech. Rep. No. CAS/CNS-98–012). Boston: Boston University Center for Adaptive Systems, Department of Cognitive and Neural Systems.

Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception, 31,* 1415–1437.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340.

Received April 29, 2004
Accepted August 6, 2004 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!