

# Attentive Navigation for Viewpoint Control in Virtual Environments

Stephen Hughes and Michael Lewis, University of Pittsburgh, Pittsburgh, Pennsylvania

Three-dimensional virtual environments and teleoperation activities depend on the ability to position an egocentric viewpoint at meaningful locations. Evidence from foundational research has led to the development of control strategies that generally adopt an all-or-nothing approach to guiding the viewer to useful vistas. This work examines *attentive navigation*, a technique for partially automating viewpoint control to promote a supportive yet unscripted exploration of a virtual environment. A collection of three experiments was designed to assess the effectiveness of attentive navigation versus commonly used free navigation techniques. Results show that this approach has promising consequences for improving object recognition, developing an understanding of the configuration of objects, and searching for target objects. Actual or potential applications of this research include improved interaction techniques for extracting information from visual representations, virtual environments, and teleoperation experiences.

## INTRODUCTION

Three-dimensional virtual environments (VEs) and teleoperated systems are frequently characterized by the use of an egocentric viewpoint that moves through the environment, offering viewers multiple perspectives on a visual information display. Numerous applications have been developed that use VEs to allow exploration of real-world phenomena, including galaxies, molecules, architectural layouts, and battlefields (Hix et al., 1999). Teleoperation applications include terrestrial, undersea, and aerial robots as well as long-latency planetary rovers (Sheridan, 1992). Searching for and learning from objects in a VE or teleoperation requires both movement and control of gaze. Although people routinely scan their surroundings using eye and head movements as they walk, there are no corresponding natural movements of the viewpoint in desktop VE or teleoperation systems. As a consequence it is more difficult to search for objects and landmarks. This problem is exacerbated in teleoperation activities, which are also frequently constrained with a much narrower field of view. Although these reported experiments involve

VEs, our results can be readily extended to teleoperation in which automatic target recognition or other feature detection information is available.

Whether the motivation is danger, cost, or constraints of the physical world, VEs are well suited to provide viewers with two types of information: survey knowledge (where things are) and object knowledge (what is there). Siegel and White (1975) identified survey knowledge as the ability to understand the global organizational structure of a collection of objects. Survey knowledge is often likened to having a mental representation of a map that allows viewers to mentally navigate the space. Although survey knowledge can be obtained solely by studying maps, a more thorough understanding can be achieved by actually navigating in the environment (Arthur, Hancock, & Telke, 1996; Thorndyke & Hayes-Roth, 1982). It is important to note that true navigation requires the active engagement of the viewer: it is not enough to automatically move a viewer through an environment as if on a tour bus. Peruch, Vercher, and Guthier (1995) determined that self-controlled viewers tended to develop a rich survey knowledge more quickly than did passive observers.



Helping users acquire object knowledge represents a challenge that is common to the design of all graphical displays: facilitating the viewer's ability to extract relevant information from the surrounding context. This is commonly known among visualization researchers as the "focus-plus-context" problem (Card, Mackinlay, & Shneiderman, 1999). This issue has been well researched in the domain of two-dimensional (2-D) displays. Many approaches to this issue allow the viewer to distort or magnify a portion of the display to allow for closer inspection (Keahey, 1998). Other methods attempt to provide visual cues to highlight or augment specific areas of interest (Zhai, Wright, Selker, & Kelin, 1997). Although these approaches work well if the important features are always within the field of view, they may not be suited to large VEs, in which part of the model can be occluded or otherwise out of sight. Furthermore, deliberately introducing distortions may impair the viewer's ability to gain accurate survey knowledge.

To locate objects in 3-D environments requires that (a) the viewer is positioned in a location from which the object can potentially be viewed, (b) the viewpoint is subsequently oriented such that the relevant objects are prominently displayed within the field of view, and (c) Objectives a and b are met while maintaining the spatial integrity of the environment. Unfortunately, this is a difficult task for a person to do alone. The operator is often faced with the cognitively demanding task of controlling six viewing parameters: position ( $x$ ,  $y$ ,  $z$ ) and orientation (yaw, pitch, and roll). Arthur, Hancock, and Chrysler (1993) pointed out that "when interaction becomes highly attention demanding, memory for the present location frequently decays, with the result that the individual becomes lost in space" (p. 328). Moreover, Goerger et al. (1998) found that self-navigating viewers in information-rich VEs are particularly susceptible to superfluous data and are easily distracted. If viewers are left to control their own viewpoint, these distractions can significantly impair their ability to complete searching tasks.

Although motion through scenes is essential to both kinds of information, the amount of self-determination the viewer has over the viewpoint is critical. Survey knowledge flourishes when the viewer is able to freely explore the environment,

but important details may be overlooked if guidance is not provided. Numerous methods have been developed for viewpoint control in VEs, but they tend to cater to one of two extremes. On one hand, guided navigation techniques automatically move the viewpoint through the environment, often relying on guidelines from cinematography to ensure that relevant imagery is presented from the model (Bares, Thainimit, & McDermott, 2000; Drucker & Zeltzer, 1994; Halper & Oliver, 2000; He, Cohen, & Salesin, 1996). On the other hand, research into free navigation attempts to minimize cognitive overhead and physical interaction in the hope that the viewer's resources will be more directed to positioning the viewpoint in meaningful locations. Several dominant metaphors for viewpoint control have emerged, including "eyeball in hand," "flying vehicle" (Ware & Osborne, 1990), and "gaze-directed steering" (Bowman, Koller, & Hodges, 1997).

These approaches present an all-or-nothing approach to the amount of self-determination that the viewer can exert when exploring the environment. Some recent investigations attempted to provide hybrid systems, allowing users to toggle between exploration and guided modes. Beckhaus, Ritter, and Strothotte (2001) described a guided tour system that can be interrupted to allow viewers to explore an area of local interest. When the viewer is finished, the tour is capable of automatically resuming from wherever the viewer left off.

Very little attention has been paid to the idea of partially automating viewpoint control to promote a supportive yet unscripted exploration of a virtual environment. This work explores this possibility through an approach known as *attentive navigation*. This technique for viewpoint motion grants the viewer an "appropriate" level of control while the system suggests optimal viewing parameters. The following sections of this paper discuss specific details of the interaction model and present the results of three experiments designed to evaluate the effectiveness of this technique for promoting spatial knowledge.

## ATTENTIVE NAVIGATION

Attentive navigation attempts to reconcile the issues surrounding self-control in VEs by sharing control of the viewpoint between the viewer



and the designer of the environment. The viewer effectively controls the position in the environment, and the system offers viewpoint suggestions based on the user's context within the environment. The self-determined motion allows the viewer to benefit from multiple orientations and receive the information on demand. At the same time, the technique can support the viewer's goals by focusing the viewpoint on elements of the scene that build knowledge while impeding perspectives that detract from learning.

Attentive navigation is an implementation of a method proposed by Hanson and Wernert (1997) in an attempt to facilitate 3-D navigation using 2-D controllers. Their technique divides the navigation space into two components: The constraint space defines the range of positional values that can be reached by the input device, and the camera model field (CMF) describes a set of ideal viewing parameters for each point in the constraint space. As the viewer controls motion through the constraint surface, the system determines the corresponding values in the CMF and presents them to the viewer.

Practically speaking, the CMF is not explicitly represented, as the number of points that constitute the constraint surface can be arbitrarily large. Hanson and Wernert (1997) proposed sampling control points of the CMF at a fixed resolution and then interpolating the viewing parameters using a bicubic Catmull-Rom spline. This method has the desirable property that the interpolated values will include the control points and ensures smooth, continuous transitions between the viewpoints (Dunlop, 2002).

This framework effectively describes a recommendation engine capable of guiding the viewer to ideal viewpoints based on the context of their exploration. Figure 1 shows a simple 2-D environment map with ideal gaze-direction vectors overlaid. This example shows two points of interest – the lower left and the right center – that will be recommended if the viewer is nearby. It should also be noted that this example shows only 2-D vectors, effectively adjusting the yaw of the viewpoint. Attentive navigation readily extends to manipulate any number of the viewing parameters.

The construction and use of these ideal viewing parameters can lead to the varying levels of support. For example, one design question asks,

“when should the system initiate its suggestions?” It may be the case that the viewpoint should remain aligned with the direction of motion, and the system automatically adjusts the viewpoint only when the viewer stops moving. Alternatively, if the system can actively adjust the viewpoint while the viewer is moving, perhaps it is desirable to let the velocity modulate how much input the system makes. A rich taxonomy of some of these design decisions can be found in Wernert and Hanson (1999). The goal of the present paper is to explore shared control of the viewpoint as a proof of concept, not to provide an exhaustive investigation of variations. Therefore, the experiments presented in this paper adopted two major approaches.

The first technique is called the *attentive camera*. It is characterized by the system continuously aligning the viewpoint with the ideal gaze vector as the viewer moves through the environment. Using the gaze-directed steering metaphor, the viewer controls two degrees of freedom: heading and forward motion. When the user initiates a movement forward, a motion vector is established. All movement occurs along that vector until the user stops moving. While the user is moving, the system uses the ideal viewing parameters from the CMF to adjust the orientation of the user's gaze. This decouples the direction of gaze from the direction of motion

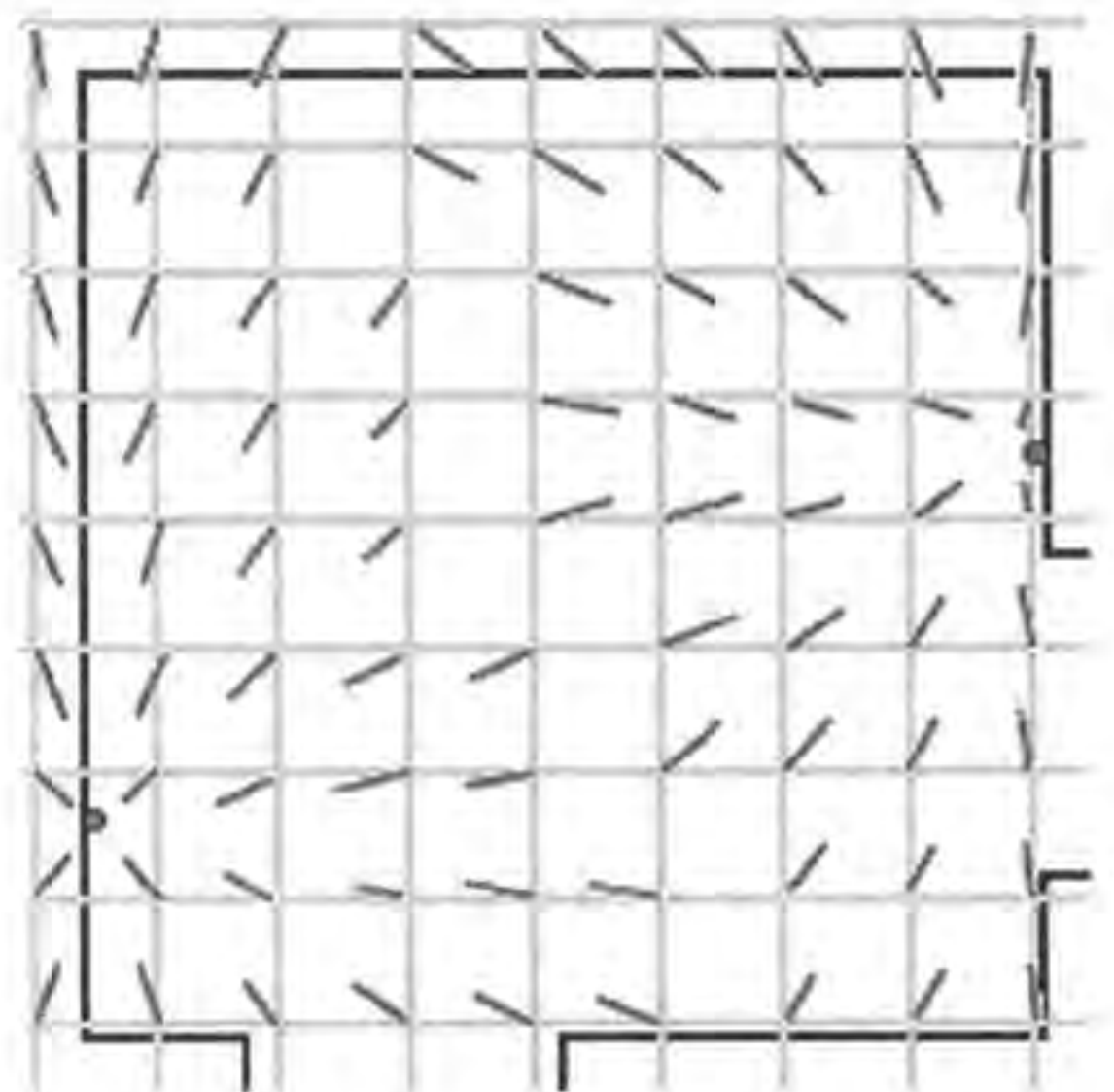


Figure 1. Two-dimensional ideal gaze vectors.



and allows the user to fixate on significant objects rather than the direction of travel. Although the gaze is manipulated to fixate on certain objects, the viewer can override the suggestions by simply stopping forward motion and readjusting the orientation of the viewpoint to look elsewhere.

A brief pilot study with this approach motivated two additional restrictions: (a) The gaze redirection should be disabled during the initial steps, and (b) the gaze should not be redirected to exceed the original peripheral vision limits, in order to provide good optical flow. Figure 2 demonstrates this approach. Motion starts at Point A. From A to B, the gaze remains aligned with the motion vector to allow the user to establish a sense of the motion direction. From B to C, the gaze is smoothly redirected from the motion vector to the ideal gaze vector, fixating on an object of interest. From C to D, the intended gaze vector exceeds the initial field of view, so the attention is gradually shifted back to the motion vector.

The second technique is known as the *attentive flashlight*. In this case, the ideal viewing parameters are externalized to compute the direction in which to shine a spotlight. This is similar to the effect that was proposed by Zhai et al. (1997) to expedite training with complex graphical user interface panels. As the viewer walks through the environment, the flashlight fixates on objects of interest. When the viewer stops moving, the viewpoint is pivoted until it is aligned with the ideal gaze vector. This approach allows the viewers to watch where they are going but gives them an external cue to where they should be looking or moving. A brief pilot study motivated one modification: If the object of interest is behind the viewer, the spotlight remains

in the peripheral edge of the image. This ensures that a portion of the flashlight is visible at all times and alerts the viewer to relevant information that is currently outside the field of view. Figure 3 shows a snapshot of the environment using the attentive flashlight.

## OVERVIEW OF EXPERIMENTS

A collection of experiments was designed to assess the effectiveness of attentive navigation versus free navigation techniques. These experiments present a range of comparisons that reflect various objectives and interests in the topic of assisted viewpoint control. The initial experiment contrasted the effects of using a mouse with automated viewing parameters with the robust control afforded by the use of a six-degrees-of-freedom (6DOF) controller. Acknowledging the potential confound introduced by an unfamiliar input device, we directed subsequent studies at isolating the effects of automated viewing parameters on spatial awareness and search tasks. Specifically, the second experiment studied the effect of the attentive camera on survey knowledge, and the third experiment examined how attentive techniques can aid both finite and exhaustive searches. Despite their differences, the three experiments belong together because they are complementary in providing a more complete picture of the effects and potential uses of assisted viewpoint control.

### EXPERIMENT 1

The first experiment was formulated as an attempt to show the viability of using a mouse with the attentive camera as a 3-D navigation tool. Previous studies have suggested that a mouse

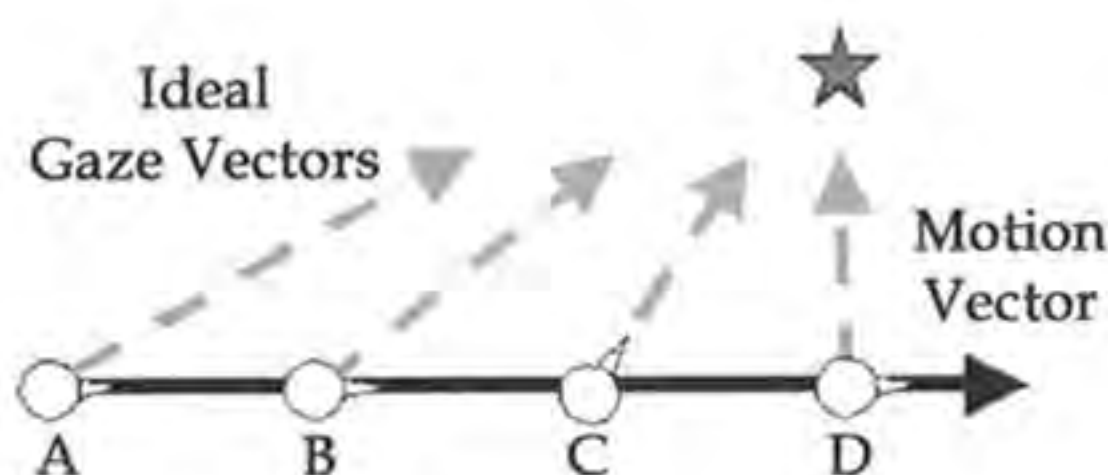


Figure 2. Attentive camera.



Figure 3. Attentive flashlight.



is undesirable for 3-D navigation, as compared with special 6DOF input devices such as the space ball or floating mouse. Although it is possible to simulate 6DOF with a mouse, to do so requires an artificial mapping, resulting in an extra cognitive burden on the user and slower performance (Beaton, DeHoff, Weiman, & Hildebrandt, 1987). It is our hypothesis that the attentive camera can effectively constrain the viewpoint to meaningful orientations and yet still allow operators the freedom to explore on their own terms.

## Method

**Participants.** Twenty undergraduate students were randomly assigned to navigate with either the attentive camera or free navigation. Each was paid \$10 for his or her involvement in this study.

**Stimuli.** All of the experiments in this study were conducted using a 17-inch (43-cm) color display rendering environments generated in C++ using OpenGL libraries (Silicon Graphics, Inc., Mountain View, CA).

For this experiment, the environment consisted of a large square room divided into even quadrants. The four exterior walls were shaded light green and tiled with a brick-like pattern. The four interior walls were light gray and had the same brick pattern. The floor and the ceiling both had an  $8 \times 8$  checkerboard pattern of black and gray squares. All eight walls extended from the floor to the ceiling. The environment also contained a variable configuration of objects. Each object was one of three shapes (sphere, pyramid, or a four-sided box) and one of nine colors (red, blue, green, yellow, violet, cyan, orange, gray, and light green). Between seven and nine objects were added to the room but ensuring that no color-shape pair was repeated. Figure 4 displays a sample map of the environment to give a sense of density of objects with respect to the size of the environment.

The environment was designed to prevent the user from being able to see the contents of an entire quadrant from a single viewpoint. To be confident about the contents of a quadrant, the participant needed to make use of all directions of motion. This was accomplished by the existence of three types of objects. An object was classified as *elevated* if it was positioned such

that it was viewable only if the viewpoint was adjusted vertically, either through position (increased  $y$ ) or orientation (increased pitch). An object was *occluded* if it was contained within another object (e.g., a four-sided box). Unless the viewpoint was properly positioned, the viewer would see only the container object. For example, if an open face was on top of a four-sided box, an interior object could be viewed only if the position was elevated (increased  $y$ ) and the viewer was looking down (decreased pitch). Finally, a *plain* object was defined as any object that was not elevated or occluded. Most of the objects that were encountered were plain, with elevated and occluded objects accounting for two to four objects per environment.

**Apparatus.** Participants exploring with the attentive camera used a standard mouse as the input device, adopting the gaze-directed walking metaphor (Bowman et al., 1997). Movement was registered by displacement from the initial starting position while the mouse button was depressed. The magnitude of the displacement was translated into a velocity in the VE. Moving the mouse forward or backward resulted in motion in the environment, whereas right or left motion caused the user to pivot clockwise or counterclockwise in place. Users were restricted from moving and pivoting simultaneously. Vertical position and pitch and yaw orientation were influenced by ideal vectors stored in the CMF. The

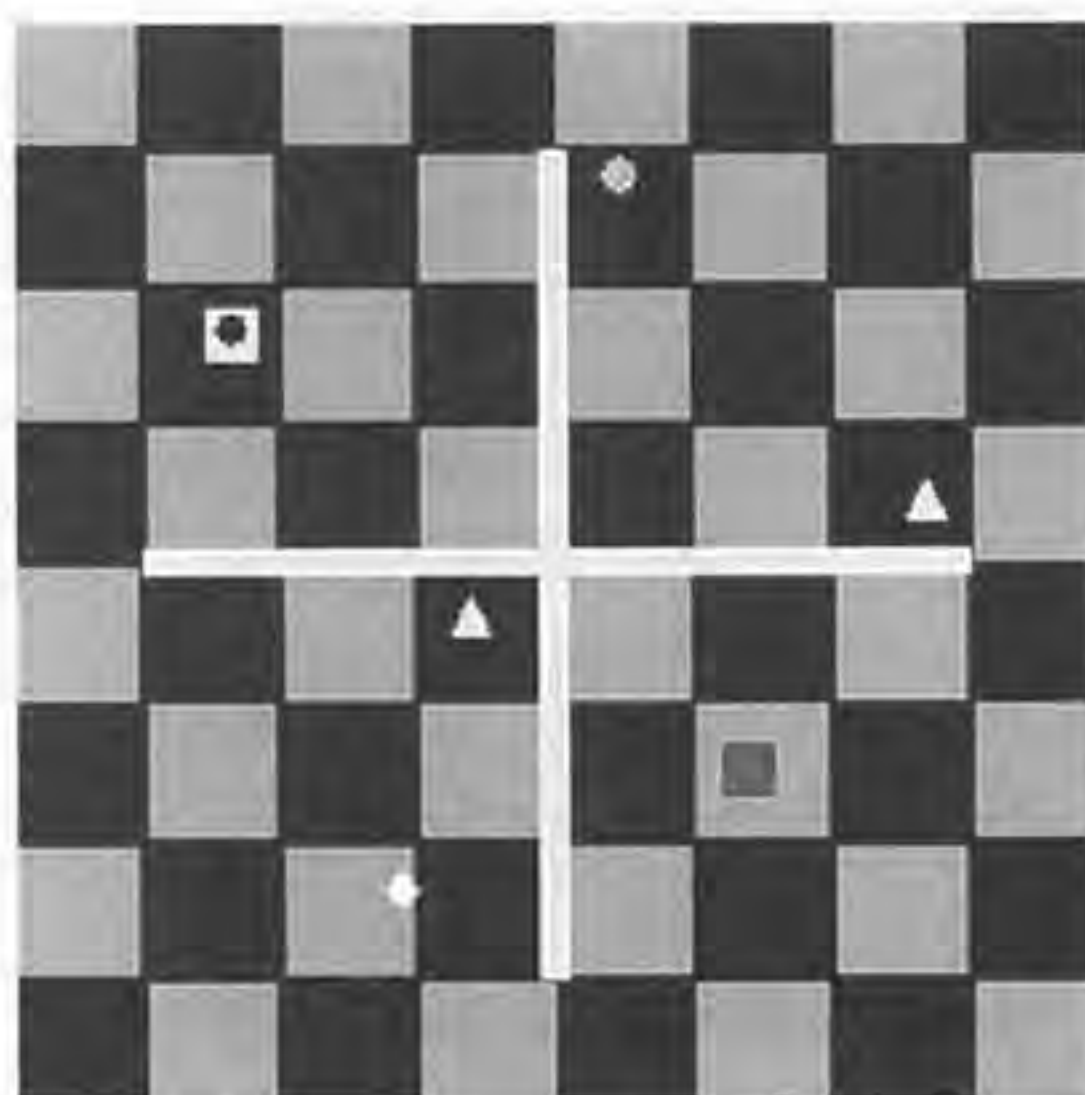


Figure 4. Sample Environment 1.



CMF was constructed so that the attentive camera would fixate on the object closest to the viewer. This strategy afforded the opportunity for the viewer's attention to be redirected to each object in the environment.

Free navigation adopted a standard flying metaphor and used a space ball to register input. Participants were allowed the following control: move forward/backward, move left/right, elevate/descend, rotate gaze to the left/right, and rotate gaze up/down. The velocity was a function of the amount of pressure applied to the space ball; more force exerted by the participant resulted in faster movement. Roll transformations were not included for either condition.

*Procedures.* The participants were given a verbal description of the technique and instruction on how to use the input device to perform certain actions. Additionally, to give a clear mental image of the technique's operations, the experimenter then physically demonstrated how each control movement corresponded to movement and orientation of a person's viewpoint.

Before testing, participants were placed in a sample environment and directed to familiarize themselves with the control of the navigation technique. They were instructed to train until they indicated that they felt completely comfortable with the operation of the technique. Participants were required to complete a simple set of motions to verify that they had obtained a baseline of control before proceeding. This verifica-

tion called for the participant to demonstrate motion in all directions followed by travel back to a specific object near the starting position.

After the training was completed, the participants were asked to explore a series of four environments with the task of locating and remembering as many objects as possible. No instruction about the nature or location of the objects was provided to the participants. The exploration session started when the participant pressed the space bar and lasted for 4 min. When the time expired, the screen automatically went blank. Immediately following the exploration, participants were given a list of 15 objects (color-shape pairs). They were asked to identify which of these objects they recalled seeing during their explorations by checking "yes" or "no" next to each entry on the list. The lists included all the objects that were present in the environment, with the balance assigned randomly from the remaining color-shape pairs. To assist with the recognition, participants could also scroll through a display that rendered each object on the list, providing a visual representation of the color-shape pair.

## Results

Data were analyzed for the likelihood that an object was overlooked or forgotten for each of the three types of objects: plain, elevated, and occluded. Figure 5 shows that there was no difference between attentive and free navigation for

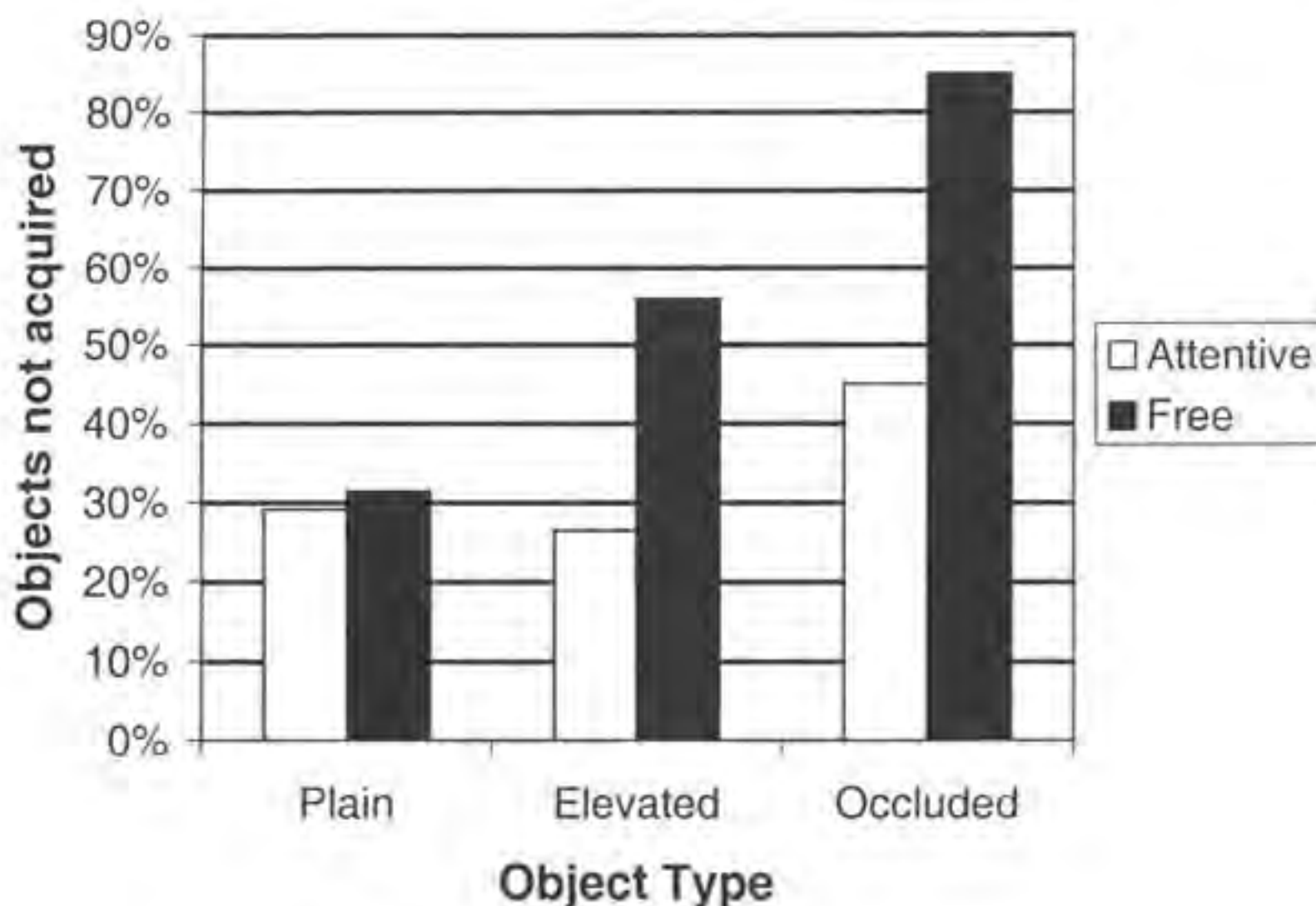


Figure 5. Errors in object identification.



plain objects. However, a single-factor analysis of variance (ANOVA) showed that elevated and occluded objects were significantly more likely to be missed using free navigation,  $F(1, 18) = 16.87$  and  $F(1, 18) = 20.53$ ,  $p < .01$ , respectively.

In order for the viewer to perform a correct identification, an object must have been in the field of view and the viewer must have noticed and remembered it. A breakdown of the data into these components provided additional insight to the advantages of the attentive camera.

The display accuracy is the ratio of objects in the room versus objects that were displayed on the screen. Objects could have been overlooked if the viewpoint did not include objects of interest in the field of view. With free navigation, this burden rested with the viewer, whereas the attentive techniques assumed this responsibility. Accordingly, the attentive techniques provided a significantly higher display accuracy for elevated objects (+21%),  $F(1, 18) = 7.05$ ,  $p < .02$ , and occluded objects (+35%),  $F(1, 18) = 10.36$ ,  $p < .01$ .

The attentive techniques did an enormous service by getting the objects in the field of view. In some cases this was sufficient. Given that an elevated object was on the screen, there was no difference between attentive and free navigation,  $F(1, 18) = 1.51$ ,  $p = .27$ , in the ability to recall it. However, for occluded objects, the attentive camera provided additional conspicuity; 82% of occluded objects that were displayed were remembered, versus only 35% with free navigation,  $F(1, 18) = 10.18$ ,  $p < .01$ .

These results confirm the initial hypothesis. The attentive camera limited the navigation to relevant viewpoints, automatically redirecting the gaze to targets while discouraging fruitless searches. It was notable that users of free navigation rarely made use of all the degrees of freedom provided to them; the vertical components of control were often neglected. Generally, users would occasionally remember to search for elevated objects, but they often discontinued this practice after lack of success. On the contrary, attentive camera users could focus on moving the viewpoint through the environment and relied on the automation to indicate when special targets were nearby.

We acknowledge that this experiment potentially confounds the benefits of assisted view-

point control with physical operation of the controller. The superior performance of the attentive camera could come from either the automatic gaze redirection or a lack of familiarity with the use of the space ball. Regardless of this confound, these results show that the attentive camera is an effective control technique that allows users to operate in desktop virtual environments with low-degree-of-freedom controllers.

## EXPERIMENT 2

The second experiment was designed to assess the impact of the attentive camera on survey knowledge – the understanding of the configuration of landmarks in the environment. Although this kind of spatial knowledge is facilitated by active, self-controlled navigation, distractions in the environment can be a significant impediment. We propose that acquisition of survey knowledge can be hastened by using the attentive camera to focus on key landmarks.

### Method

*Participants.* Twenty participants, unique to this study, were randomly assigned to navigate with either the attentive camera or free navigation. Each was paid \$10 at the conclusion of the study.

*Stimuli.* The ability to gain survey knowledge depends somewhat on the topology of the environment. For this experiment, a simple space was constructed that adhered to the following principles:

- A survey of the entire room could not be obtained from a single vantage point. Two large pillars separated the rectangular room into a figure-eight-shaped collection of corridors.
- A checkerboard floor pattern was in place to assist the user in estimating distance traveled. This feature also guaranteed that a visual flow would always be present whenever the user was mobile.
- Different-colored walls allowed the user to maintain orientation in a symmetric room.

Each environment contained 18 objects: three shapes (sphere, cube, and cone) by six colors (red, yellow, orange, green, blue, and violet). Unlike the previous experiment, here all 18 combinations were present in every environment. Furthermore, all objects could be considered plain; there were no containers, and all objects



rested directly on the floor. A sample map is shown in Figure 6.

*Apparatus.* Because all of the objects were situated on the ground, there was no need for the operator to adjust the vertical components of the viewpoint. Therefore, both groups could navigate using a standard mouse, implementing the gaze-directed walking metaphor described in the previous experiment.

The CMF for the attentive camera was generated in order to manipulate the yaw orientation of the operator's gaze based on proximity to target objects. For this experiment, a subset of the objects in the room was designated as targets of gaze redirection (attended objects). The remainder of objects did not influence navigation (unattended objects). In the first trial, ideal viewing information was provided to focus on the six spheres. In the second trial, the viewpoint was manipulated to fixate in the direction of nearby cubes. Participants were given no a priori information about what the attentive camera would highlight. Although there was nothing inherently important about the attended objects, this approach can simulate an exploration in which a naive viewer lacks the knowledge to discriminate between relevant and irrelevant features.

*Procedures.* As in the previous experiment, participants were given a verbal description, a physical demonstration, and an opportunity to practice using the assigned technique. Participants were exposed to two timed trials in which

they were asked to explore the environment with the goal of being able to reconstruct a map. To compensate for the larger environment and to allow adequate time for initial survey knowledge to develop, each trial lasted 15 min. Upon completion of the task, the screen went blank, and participants were given an electronic map framework (boundary lines with the floor pattern) similar to that shown in Figure 6. Participants were asked to drag and drop the 18 objects from the bottom of the screen to their corresponding positions on the map.

## Results

The reconstructed maps were analyzed for object misplacement. For each object, the Euclidean distance between the actual location and the reported location was calculated. The unit of measure used for this analysis is based on the unit distance in the software package that was used. However, for clarity of this discussion, these measurements (shown in Figure 7) have been normalized with the floor tiles.

There were no differences in the overall error measure between participants using the attentive camera technique and those using the free navigation technique. Participants were able to position an average of approximately eight objects within two tiles of the actual location. However, further analysis revealed a systematic difference in the accuracy with which objects were positioned. Although there is no apparent

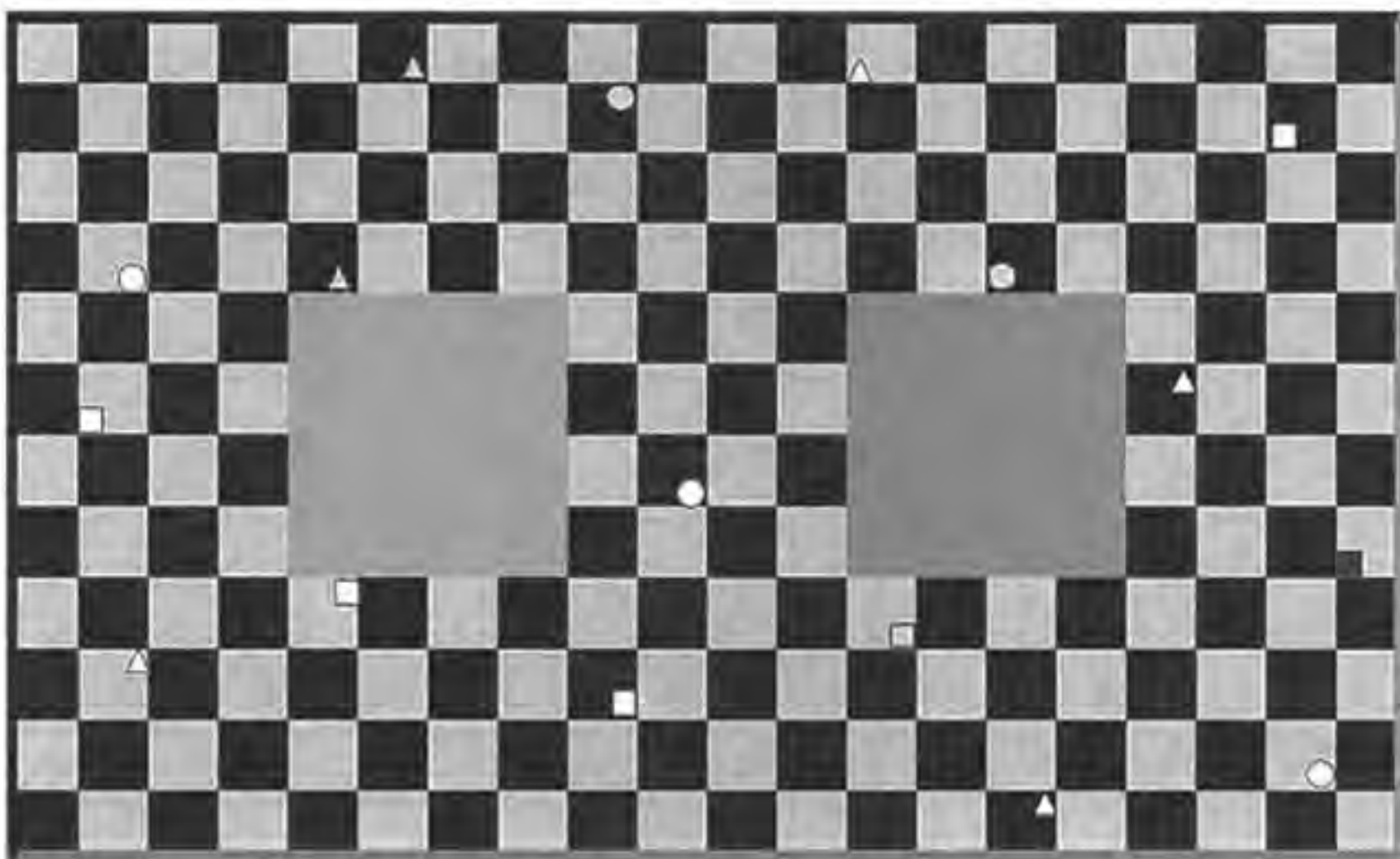


Figure 6. Sample Environment 2.



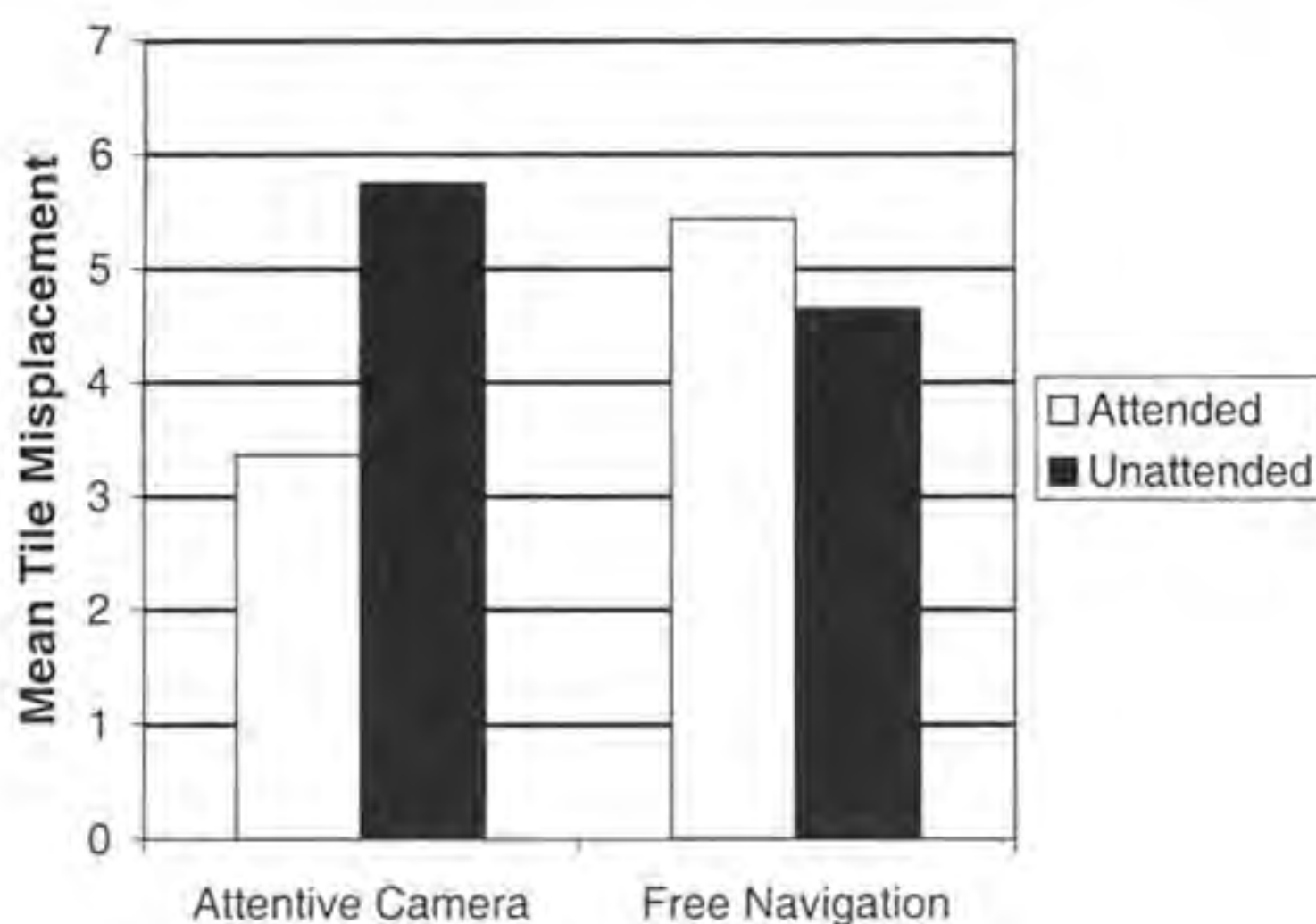


Figure 7. Object displacement.

pattern to the objects that were well positioned by the free navigators, the objects correctly positioned by viewers navigating with the attentive camera correspond to the objects targeted by the CME. Figure 7 indicates that users of the attentive camera technique made significantly fewer placement errors on attended objects than on unattended objects,  $F(1, 18) = 5.24, p < .04$ . With no distinction between attended and unattended objects presented to the free navigation group during exploration, it is not surprising that no difference is observed in map placement,  $F(1, 18) = 1.58, p = .26$ . Users of the attentive camera also proved significantly better at placing the attended objects when compared with the free navigation users,  $F(1, 18) = 6.39, p < .05$ . These results suggest that subtly redirecting attention to a subset of important features can bias the order in which survey knowledge is acquired.

It is also worth noting that a partial ceiling effect was observed in the success of the attentive camera. Of the 10 participants using the attentive camera technique, 2 actually placed the unattended objects as well as the attended objects, with a total average error of just over one tile per object. Had the environment been more complex, it is expected that the attended score would have stayed low, whereas the unattended score would have increased.

### EXPERIMENT 3

In an environment where fine details need to

be processed, it is critical that viewers can quickly home in on appropriate information. The third experiment was designed to assess the ability of the attentive techniques to filter out extraneous details and focus on attributes that are relevant to a given task. By using the attentive techniques in an environment that calls for judgments to be made based on discriminating features, we predicted that more accurate and thorough searches would be performed.

### Method

*Participants.* Twenty-four users, unique to this study, were randomly assigned to one of three navigation techniques: attentive flashlight (9), attentive camera (8), and free navigation (7). Three participants were excluded from the study. One participant terminated participation prior to completing the full experiment. A data collection malfunction caused the data files for 2 additional participants to be overwritten, necessitating their exclusion. Participants were each paid \$15 at the conclusion of the experiment.

*Stimuli.* The nature of this experiment required individuals to evaluate details of objects contained in the virtual environment. Because the simple geometric shapes used in the previous experiments were not sufficient for this task, viewers were asked to navigate through a virtual art gallery. The gallery was composed of 40 paintings evenly distributed over a three-room environment (see Figure 8). For all trials, the



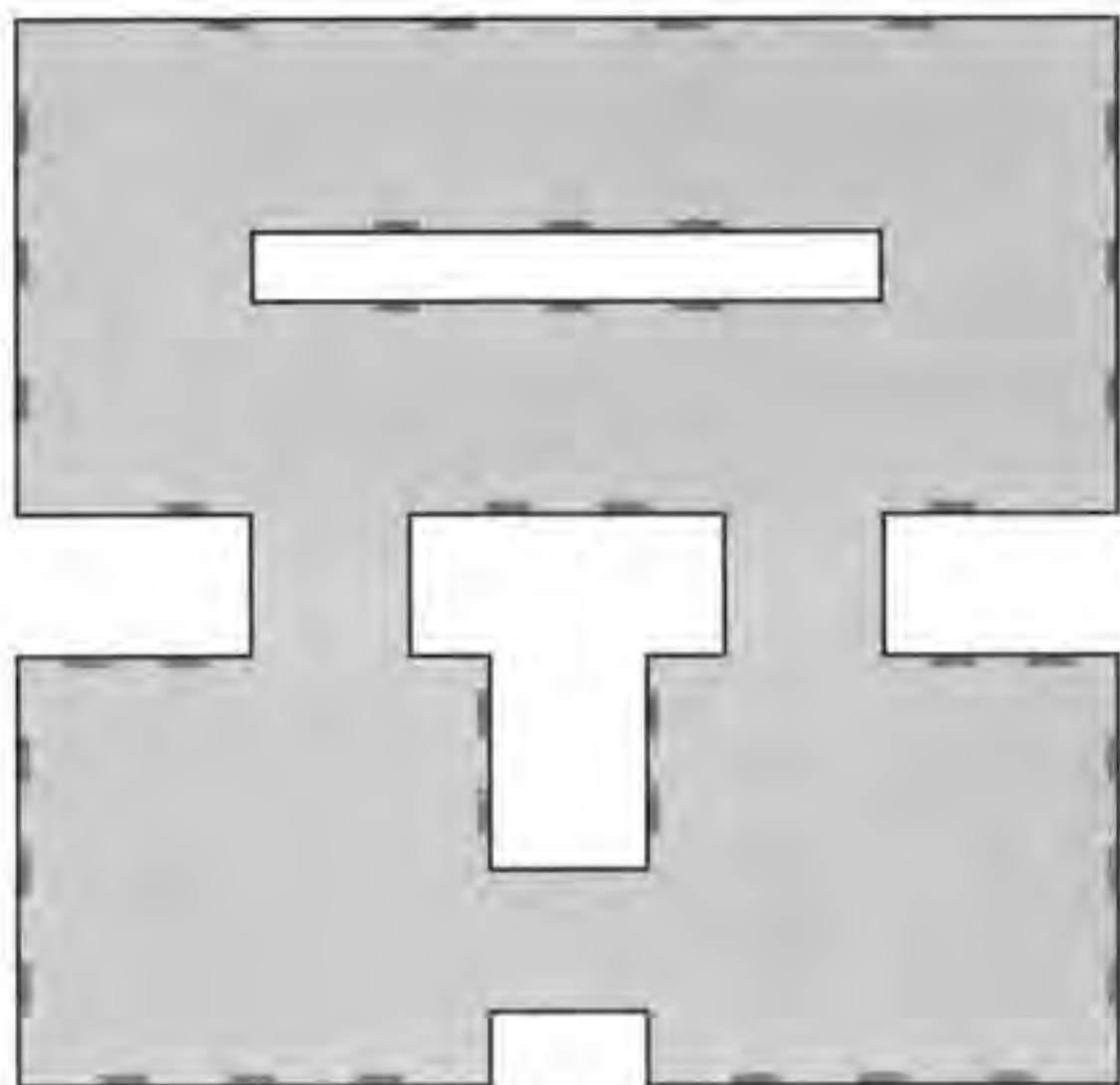


Figure 8. Sample Environment 3.

floor plan of the environment was the same, but the artwork was changed for each task. The participants were advised that it was similar to visiting the same museum but that a different exhibit was on display. Below each of the paintings, a label indicated the title of the work, the artist, and the year it was painted.

**Apparatus.** All participants navigated using a standard mouse, implementing the gaze-directed walking metaphor described in Experiment 1. The CMF was used to provide guidance for the yaw orientation in the direction of a subset of target paintings consistent with the task given to the operator. As described previously, the CMF was used to redirect the gaze in the case of the attentive camera, whereas the attentive flashlight used the same CMF to establish the direction for a projective texture, simulating a spotlight.

**Procedures.** As with the previous two experiments, descriptions, instructions, and training were provided before the commencement of the experiment. The controls for the attentive flashlight were identical to those for the other two techniques.

After training, participants were exposed to four separate search tasks in a virtual art gallery. The participants were instructed to complete each search task as quickly and accurately as possible. When the participant was confident that the search was complete, the space bar was pressed, ending the trial.

Three of the tasks required the participants to find multiple paintings that shared a common characteristic. They were given a brief description of the characteristic, followed by the goals: "What are the titles of the works by *Cézanne* that are displayed?" "What are the titles of the *Portraits* that are displayed?" and "What are the titles of the *Abstract Expressionist* works that are displayed?"

Because the viewers did not know in advance how many paintings they were looking for, these tasks required an exhaustive search of the space to complete. The fourth task involved a finite search; the viewer was asked to locate a single painting based on the title: "What animal is featured in the work 'Three Worlds'?"

During each of the four explorations, the following data were collected: The participants recorded answers to the search questions on paper. A log file was generated that time stamped all the actions made by the viewer as well as all supplemental actions taken by the attentive techniques. The CMF was created to direct attention to support these search tasks. For the exhaustive searches, the attentive techniques once again fixated on nearby targets. In the case of the finite search, the suggestions offered by the attentive techniques aligned with the ideal route until the actual painting was nearby. This prevented the system from suggesting that the viewer stare at an obstructing wall when the actual picture on the other side.

## Results

The analysis of the results is divided into the two task groupings defined previously: exhaustive searches and finite searches. For the exhaustive searches there were no differences in the duration of the search; however, the accuracy of the search showed significant differences. Figure 9 displays the two types of errors that were observed: either targets could be missed (omissions) or nontargets could be incorrectly identified (false alarms).

An ANOVA showed a main overall effect on the number of omissions,  $F(2, 69) = 6.15, p < .01$ . Paired comparisons using the Bonferroni technique showed that both attentive techniques were effective at reducing the number of omissions: attentive camera,  $p < .01$ , and attentive flashlight,  $p < .013$ . As observed in the first



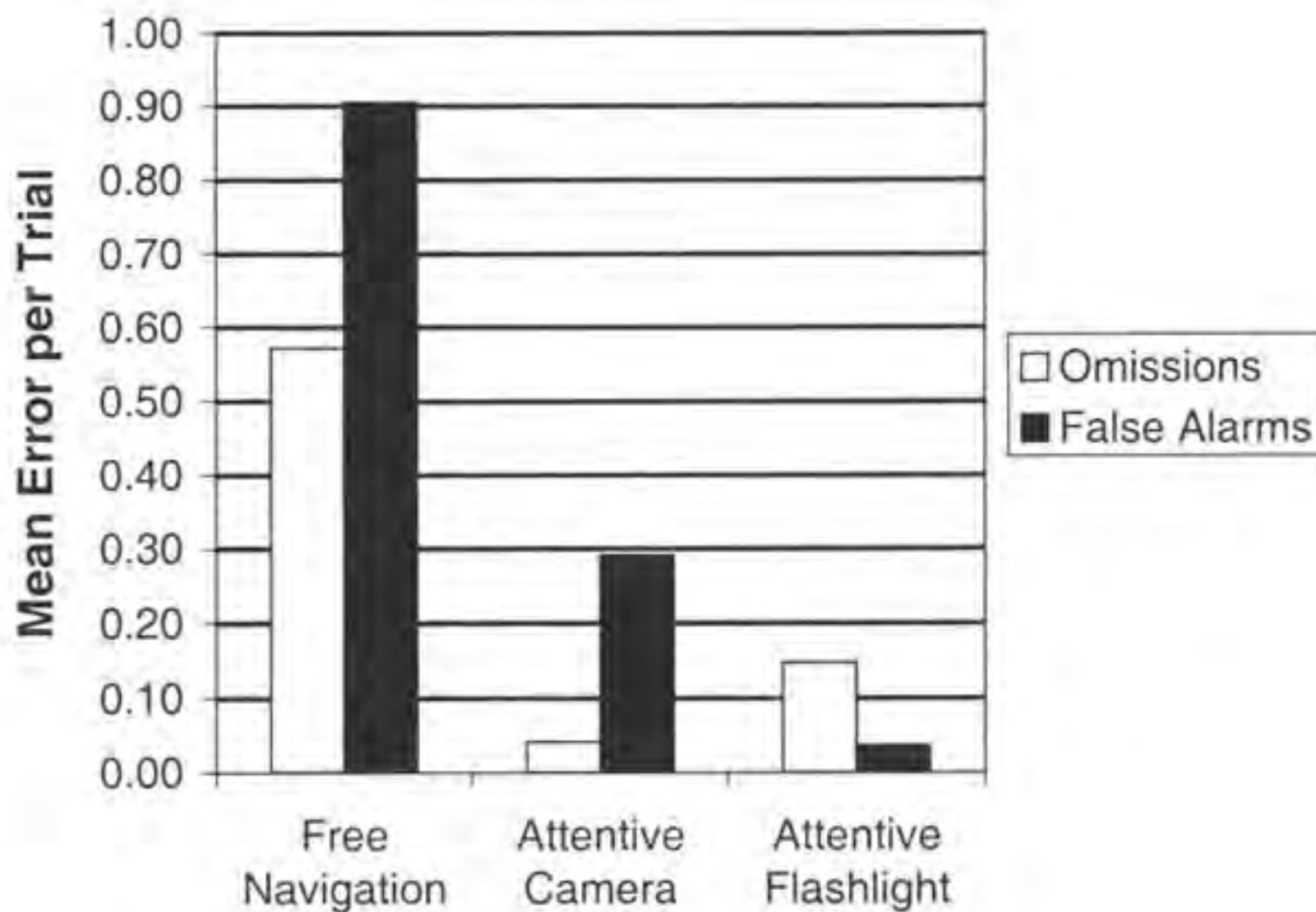


Figure 9. Errors by condition.

experiment, the suggestions offered by the attentive techniques effectively draw the targets into the field of view, causing more attention to be devoted to the target. Similarly, there was an overall main effect on the number of false alarms,  $F(2, 69) = 4.80, p < .02$ , specifically attributed to the difference between free navigation and the attentive flashlight ( $p < .01$ ). This result can probably be attributed to the visual reinforcement provided by the flashlight. If the flashlight did not fixate on the object, it was clearly not part of the target set. However, the gaze fixation from the attentive camera was subtler and thus more ambiguous.

Not only did the attentive techniques prove more accurate, they were also easier to control. Every time the viewer wanted to change the method of viewpoint motion between moving and pivoting, he or she was required to release and repress the mouse button. Therefore, the number of mouse clicks approximates the effort that the user had to exert to control the technique. Both attentive techniques resulted in significantly fewer mouse clicks as compared with the no-assistance condition,  $F(2, 69) = 11.17, p < .01$ . Post hoc analysis indicated that the attentive flashlight users made even fewer clicks than did the attentive camera users,  $p < .01$ .

Moreover, without assistance, viewers had to explicitly look around to find the target objects, averaging 800 commands. Alternatively, the at-

tentive techniques automatically oriented the viewer in the appropriate direction. The success of the flashlight technique can be seen by a significant reduction in user-initiated pivot motion,  $F(2, 69) = 12.32, p < .01$ , from either the no-assistance,  $p < .01$ , or the attentive camera condition,  $p < .01$ . Interestingly, users of the attentive camera and the no-assistance technique did not differ significantly in the number of "looking motions." Figure 10 shows the relative number of user actions devoted to pivoting the viewpoint.

For finite searches, the viewer knew when the task objective had been met. Therefore, an exhaustive search of the space was not needed. In this case, it was simply a matter of finding the appropriate painting as quickly as possible. An ANOVA showed a main overall effect on the search time,  $F(2, 21) = 3.74, p < .05$  (Figure 11). Paired comparisons using the Bonferroni technique shows that users of the attentive flashlight were significantly faster than free navigation users,  $p < .016$ . The difference in search time between the attentive camera and free navigation conditions was not significant.

## DISCUSSION AND CONCLUSIONS

These experiments demonstrate an approach to exploring virtual environments that lies between the extremes of guided tours and unassisted free navigation, capitalizing on the strengths



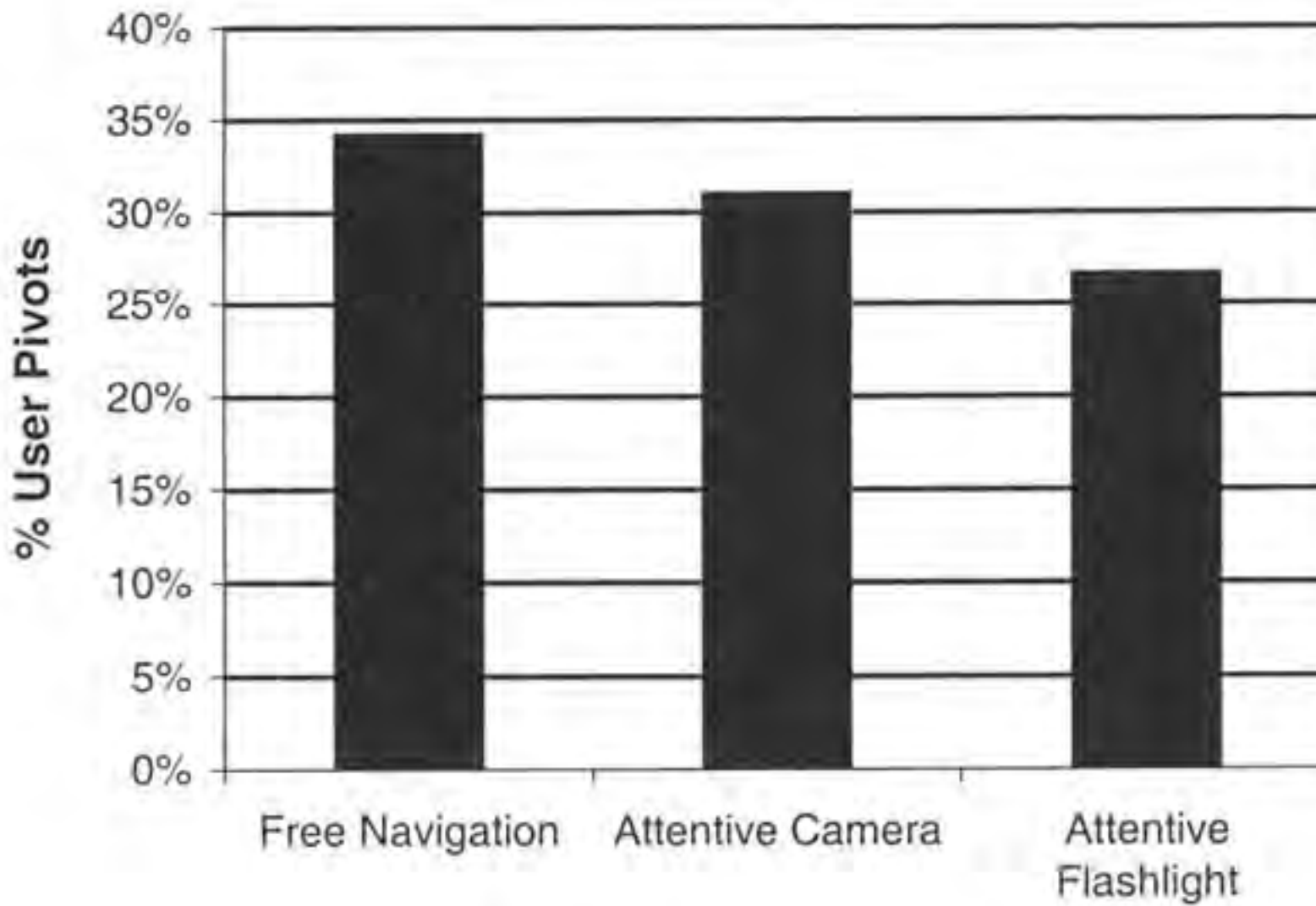


Figure 10. User-initiated looking.

of each. By sharing control of the viewpoint, the attentive techniques reduce the amount of interaction required to successfully extract the same amount of information. Therefore, the viewer is allowed to direct more attention on what is being observed during an exploration and less on how to manipulate the controls. At the same time, the viewer can still engage the display according to his or her needs. Unlike fully guided tours, this technique allows the viewer to make decisions about the sequencing and observation

time. Thus the viewer retains a sense of self-determination, and the designer can ensure that certain information will be displayed.

At a very basic level, the attentive camera, using automatic gaze redirection, increases the probability that key viewpoints will be utilized. This is supported by the inordinate number of missed objects and omissions registered by the free navigators throughout the experiments. Naive explorers in the first experiment may not have been aware that some objects would be

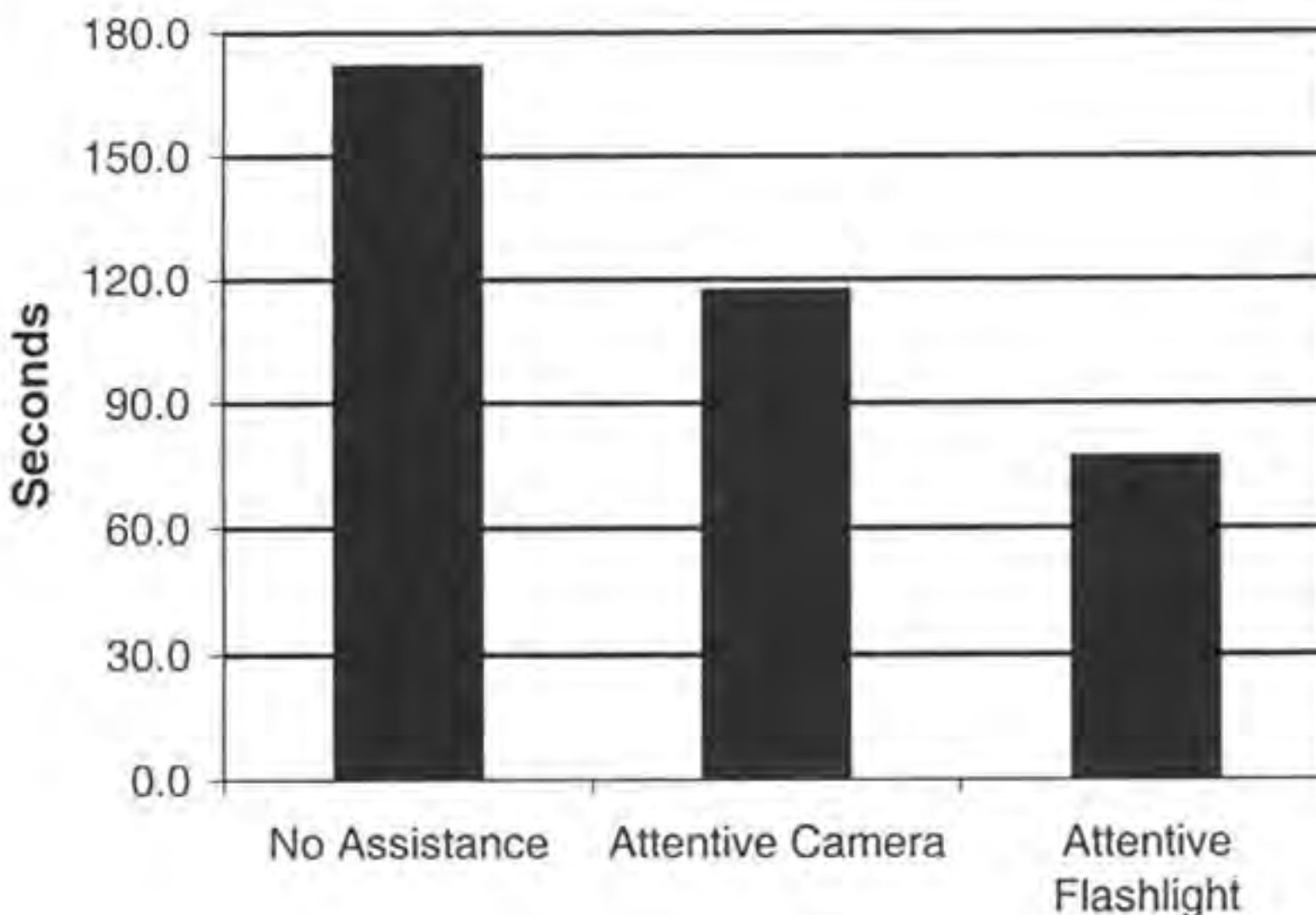


Figure 11. Finite search time.



elevated or contained inside other objects. Free navigators remained oblivious, whereas the attentive camera successfully drew attention to these “hidden” elements. These benefits were evident with relatively small environments and should be expected to scale to large and more complex environments with even more pronounced effects.

Attentive navigation helped viewers understand the configuration of key elements, as well as their presence. Development of survey knowledge is largely dependent on repeated exposure to the landmarks in the environment. Unfortunately, given the complexity of a realistic VE, an explorer may focus on irrelevant or redundant landmarks, prolonging the development of useful survey knowledge. The attentive camera technique is effective at focusing user attention on significant elements in the environment, maximizing their exposure and thus maximizing the user’s knowledge of their configuration.

Thorndyke and Hayes-Roth (1982) argued that motion greatly facilitates obtaining orientation-independent survey knowledge. One explanation they offered is that the navigator is being exposed to multiple perspectives. The attentive camera takes this exposure a step further. By fixating the gaze on an object while moving past it, the viewer sees not only multiple facets of the object but an extended context as well. In effect, the object remains stationary while the environmental background rotates around the object, maximizing the number of perspectives for the viewer. Because gaze fixation is built into the CMF, the viewer does not need to take any overt actions to take advantage of this optimized information.

The attentive techniques also foster the establishment of connections between elements of the scene, another key factor to learning survey knowledge. When using gaze-directed steering, viewers establish the intent to move by looking at an object they wish to explore. The same is true for this implementation of the attentive techniques. As the viewer moves toward his or her objective, unlike assisted navigation, the attentive techniques may shift the viewpoint to focus on another object. This redirection may forge a connection between the intended destination and the attended object: “As I walked toward the far wall, I looked to my left and saw

X”. Connections such as this offer insight to the overall configuration of the environment.

The attentive flashlight technique outperformed the other techniques with regard to explicit information seeking. Two major factors contribute to this overall success. Primarily, guidance offered by the attentive flashlight is continuous, whereas the attentive camera has some gaps in its feedback. Recall that the gaze redirection is temporarily disabled when the viewer starts moving and again when the ideal gaze is beyond the peripheral vision. Although these constraints are necessary to make the attentive camera viable, they ultimately impair the overall performance for this type of task. In the same vein, the viewer is able to see the recommendation of the attentive flashlight explicitly at all times, whereas the attentive camera offers advice implicitly and only while the viewer is in motion.

The success of the attentive flashlight may also be explained by the fact that the recommendation is external to the viewer, giving a greater sense of control. Users of the attentive flashlight were able to effectively ignore the system’s suggestions until they stopped moving. In contrast, the attentive camera began influencing the gaze shortly after motion was initiated, regardless of the viewer’s intentions. The viewers’ conflict with the automatic gaze redirection may have been more pronounced in the third experiment because of the more explicit goals of their task. Anecdotally, some users of the attentive camera seemed to struggle for control of the gaze, frequently stopping to manually realign their gaze with the direction of motion. It is likely that the measure of user-initiated pivots suffered because of this. A more in-depth analysis is needed to discriminate whether the intention to turn was to try to find something or to maintain a desired heading.

These experiments have shown that attentive navigation was an effective method for sharing viewpoint control. This approach can lead to development of targeted survey knowledge, better recognition of objects, and improved search times. However, it should be noted that the two implementations provide different levels of assistance. We do not wish to suggest that one of these approaches is superior for all scenarios or all viewers. It is possible that some tasks are



best suited to the extremes of guided tours or free navigation. Viewers may even benefit from initially being exposed to a guided tour and gradually assuming more control as they learn more about the environment or domain that they are exploring. When creating a VE, designers should attempt to match the tasks and user experience to the appropriate level of control – no longer an all-or-nothing decision.

### ACKNOWLEDGMENTS

This research was supported by an Office of Naval Research Grant N-00014-96-1-1222 and by Air Force Office of Scientific Research Contract Number F49640-01-1-0542.

### REFERENCES

- Arthur, E., Hancock, P., & Chrysler, S. (1995). Spatial orientation in real and virtual worlds. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 328–332). Santa Monica, CA: Human Factors and Ergonomics Society.
- Arthur, E., Hancock, P., & Telke, S. (1996). Navigation in virtual environments. In *Proceedings of SPIE-International Society for Optical Engineering* (Vol. 2704, pp. 77–85). Bellingham, WA: SPIE.
- Bares, W. H., Thainimit, S., & McDermott, S. (2000). A model for constraint-based camera planning. In *Smart graphics: Papers From the 2000 AAAI Symposium* (pp. 84–91). Menlo Park, CA: AAAI Press.
- Beaton, R., DeHoff, R., Weiman, N., & Hildebrandt, P. (1987). An evaluation of input devices for 3-D computer display workstations. In *Proceedings of SPIE-The International Society for Optical Engineering* (Vol. 761, pp. 94–101). Bellingham, WA: SPIE.
- Beckhaus, S., Ritter, F., & Strothotte, T. (2001). Guided exploration with dynamic potential fields: The CubicalPath system. *Computer Graphics Forum*, 20, 201–210.
- Bowman, D., Koller, D., & Hodges, L. (1997). Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Virtual Reality Annual International Symposium* (pp. 45–52). Los Alamitos, CA: IEEE Computer Society Press.
- Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann.
- Drucker, S. M., & Zeltzer, D. (1994). Intelligent camera control in a virtual environment. In *Proceedings of Graphics Interface '94* (pp. 190–199). Banff, Canada: Canadian Information Processing Society.
- Dunlop, R. (2002). *Introduction to Catmull-Rom splines*. Retrieved June 24, 2005, from <http://www.mvps.org/directx/articles/catmull/>
- Goerger, S., Darken, R., Boyd, M., Gagnon, T., Liles, S., Sullivan, J., et al. (1998). Spatial knowledge acquisition from maps and virtual environments in complex architectural spaces. In *Proceedings of the 16th Applied Behavioral Sciences Symposium* (pp. 6–10). Colorado Springs, CO: U.S. Air Force Academy.
- Halper, N., & Oliver, P. (2000). CAMPLAN: A camera planning agent. In *Smart graphics: Papers From the 2000 AAAI Spring Symposium* (pp. 92–100). Menlo Park, CA: AAAI Press.
- Hanson, A., & Wernert, E. (1997). Constrained 3D navigation with 2D controllers. In *Visualization '97* (pp. 175–182). Los Alamitos, CA: IEEE Computer Society Press.
- He, L. W., Cohen, M. F., & Salesin, D. H. (1996). The virtual cinematographer: A paradigm for automatic real-time camera control and directing. In *SIGGRAPH 96* (pp. 217–224). New York: Association for Computing Machinery Press.
- Hix, D., Swan, E., Gabbard, J., McGee, M., Durbin, L., & King, T. (1999). User-centered design and evaluation of a real-time battlefield visualization virtual environment. In *IEEE Virtual Reality '99* (pp. 96–105). Washington, DC: IEEE Computer Society Press.
- Keahey, T. A. (1998). The generalized detail-in-context problem. In *IEEE Symposium on Information Visualization* (pp. 44–51). Washington, DC: IEEE Computer Society Press.
- Peruch, P., Vercher, J., & Guthier, G. (1995). Acquisition of spatial knowledge through visual exploration of simulated environments. *Ecological Psychology*, 7, 1–20.
- Sheridan, T. B. (1992). *Telerobotics, automation and human supervisory control*. Cambridge, MA: MIT Press.
- Siegel, A. W., & White, S. (1975). The development of spatial representations in large-scale environments. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 9–55). New York: Academic Press.
- Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14, 560–589.
- Ware, C., & Osborne, S. (1990). Exploration and virtual camera control in three dimensional environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (pp. 175–185). New York: Association for Computing Machinery Press.
- Wernert, E., & Hanson, A. (1999). A framework for assisted exploration with collaboration. In *Visualization '99* (pp. 241–248). Los Alamitos, CA: IEEE Computer Society Press.
- Zhai, S., Wright, J., Selker, T., & Keln, S. (1997). Graphical means of directing users' attention in the visual interface. In *Interact '97* (pp. 59–66). London: Chapman & Hall.

Stephen Hughes is an assistant professor in the Department of Mathematics, Computer Science and Physics at Roanoke College. He received his Ph.D. in information science from the University of Pittsburgh in 2005.

Michael Lewis is an associate professor in the Department of Information Science and Telecommunications in the School of Information Sciences at the University of Pittsburgh. He received a Ph.D. in engineering psychology from Georgia Institute of Technology in 1986.

Date received: May 19, 2005

Date accepted: September 2, 2004