

# The Impact of Three Interfaces for 360-Degree Video on Spatial Cognition

**Wutthigrai Boonsuk**  
Iowa State University  
Ames, Iowa, USA  
wutthigraib@gmail.com

**Stephen B. Gilbert**  
Iowa State University  
Ames, Iowa, USA  
gilbert@iastate.edu

**Jonathan W. Kelly**  
Iowa State University  
Ames, Iowa, USA  
jonkelly@iastate.edu

## ABSTRACT

In this paper, we describe an experiment designed to evaluate the effectiveness of three interfaces for surveillance or remote control using live 360-degree video feeds from a person or vehicle in the field. Video feeds are simulated using a game engine. While locating targets within a 3D terrain using a 2D 360-degree interface, participants indicated perceived egocentric directions to targets and later placed targets on an overhead view of the terrain. Interfaces were compared based on target finding and map placement performance. Results suggest 1) non-seamless interfaces with visual boundaries facilitate spatial understanding, 2) correct perception of self-to-object relationships is not correlated with understanding object-to-object relationships within the environment, and 3) increased video game experience corresponds with better spatial understanding of an environment observed in 360-degrees. This work can assist researchers of panoramic video systems in evaluating the optimal interface for observation and teleoperation of remote systems.

## Author Keywords

360-degree view; panorama; virtual navigation; spatial cognition.

## ACM Classification Keywords

H.5.2 User Interfaces - Evaluation/methodology, Graphical user interfaces (GUI), User-centered design;

## General Terms

Evaluation, Interface Design, Navigation, Spatial cognition

## INTRODUCTION

A surveillance system typically involves monitoring multiple video feeds of cameras in an extensive area. We can divide this system into two categories: stationary systems and mobile systems. In this paper, we focus on the mobile system in which cameras are attached to dynamic objects such as persons, vehicles, or aircrafts to provide a

remote 360-degree view. The video feeds from the mobile system are usually monitored in real-time and require significant vigilance to examine their contents. The design of the view interface can be essential in providing the effectiveness of observation. Traditional interfaces for remote camera systems (e.g., Google Street View [25] or HMDs connected to 360-degree environments [7] or immersive systems with displays surrounding the user [5, 12, 19, 24] require multiple views or active remote panning of the view to observe occluded regions. An interface that provides observers with a complete view at a single glance without leading to perceptual distortion of spatial information would represent an improvement.

Human eyes have a horizontal field of view (FOV) of approximately 200 degrees, and so displaying a 360-degree panorama to a human observer requires compressing the display horizontally. The resulting horizontal distortion could disrupt the viewer's ability to accurately perceive spatial relationships between multiple objects in the camera's view. Egocentric (self-to-object) directions of objects in the display will not necessarily correspond to the egocentric directions of the objects relative to the person wearing the camera. Human spatial orientation is thought to be largely based on the egocentric directions and distances to known landmarks [8, 26]. Misperception of those egocentric directions could result in large errors in determining one's position within a remembered space. In light of the potential disruption of normal spatial cognitive processes, the interface should augment the view to leverage our natural sense of presence and spatial awareness. In this paper, we describe an experiment investigating the effectiveness of three designs for a 360-degree view interface, part of an ongoing project to build a wireless 360-degree life-sharing system.

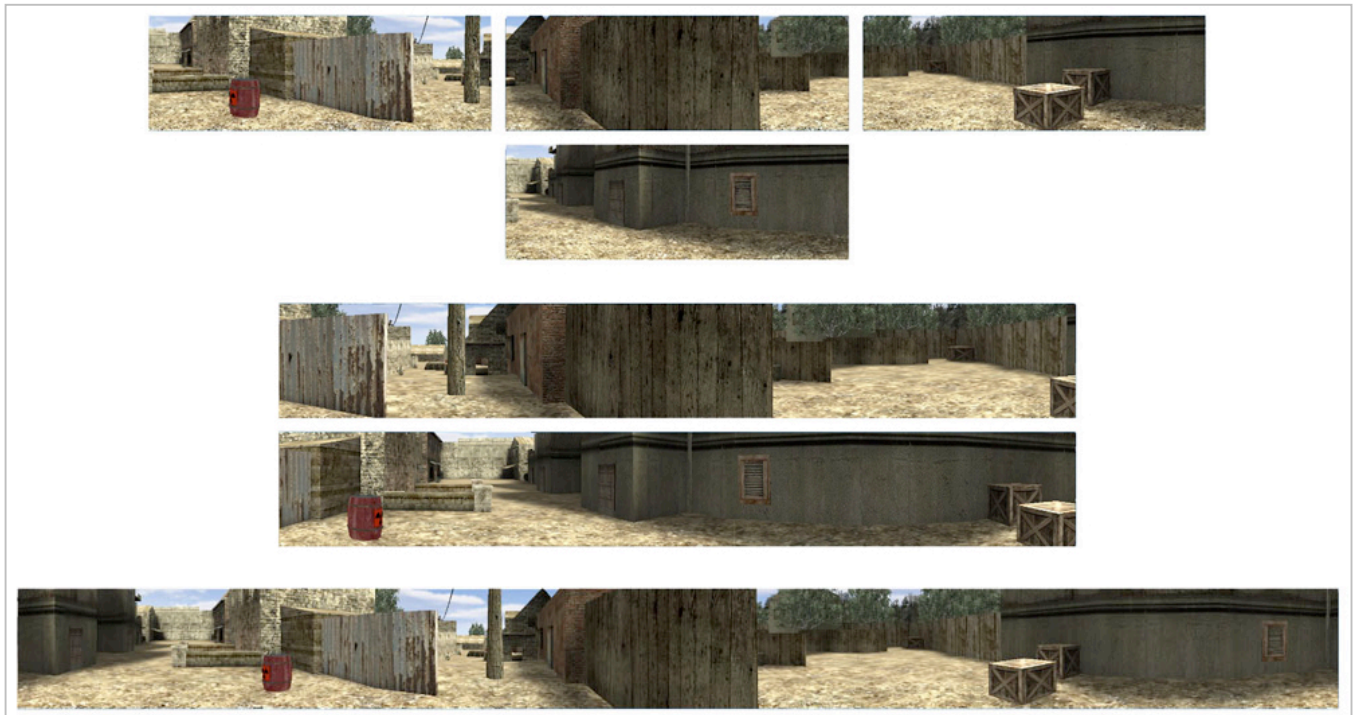
## Acquiring 360-Degree Video

The two approaches to producing 360-degree video are combining images from multiple cameras, each with limited field of view (FOV), and using a single panoramic camera. The first approach requires use of software techniques such as image blending [4], piecewise image stitching [9, 22], or 2D projective transformation [17, 20, 23]. The single camera approach may use a fish-eye lens [28], omnidirectional camera [14], or a conic mirror camera [3]. The current study simulates video feeds within a virtual game

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.



**Figure 1. Three 360-degree interfaces from the same viewpoint: (a) 90-degree x 4, with left, front, right, rear; (b) 180-degree x 2, with front and rear and (c) 360-degree x 1, panorama. Original 3D model created by HORSE-A-F; used with permission.**

engine environment by utilizing multiple cameras situated around the head of a virtual character. This approach is similar to piecewise image stitching, and the choice of multiple cameras enabled tight control of distortion, which is reduced by using multiple cameras with small FOV (25-30 degrees). The interfaces were designed to maintain the same size of objects in the 3D scene across all three interfaces.

#### **Previous 2D Displays of 360-Degrees**

The primary reason to transform a 360-degree view to a 2D flat interface is to offer observers the ability to use standard computer displays and to enable monitoring of multiple video feeds simultaneously (e.g., from multiple people or vehicles using 360-degree camera rigs). The 360-degree interface should help observers attend to both center and peripheral information and maintain spatial orientation within the displayed environments.

Past research has used a variety of interfaces to display a 360-degree video feed on a single monitor. Three examples highlight the diversity of interface designs. Kadous et al [13], for a robot search and rescue system, displayed a main front view with smaller additional views (left, right, and rear) arranged around the border of the main view. Meguro et al [15], for a mobile surveillance system for an autonomous vehicle, presented two greens, each with 180-degree FOV (front and rear). Greenhill and Venkatesh [11], for a mobile surveillance system for city buses, presented a single uncut 360-degree view. While a variety of interface arrangements exist, the authors are not aware of

a thorough investigation of the characteristics of an effective 360-degree display.

#### **Potential Impact on Spatial Cognitive Tasks**

A remote 360-degree surveillance system allows the surveillance operator to detect objects of interest in any direction, thereby exceeding the abilities of the camera-wearer. For example, a surveillance operator could detect a navigational goal or target behind the wearer and direct him/her to approach the goal. Alternatively, the operator could remember the goal location for later use. In both of these tasks, performance might be impaired by displaying 360-degree video on a 2D monitor because of disruptions in judging egocentric directions.

Accurate perception of the egocentric direction of a goal relative to the camera-wearer requires mapping the 360-degree image onto a representation of the wearer's body. When displaying a 360-degree view in a single, uninterrupted image on a monitor, an object appearing on the far left or right of the image is actually directly behind the camera-wearer. This mapping between display and body coordinates is unnatural and will likely cause errors in estimating egocentric object locations. Adding visual borders to the image might improve judgments of egocentric direction by providing multiple points of reference between the image and the wearer's body. For example, segmenting the display into four 90-degree sections separated by borders indicating front-left, front-right, back-left and back-right in body coordinates would

facilitate mapping between image and body coordinates. This hypothesis is tested in the current experiment.

### **Direction Performance Tied to Mapping Performance?**

Errors in judging egocentric directions to objects may or may not lead to errors in generating a cognitive map of the environment, and different spatial cognitive theories make different predictions regarding this relationship. For the purposes of this study, a cognitive map is considered to be a mental representation of object-to-object, or exocentric, locations in the environment. A cognitive map is required when retrieving remembered locations, such as when drawing a map of the environment or navigating to a remote goal. If accurate perception of egocentric direction is necessary in order to construct an accurate cognitive map, then there should be a direct relationship between egocentric and exocentric errors. That is, misperception of a target's egocentric direction should lead to errors in placing that object within a cognitive map [8, 26]. This would result in a positive correlation between egocentric pointing errors made within the environment and map placement errors made when reconstructing the layout from memory.

However, if exocentric relationships are stored independently of egocentric directions, then errors in egocentric and exocentric judgments should be unrelated. If cognitive maps are constructed on the basis of object-to-object relationship (e.g., the barrel is on the north side of the shed), then an accurate cognitive map could be formed even if egocentric directions to individual objects are misperceived. This would result in no correlation between egocentric pointing errors made within the environment and map placement errors made when reconstructing the layout from memory. These alternative hypotheses are also tested in the current experiment.

### **RESEARCH APPROACH**

The goal of this project is to identify which of three design interfaces offers the best spatial task performance in an observed remote environment. For this study, active navigation is used instead of passive observation since active navigation benefits peripheral perception in a large FOV display [18]. Participants performed several spatial tasks using each interface, and their performance was recorded and used to determine the effectiveness of the interfaces. The best interface should provide appropriate utilization of the edges of the display, including the corners, as is commonly used in first-person shooter games to display status information, and should promote accurate judgment of egocentric object directions as well as accurate creation of a cognitive map of the virtual environment.

### **Interfaces**

The three interface configurations chosen for comparison were based on the previous work described above and on an informal pilot study of these and additional interfaces. The

interfaces are shown in Figure 1 and the accompanying video figure.

The first interface, "90-degree x 4," is a combination of four views: front, left, right, and rear. As illustrated in Figure 1a, each view has 90 degree FOV and is placed 10 pixels apart from the other. The rear view is placed underneath the front, left, and right views. This first interface is designed based on the common size of FOV for a video game with additional views to cover 360 degrees. The borders of the four views provide visual landmarks indicating the camera boundaries at the front-left, front-right, back-left, and back-right relative to the user's heading within the virtual environment.

The second interface, "180-degree x 2," comprises two views with 180 degree FOV (Figure 1b). We designed this interface based approximately on the natural horizontal FOV of human eyes. The left and right borders of the images provide visual landmarks indicating the camera boundaries directly left and right relative to the user's heading within the virtual environment.

The last interface, "360-degree x 1," is a single 360-degree panoramic view as illustrated in figure 1c. We believe this interface may reduce visuospatial working memory load since the views are grouped into a single element. While this interface has been perceived by some to be the obvious "winner" because of its familiarity, the results below suggest that a careful analysis is merited. The left and right borders of the image both provide visual landmarks indicating the camera boundary directly behind the user's heading within the virtual environment

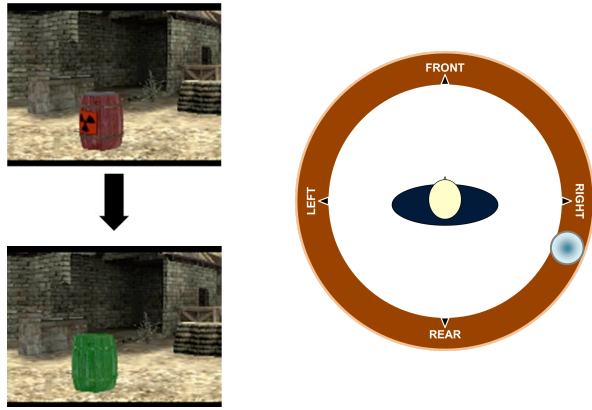
Each interface was displayed on a 22-inch monitor with the participant sitting approximately one foot away, yielding a visual angle on the eye of ~30-40 degrees horizontally and ~10 degrees vertically.

### **METHODS**

The virtual environment used in this study was created using the graphics game engine Irrlicht [1] with C++ and OpenGL. Our experiment utilized a 22-inch 3M multi-touch display to present the views from virtual cameras of the game engine and receive participants' responses in given tasks. Participants used the keyboard arrow keys to navigate through the virtual environment.

### **Tasks**

For each interface, the experiment was broken up into three phases, a familiarization phase, a target search phase, and a map reproduction phase. Interface order was counterbalanced to prevent order effects from contaminating the analyses. The familiarization phase occurred at the beginning of each interface session. During familiarization, participants were given 5 minutes to familiarize themselves with the navigation controls and the interface. Familiarization took place within a 3D model of an urban environment; the subsequent target search phase



**Figure 2, left:** Participants tapped 10 red barrels in the environment to "clear" them to green. **At right:** Compass rose, showing front, right, left, and rear, where a participant has tapped to indicate that a target barrel sits about 100° to the right of her. Original 3D model at left created by HORSE-A-F; used with permission.

took place within a different 3D model of a small hilly village (measuring 360 x 360 feet). During the target search phase, participants had 10 minutes to locate 10 targets (red wooden barrels). Participants were instructed to select the target as soon as it appeared anywhere on the display. To select a target, the participant tapped it with a finger, and the target color changed from red to green (Figure 2, left). To accommodate colorblindness, red barrels also lost their radioactivity logo when touched. Immediately after selecting the barrel target, participants were asked to identify the target direction relative to their current heading in the virtual environment by tapping a compass rose (Figure 2, right), herein referred to as the pointing response.

The map phase began after participants had located all 10 targets or after 10 minutes had elapsed. Participants were asked to place 10 targets on an overhead map (Figure 3), herein referred to as the map response. Barrels were not uniquely identifiable except by their spatial location. After placing all 10 targets on the map, participants were introduced to the next interface and the entire process was repeated.

Three different target layouts were created and randomly paired with the three interfaces so that each participant experienced all three layouts. The primary dependent measures were errors committed on pointing and the map responses. We also recorded the total number of targets found during the target search phase, the total time required to find all targets during the target search phase, and the egocentric directions and distances of selected targets during the target search phase.

#### Questionnaire

At the end of the three interface sessions, a questionnaire was administered to assess participants' satisfaction with



**Figure 3:** Overhead map of terrain, on which participants located the target barrels that they had selected in first-person view. Original 3D model created by HORSE-A-F; used with permission.

each interface, and also to identify whether participants reported using the display periphery on each interface.

#### Participants

A total of 20 participants (4 females and 16 males), ages 18 - 35, were recruited. To gauge their ability to understand a video-game-like virtual environment and navigate using the keyboard control, participants were asked for number of hours of video game playing per week. The median number of hours was one; most participants did not routinely play video games.

#### RESULTS

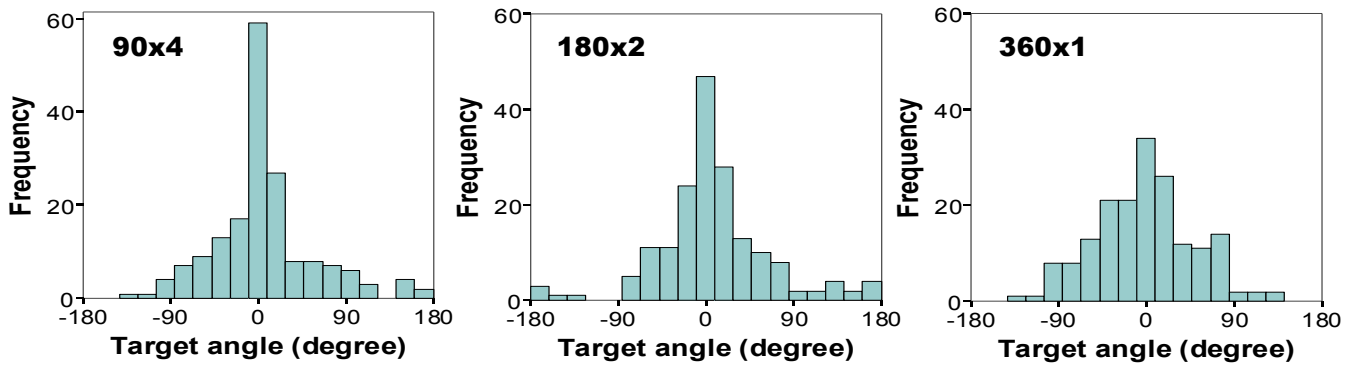
Data from two of the 20 participants were excluded from all analyses because their pointing errors were extremely high (more than  $6\sigma$ ) and inconsistent with errors made by the remaining participants.

Across all three interface sessions, 13 of the 18 participants found all 10 targets. The remaining 5 participants sometimes found only 8 or 9 targets. Participants who found all 10 targets spent 5 minutes on average. In particular, average time for 90-degree x 4 was 4.8 minutes, 180-degree x 2 was 5.0 minutes, and 360-degree x 1 was 5.3 minutes. However, times could range from 2 to almost 10 minutes.

#### Target Location

The degree to which participants made use of the entire 360-degree display can be determined by analyzing the actual angles of target barrels at the time they were selected during the target search phase.





**Figure 4: Histograms of actual barrel target angle at the time of participant selection. More barrels were chosen in the periphery in the 360-degree x 1 interface.**

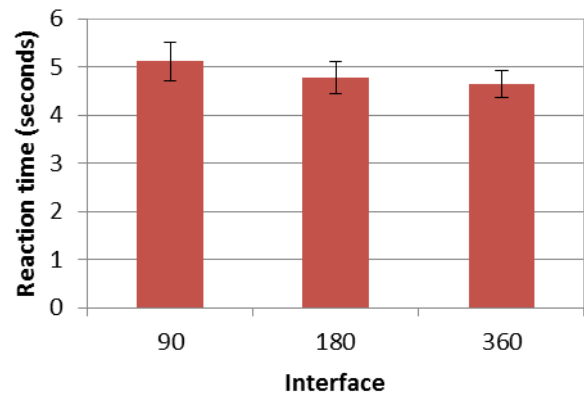
The frequency distribution of actual target directions at the time they were selected is shown in Figure 4. Directions displaced rightward from the participant's heading at the time of selection were arbitrarily labeled as positive angles, and leftward directions were labeled as negative angles, where 0 degrees is straight ahead and 180 degrees is directly behind. In all three interfaces, the peak frequency occurs at 0 degrees and decreases with increasing distance from 0 degrees. This indicates that participants tended to select targets when they were near the center of the display. Furthermore, the distributions of target directions vary across interface. When using the 90-degree x 4 interface, 52% of targets were selected when they fell in the center 45 degrees of the display. This value drops to 50% in the 180-degree x 2 interface and 41% in the 360-degree x 1 interface. Considering an even smaller window defined by the center 15 degrees of each interface accounts for 31% of targets in the 90-degree x 4 interface, 23% of targets in the 180-degree x 2 interface, and 15% of targets in the 360-degree x 1 interface. These differences in target direction distributions indicate that the interface influenced participants' focus of attention, such that attentional distribution across the display was broadest in the 360-degree x 1 interface and most narrow in the 90 x 4 interface (and intermediate in the 180 x 2 interface).

Target distance (the distance from the participant's point of view to the barrel at the time it was selected) affected the size of the target in pixels within the interface. The size of the target was expected to have little impact on pointing performance (angular pointing error or the time to choose the pointing angle on the compass rose), since the pointing action is based on the horizontal position of any green barrel pixels on the display. As predicted, target distance was not significantly correlated with pointing error,  $r(18) = .135$ . However, target distance was positively correlated with the time required to indicate the target's direction on the compass rose,  $r(18) = .570$ . This suggests that participants processed the scene three-dimensionally during the pointing task, noting a target's distance from themselves, or possibly associating the distant barrels with

distant landmarks. This spatial processing implies that, as desired, participants are engaged with a 3D environment rather than a 2D projected image.

#### Reaction time

Reaction time was evaluated using a log of participants' actions in the virtual environment. This log was an output text file that contained participants' directions and positions that were recorded every second during the experiment. The reaction time was analyzed by replaying this data and then measured the time delay between target appearance and target selection.



**Figure 5: Time delay between target appearance and target selection**

The differences of average reaction times were small across three interfaces (less than 1 second) as shown in Figure 5. Reaction time was analyzed in a one-way repeated-measures ANOVA to determine the effect of interface on reaction time. The main effect of interface was not significant indicating that reaction times did not differ across the interfaces,  $F(2,34) = 0.871$ ,  $p = 0.428$ .

### Pointing error

Absolute pointing errors were evaluated in order to assess the accuracy with which participants could map the 360 degree images onto egocentric target directions relative to the participant's heading in the virtual environment. Pointing errors were computed by finding the minimum absolute difference between the actual angle of the target and the angle indicated by the participant. Pointing errors (shown in Figure 6) otherwise were generally low with approximately 75% of pointing responses being within 20 degrees of the actual target direction. Absolute pointing error was analyzed in a one-way repeated-measures ANOVA to determine the effect of interface on pointing error. The main effect of interface was significant,  $F(2, 34)=5.54$ ,  $p = 0.008$ . Post-hoc analyses indicated that pointing errors in the 90-degree x 4 and 180-degree x 2 interfaces did not differ from one another ( $p=.809$ ), but pointing error when using the 360-degree x 1 interface was significantly greater than when using the 90-degree x 4 ( $p=.032$ ) and the 180-degree x 2 interface ( $p=.003$ ).

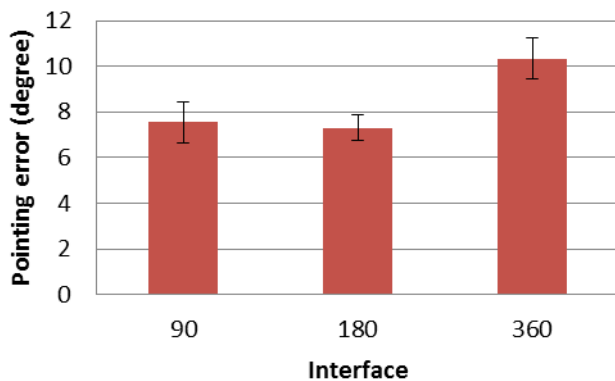


Figure 6: Pointing errors by interface

### Map error

Map error was calculated as the mean Euclidean distance between the participant's selected locations on the map task and the actual target locations. In many cases there was a clear correspondence between the barrel placed by the participant and the actual location of a barrel in the environment (Figure 7, top). In other cases, participants placed barrels in locations midway between two actual barrel locations (Figure 7, bottom). Because the barrels were non-unique, it was impossible to determine which barrel response corresponded to which actual barrel location. As such, we calculated the sum of squared distance errors for all possible pairings of barrel responses with actual barrel locations for each participant on each map. The barrel pairings which minimized the sum of squared errors (SSE) were used to calculate average barrel placement error (see Figure 7 for an example of barrel pairings).

The map task was challenging for some participants, and map error varied considerably across individuals. The virtual environment was 360 feet x 360 feet in dimension.

- Actual target
- Participants' placed targets

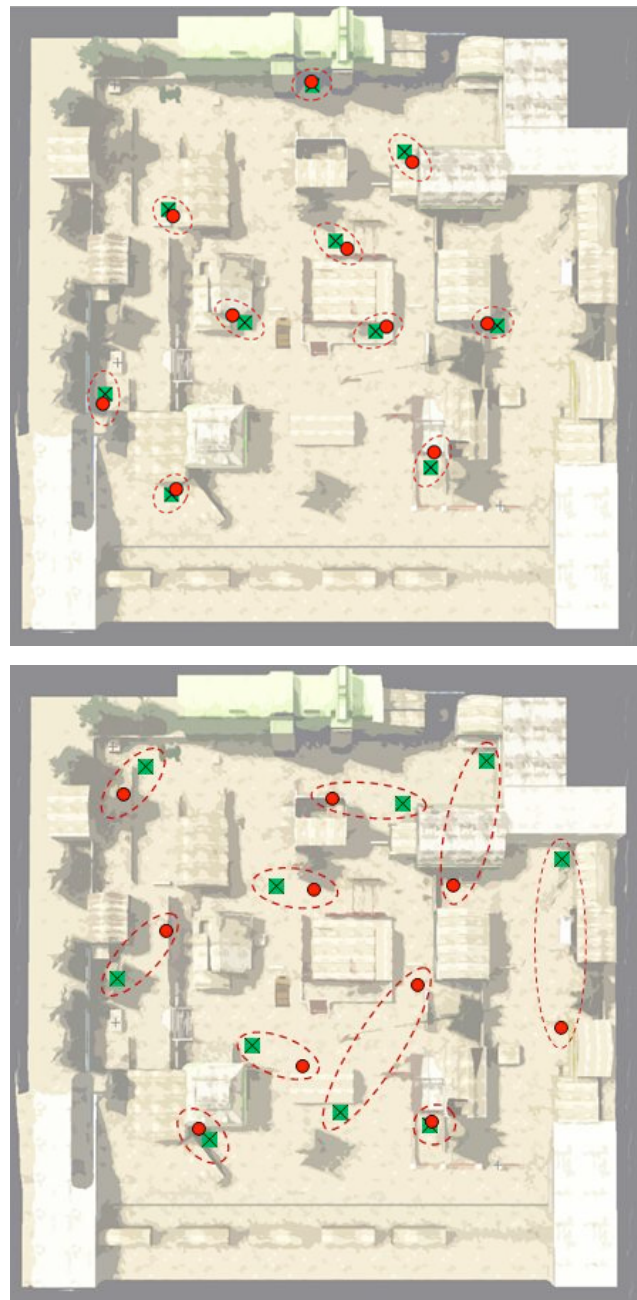
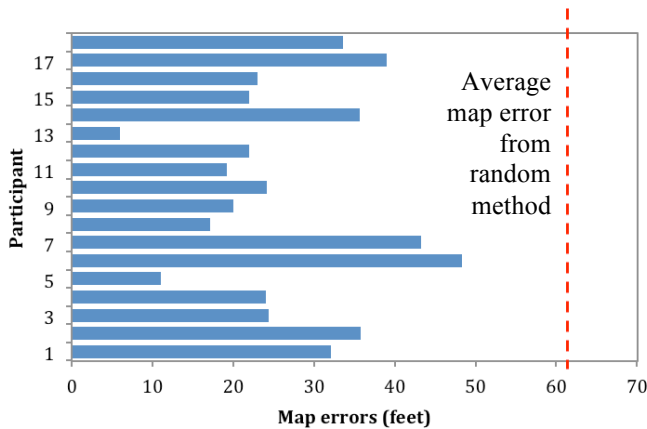


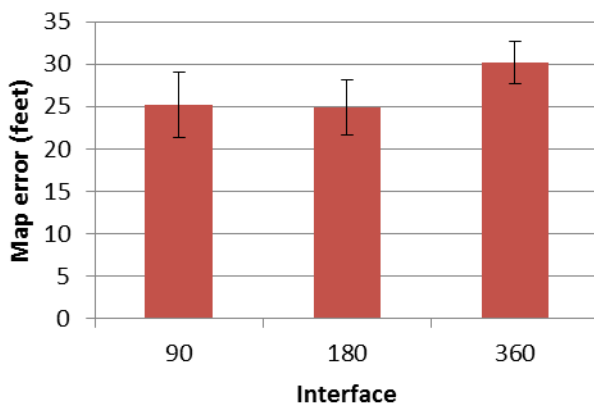
Figure 7: Two maps showing large differences in map error. The top map shows targets placed highly accurately, while the bottom map shows a more typical placement. Pairings show the map error scoring method, minimizing the sum of squared errors across all possible pairings. Original 3D model created by HORSE-A-F; used with permission.

Average map error for all participants ( $n=18$ ) was 26.7 feet, but average map error for individuals (averaged across their 10 barrel placements) ranged from 6 – 48 feet. However, a simulation of repeated random barrel placements yielded a



**Figure 8: Participants' map errors vs. random map error**

map error of greater than 60 feet (Figure 8). Thus, individuals varied but were always better than chance performance. Also, because participants varied in how far they travelled through the environment during the target search phase (even though the experience was capped at 10 minutes) it could be posited that a longer travel path would lead to more exposure to relevant landmarks and object-to-object relationships, which in term may lead to more familiarity with the map and lower map error. However, map error and travel distance were not significantly correlated,  $r(18) = -.172$ . Map error was analyzed in a one-way repeated-measures ANOVA to determine the effect of interface on map error (Figure 8). The main effect of interface was not significant, indicating that map errors did not differ across interfaces ( $F(2,34)=1.589$ ,  $p = .219$ ).



**Figure 9: Map errors by interface**

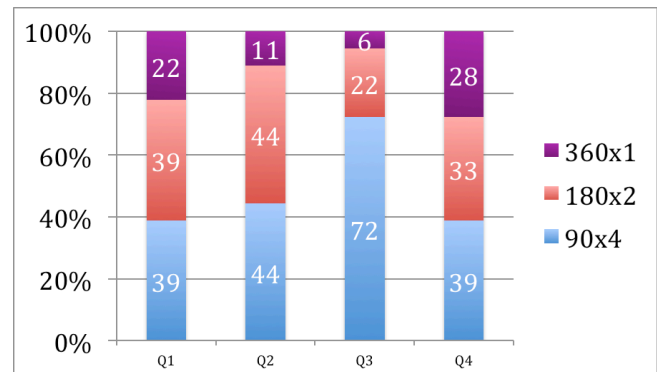
#### Correlation between pointing error and map error

If accurate perception of egocentric direction is necessary in order to construct an accurate cognitive map, then there should be a positive correlation between egocentric pointing errors and map errors. Spearman's rank correlations were calculated between absolute pointing

errors and map errors. The between-participant correlation was not significantly different from zero,  $r(18)=.164$ .

#### Survey Results

Survey results (Figure 10) serve as additional data for triangulation of participants' experiences using the interfaces. Answers to Question 1 (Which interface for 360-degree viewing did you prefer?) and Question 2 (Which interface for 360-degree viewing provided the most natural feel for navigating through the environment?) both indicate a minority of preference for the 360 degree x 1 interface (preferred by 22% and 11% of participants in Questions 1 and 2, respectively). The remaining responses are evenly split between the 90 degree x 4 interface and the 180 degree x 2 interface. Question 3 (Which interface for 360-degree viewing allowed you to determine the direction of objects in the scene accurately?) shows that a strong majority preferred the 90-degree x 4 interface (72%), and only a single participant (6%) preferred the 360-degree x 1 interface for target angle identification (pointing). Question 4 (Which interface for 360-degree viewing allowed you to place the barrels on the top-down map most accurately?) shows a somewhat even distribution with no clear preference of the best interface to help build a cognitive map.



**Figure 10: Survey results. Q1: Preferred interface. Q2: Most natural feel for navigation. Q3: Most accurate interface for pointing task. Q4: Most accurate interface for map task.**

When comparing these preferences with performance on the pointing and map tasks, Q1, Q2, and Q3 results match the pointing error performance data, in that participants performed significantly worse when using the 360-degree x 1 interface compared to the other two interfaces. Q3 highlights a slight difference between self-reported preferences and pointing errors: a majority of participants preferred 90-degree x 4 over 180-degree x 2 for accurate pointing, even though their pointing error performance did not significantly differ in the two interfaces. This result suggests that the 90-degree x 4 may be preferred for reasons other than the tasks themselves, such as familiarity of the 90-degree FOV from first-person shooter video games. Q4

results also match the map error data in that map errors did not differ by interface, and neither did preferences.

### Individual Differences

Previous work has shown that video game experience improves general spatial skills such as spatial allocation of attention and mental rotation [6]. We investigated whether individual differences such as video game experience influenced performance on the pointing and map tasks. Participants with greater or equal to 3 hours per week of video game experience were considered gamers, and those with fewer than 3 hours per week of video game experience were considered non-gamers. Video game players exhibited overall lower errors on the map task ( $F(1, 16) = 6.174, p = .024$ ), but the pattern of map errors across the three interfaces was no different for gamers and non-gamers (i.e., the interaction between video game play and interface was non-significant,  $F(2, 32) = 1.4, p = .261$ ). Video game play did not affect pointing performance ( $r(18) = .026$ ).

Although the primary analysis of map performance data revealed no differences between the three interfaces (see Map Error section, above), we also considered the possibility that there may be individual differences in the effect of the three interfaces on map error. To evaluate this, we divided participants into two groups based on their average map error. A cutoff of 25 feet was used to distinguish participants who committed high and low map errors. This cutoff was chosen because it was near the median error score, and because there was a large gap between the next largest and next smallest error scores. Map errors were analyzed in a mixed-model repeated-measures ANOVA with terms for map score (high or low) and interface. The interaction between the two variables was not significant,  $F(2,20)=2.143, p=.143$ , nor was the main effect of interface (reported previously in the Map Error section), indicating that participants who scored high or low on the map task were not differentially affected by the three interfaces. These results show that the lack of significant difference in map error across interfaces is not likely due to individual differences.

### DISCUSSION

During the target search phase, participants attempted to find 10 target barrels distributed throughout the environment. Upon finding each barrel, participants judged its egocentric direction relative to their heading in the virtual environment.

The actual egocentric directions of selected targets varied as a function of the interface used to display the 360-degree view of the environment. Participants tended to select targets close to their heading at the time of selection, and this tendency was most pronounced in the 90 degree x 4 interface, somewhat reduced in the 180 degree x 2 interface, and further reduced in the 360 degree x 1 interface. One possible explanation for this finding is that image segmentation affects spatial allocation of attention.

In all three interfaces, participants preferred to allocate their attention in the direction they were moving. In the 360 degree x 1 interface, participants were able to spread their attention horizontally over a relatively wide area of the display, centered on their heading. In the 180 degree x 2 interface, participants needed to attend to the upper 180 degree view in order to center their attentional focus around their heading, which reduced their ability to detect targets in the far periphery (e.g., targets appearing on the lower view). In the 90-degree x 4 interface, participants focused their attention on the central (forward) view, and the visual borders around the central view restricted their attention to that view. Furthermore, since the reaction times between target appearance and target selection were not influenced by different interfaces, it suggests that participants were able to focus their attention to detect targets in the periphery of the wider view. We interpret these findings as evidence in favor of the 360 x 1 interface, which allowed participants to make better use of the 360-degree view of the environment.

Pointing errors when judging the egocentric directions of targets also depended on the interface. Errors were lowest in the 90 x 4 and 180 x 2 interface, compared to the 360 x 1 interface. This difference might be due to a benefit conferred by the visible image borders, which could serve as landmarks to identify key orientations relative to the participant's heading. However, the additional borders in the 90 degree x 4 interface did not lead to superior pointing performance compared to the 180 x 2 interface. One possible explanation for this is that the borders in the 180 x 2 interface were more helpful than those in the 90 x 4 interface because they distinguished between the front and back hemi-fields. The front-back distinction in body coordinates is more salient than the left-right distinction [10].

Unlike the pointing errors, which favored the 90 x 4 and 180 x 2 interfaces, map errors did not differ as a function of interface. This suggests that participants were able to create a cognitive map of the environment equally well in all three interfaces. The distinct error patterns in the pointing and map tasks suggest that the processes might be dissociable. Specifically, the data indicate that cognitive maps may have been formed on the basis of exocentric (object-to-object or object-to-environment) directions rather than egocentric directions. This conclusion is contrary to the notion that cognitive maps are formed by integrating egocentric directions with perceived self-position and orientation [8, 26]. The lack of correlation between egocentric pointing errors and map errors provides further evidence for the dissociation between perception of egocentric object directions and creation of a cognitive map.

### SUMMARY AND FUTURE WORK

The goal of this study was to investigate the effectiveness of three designs of a 360-degree view interface for a mobile surveillance system. Interfaces were evaluated based on



participants' performance of two spatial tasks and their surveyed preferences.

We plan to extend our findings on the usage of peripheral views to evaluate the mechanism underlying the observed differences across the three interfaces. By experimentally placing targets in the periphery and measuring eye and head movements, we will be able to evaluate whether selected target direction is dependent on eye gaze direction. Adding visual boundaries (e.g., dash lines) on a wider view interface such as 360-degree x 1 may significantly improve the ability to determine target directions. However, it may also interrupt the distribution of attention across the display. This enhancing technique will need to be carefully investigated to the extent of designing an effective interface. We also plan to expand our work to a passive viewing situation rather than active navigation, since passive viewing corresponds more closely with a real-world surveillance task (e.g., "Watch this 360-degree video from a panoramic-camera wearer in the field and report anything suspicious."). Additional field agents can be added for observation to explore the maximum number of agents that can be observed reliably and the ability of an observer to build a cognitive map successfully by integrating 360-degree views from multiple field agents. Results of such a passive study may differ significantly from the current study given Noë's suggestion that the ability to act on an environment is critical to perception of it [16].

An important extension of this work will be to evaluate how experience and training with each interface affects task performance. Participants in the current study experienced each interface for a total of 15 minutes (5 minutes during the familiarization phase and 10 minutes during the target search phase), but professional operators might acquire hundreds of hours of experience with a given interface. Therefore, more extensive training studies are needed before the design implications can be fully realized. The results of the current study that favor those with more video game experience suggest that additional time with a novel interface may lead to higher performance and an increased sense of naturalness. This idea is consonant with early prism glasses research that showed that over time participants could adapt and perform normally even if their right-left and up-down were reversed or distorted [21, 27].

A practical contribution of this work is the identification of a previously unreported tradeoff that designers face when choosing the optimal interface for observation and teleoperation of remote systems capable of capturing live 360-degree video. If accurate judgment of egocentric directions is of critical importance to the task, then the 90-degree x 4, the 180-degree x 2, or another interface with visual angle markers or boundaries will be most appropriate. This applies to situations in which an observer watches the video feed and relays information to the person carrying the video cameras about the egocentric location of an important navigational goal (e.g., "The building entrance

is to your right at 4 o'clock."). However, if surveillance of the entire display is the primary task, then the 360-degree x 1 interface, or perhaps another seamless interface, would be more appropriate, since it would allow for the broadest distribution of attention. This applies to situations in which an observer provides a simple but fast alert about the presence of a person or object near the person carrying the cameras, but egocentric location is of secondary importance (e.g., when identifying the presence of an enemy combatant and that personnel should immediately seek cover).

This project represents a first attempt to determine the necessary display characteristics that allow viewers to correctly interpret 360-degree video images displayed on a 2D screen. By examining performance on two spatial tasks and user preferences using three interfaces, we have established that the best design of the interface is not obvious and have offered practical guidance for practitioners creating such displays. Also, we have revealed evidence that egocentric and exocentric spatial tasks may be dissociable, which has broader implications for the design of any navigation system or virtual environment.

#### ACKNOWLEDGMENTS

We thank the U.S. AFOSR and U.S. Army Research Lab for support for this research. The 3D village model is based on *ut4\_blitzkrieg* created by HORSE-A-F and is used with permission.

#### REFERENCES

1. Irrlicht Engine - A free open source 3d engine.
2. Agarwala, A., Zheng, K.C., Pal, C., Agrawala, M., Cohen, M., Curless, B., Salesin, D. and Szeliski, R. Panoramic video textures. *ACM Trans. Graph.*, 24, 3 (2005), 821-827.
3. Baldwin, J., Basu, A. and Zhang, H., Panoramic video with predictive windows for telepresence applications. in *IEEE International Conference on Robotics and Automation*, (1999), 2884-2889.
4. Burt, P. and Adelson, E. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2, 4 (1983), 217-236.
5. Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V. and Hart, J.C. The CAVE: audio visual experience automatic virtual environment. *Commun. ACM*, 35, 6 (1992), 64-72.
6. Feng, J., Spence, I. and Pratt, J. Playing an Action Video Game Reduces Gender Difference in Spatial Cognition. *Psychological Science*, 18, 10 (2007), 850-855.
7. Fiala, M., Pano-presence for teleoperation. in *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, (2005), 3798-3802.
8. Foo, P., Warren, W.H., Duchon, A. and Tarr, M.J. Do humans integrate routes into a cognitive map? Map-

- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 2 (2005), 195-215.
9. Foote, J. and Kimber, D., FlyCam: Practical panoramic video and automatic camera control. in *IEEE International Conference on Multimedia and Expo*, (2000), 1419-1422.
  10. Franklin, N. and Tversky, B. Searching imagined environments. *Journal of Experimental Psychology: General*, 119, 1 (1990), 63-76.
  11. Greenhill, S. and Venkatesh, S. Virtual observers in a mobile surveillance system *Proceedings of the 14th annual ACM international conference on Multimedia*, ACM, Santa Barbara, CA, USA, 2006, 579-588.
  12. Hirose, M., Ogi, T. and Yamada, T. Integrating live video for immersive environments. *Multimedia, IEEE*, 6, 3 (1999), 14-22.
  13. Kadous, M.W., Sheh, R.M. and Sammut, C., Effective user interface design for rescue robotics. in *1st ACM Conference on Human-Robot Interaction*, (2006), 250-257.
  14. Liu, H., Javed, O., Taylor, G., Cao, X. and Haering, N. Omni-directional surveillance for unmanned water vehicles *8th International Workshop on Visual Surveillance*, 2008.
  15. Meguro, J., Hashizume, T., Takiguchi, J. and Kurosaki, R., Development of an autonomous mobile surveillance system using a network-based RTK-GPS. in *2005 IEEE International Conference on Robotics and Automation*, (2005), 3096-3101.
  16. Noë, A. *Action in perception*. MIT Press, Boston, MA, 2004.
  17. Park, J. and Myungseok, A. A novel application of panoramic surveillance system *IEEE International Symposium on Industrial Electronics*, Seoul, Korea, 2009, 205-210.
  18. Richman, J.B. and Dyre, B.P., Peripheral Visual Stimulation Benefits Heading Perception during Active Control in *Human Factors and Ergonomics Society Annual Meeting*, (1999), 1411.
  19. Schmidt, M., Rudolph, M., Werther, B. and Furstenaus, N., Remote airport tower operation with augmented vision video panorama HMI. in *2nd International Conference Research in Air Transportation*, (2006), 221-230.
  20. Shum, H.Y. and Szeliski, R. Panormic image mosaics *Microsoft Research Technical Report*, 1997.
  21. Stratton, G.M. Some preliminary experiments on vision without inversion of the retinal image. *Psychology Review*, 3, 6 (1896), 611-617.
  22. Sun, X., Foote, J., Kimber, D. and Manjunath, B.S. Panoramic video capturing and compressed domain virtual camera control *Proceedings of the ninth ACM international conference on Multimedia*, ACM, Ottawa, Canada, 2001, 329-347.
  23. Szeliski, R. Image mosaicing for tele-reality applications *DEC and Cambridge Research Lab Technical Report*, 1994.
  24. Tang, W.-K., Wong, T.-T. and Heng, P.A. A system for real-time panorama generation and display in tele-immersive applications. *Multimedia, IEEE Transactions on*, 7, 2 (2005), 280-292.
  25. Vincent, L. Taking Online Maps Down to Street Level. *Computer*, 40, 12 (2007), 118-120.
  26. Waller, D., Loomis, J.M., Golledge, R.G. and Beall, A.C. Place learning in humans: The role of distance and direction information. *Spatial Cognition and Computation*, 2, 4 (2000), 333-354.
  27. Welch, R.B. Discriminative conditioning of prism adaptation. *Perception and Psychophysics*, 10, (1971), 90-92.
  28. Xiong, Y. and Turkowski, K., Creating image-based VR using a self-calibrating fisheye lens. in *IEEE Computer Society on Computer Vision and Pattern Recognition*, (1997), 237-243.