Task-Driven Camera Operations for Robotic Exploration

Stephen B. Hughes and Michael Lewis

***Abstract*—Human judgment is an integral part of the teleoperation process that is often heavily influenced by a single video feed returned from the remote environment. Poor camera placement, narrow field of view, and other camera properties can significantly impair the operator's perceptual link to the environment, inviting cognitive mistakes and general disorientation. These faults may be enhanced or muted, depending on the camera mountings and control opportunities that are at the disposal of the operator. These issues form the basis for two user studies that assess the effectiveness of existing and potential teleoperation controls. Findings suggest that providing a camera that is controlled independently from the orientation of the vehicle may yield significant benefits. Moreover, there is evidence to support the use of separate cameras for different navigational subtasks. Third, the use of multiple cameras can also be used to provide assistance without encroaching on the operator's desired threshold for control.**

***Index Terms*—Human–automation interaction, human–robot interaction, remote perception, telerobotics functional presence, usability evaluation.**

I. Introduction

ROBOTIC navigation allows expendable, albeit expensive surrogates to explore and inspect environments that are otherwise prohibitive. Regardless of whether the robot is directly manipulated by an operator, or granted full autonomy to execute its mission, at some level, human observation, supervision, and judgment remain critical elements of robotic activity. Fong and Thorpe [15] characterize several types of interfaces used to keep humans involved in robotic activities, but observe that direct control while watching a video feed from vehicle mounted cameras remains the most common form of interaction. The ability to leverage experience with controls for traditionally piloted vehicles heavily influences the appeal for this interaction style.

Unfortunately, limiting interaction with the remote environment to the visual channel causes a breakdown of perceptual modalities, as well as a lack of important vestibular and proprioceptive cues. Moreover, the field of view provided by the video feed is often much narrower than human vision, adding to the handicaps of remote perception. This impairment at the perceptual level leaves the operator prone to numerous, well-known operational errors, including disorientation, degradation of situation awareness, failure to recognize hazards, and simply overlooking relevant information [9], [29].

Manuscript received August 1, 2004; revised February 15, 2005 and March 9, 2005. This work was supported by the AFOSR under Contract F49640-01-1-0542. This paper was recommended by the Guest Editors.

The authors are with the School of Information Science, University of Pittsburgh, Pittsburgh, PA 15260 USA (e-mail: shughes@mail.sis.pitt.edu; ml@sis.pitt.edu).

Digital Object Identifier 10.1109/TSMCA.2005.850602

Accepting that remote perception will never match direct perception, the objective of teleoperated systems should be to achieve *functional presence*. This occurs when the operator receives enough cues to maintain situation awareness and successfully conduct operations in the remote environment [40]. Unlike the conventional understanding of "presence," functional presence does not require operators to have the sense that they actually are situated at the remote location, only that they can accurately process the data that they are afforded.

The mechanisms provided to operators for manipulating this video stream will have a dramatic influence on the overall experience. However, before affordances can be discussed, it is important to reflect on two important subtasks that will engage the operators: Navigation and Inspection [36]. Navigation describes the act of explicitly moving the robot to different locations in the environment. It can take the role of exploration to gain survey knowledge, or traversing the terrain to reach a specific destination. Inspection, on the other hand, describes the process of acquiring a specific viewpoint—or set of viewpoints—relative to a particular object. While both navigation and inspection require the robot to move, an important distinction is the focus of the movement. Navigation occurs with respect to the environment at large, while inspection references a specific object or point of interest.

Switching between these two subtasks may play a major role in undermining situational awareness and functional presence in teleoperated environments. For example, since inspection activities move the robot with respect to an object, viewers may lose track of their global position within the environment. Additional maneuvering may be necessary to reorient the operator before navigation can be effectively resumed. Well thought out camera configurations and control strategies may be able to mute the disorientation, or at least hasten the recovery.

This paper seeks to understand the relationship between camera mountings, control opportunities and functional presence. Specifically, it is hypothesized that multiple cameras, delegated to explicit tasks, can be used to mitigate some of the problems with situational awareness, and increase the effectiveness of exploration.

II. Robotic Simulation With Game Engines

Milgram has observed strong parallels between the interaction required to navigate remote and artificial environments [30]. This relationship benefits our efforts in two ways. First, given the advances in realistic virtual models, teleoperation interfaces can be prototyped with high fidelity using virtual environment (VE) technology [27]. Second, design of interfaces

for robotic exploration can draw on the extensive literature of viewpoint control from VEs.

The computer-gaming market has made a dramatic impact on the development of computer graphics technology and has driven the required hardware down to the commodity level [35]. Contemporary games are designed in a modular fashion, separating the simulation code from the environmental data (levels) and the rules of the game. This modularity encourages end-users to make modifications to the environment as well as the behavior of the game, building on complex features that are common to most simulations, such as collision detection, Newtonian physics, lighting models, and fundamental inout/output (I/O) routines. This flexibility means that game engines can provide powerful, inexpensive tools for the researchers who need interactive three-dimensional (3-D) modeling and graphics [26].

One of the purported benefits to exploring an artificial environment is that constraints of the physical world can be abandoned. For example, viewers can instantaneously teleport from one spot to another. However, this kind of activity has proven disorientating to many users, pushing for design of more natural interactions—the type that are likely to be useful to robotic activities. At the same time, easing other physical restrictions may actually inform the design of robotics interfaces. For example, it is common to relax rules of collision detection such that minor disturbances in the environment do not impede locomotion. Certainly robotic operators cannot alter the laws of physics, but by granting the robot autonomy for local movements, minor obstacles could be avoided, offering a similar interaction experience to the operator.

III. Robotic Viewpoint Control Issues

Adjusting the viewpoint in VEs have been identified as having a profound impact on all other user tasks [21]. Several metaphors have been developed to reduce the cognitive load required to manipulate the viewpoint to meaningful perspectives [7], [17]. While not one technique is optimal in the general case [5], investigation in this field has framed the key factors that should be considered in the design of specific instances.

A. Control Mapping Strategies

While six degrees-of-freedom (6DOF) are necessary to fully control the position (X, Y, Z) and orientation (yaw, pitch, roll) of the viewpoint, the cognitive overhead of operating a 6DOF device may be more than the average user can handle. Arthur observes: "When interaction becomes highly attention demanding, memory for the present location frequently decays, with the result that the individual becomes lost in space" [1]. Moreover, many familiar input devices, such as joysticks and mice, do not offer six control options. This leads to several alternative strategies for control mappings.

Overloading: Extra degrees of freedom are achieved by modal operation of the device. Various combinations of control keys or button presses supplement the operation of the device to determine the mode of operation. While this technique is popular with computer-aided design (CAD) and modeling software, the increased cognitive burden of remembering the current state of the controller can negatively impact performance [4]. Alternatively, multiple instances of a lower-order control device can be used.

Constraining: Movement of the viewpoint is limited to certain operations; manipulations of other attributes are simply not permitted. The most common example of constraining is to restrict motion to a ground plane, eliminating the need for vertical translation [39]. Roll is also frequently eliminated, especially in simulations of ground vehicles.

Coupling: This approach functionally binds one or more viewpoint attributes to the state of the others. The most common example is the gaze-directed steering metaphor in which the viewer's motion is determined by the direction they are looking (as described below) [6].

Offloading: This method cedes control of certain travel operations to an external source. These sources may include a precomputed route or sequence, a collaborative operator or even an autonomous agent.

These four techniques are not exclusive; in fact, some combination is often employed to bring the control space from 6DOF to match the affordances of the controller.

B. Camera Configurations

There is appreciable variability in strategies for generating video feeds. Designers who wish to balance the appeal of direct control with the complications of remote perception may need to consider the range of techniques. McGovern provides accounts of robotic systems that include independently controlled cameras, cameras that are dependent on the steering mechanisms, and multiple fixed cameras [29].

Fixed Camera Controls: Mounting a fixed camera on the front of a robot yields the equivalent of the popular VE gaze-directed steering interface. The operator can repositions the viewpoint through a combination of adjustments to the direction and speed of motion [31]. Typically, this works in a sequential manner; the viewer selects a promising orientation, and then initiates movement, telling the camera to "move forward," until the desired viewpoint has been acquired. This coupled approach has become one of the most pervasive forms of viewpoint control in VEs, perhaps because of its intuitive nature; it is much like driving a car. However, this technique offers a strong bias toward navigation tasks at the expense of all but the most trivial inspections. Using this technique to navigate, the operator only needs to be concerned with two degrees of freedom: the orientation of the robot (which direction is it facing) and the velocity (forward or backward motion). Inspecting an object is much more difficult. Consider the task of looking at an object from all sides. Since the robot always moves forward in the direction that the camera is oriented, the operator must periodically stop moving, pivot the robot to acquire the desired view of the object, and then pivot back to resume motion. There is no guarantee that the object of interest even remains in the field of view, increasing the chances that useful viewpoints may be overlooked or missed [11]. Knowing when to turn to face the object requires that the controller have a good sense of the overall configuration and scale of the object and the environment. For many robotic exploration applications, it is unlikely that these conditions are reliably met.

Independent Camera Controls: Allowing for an independently controlled camera with constraints on elevation and roll reduces the control space to four degrees-of-freedom. This might be implemented using two joysticks (one for positioning the robot and the other for orienting the camera) or a joystick with a hat-switch. This overcomes the problem of not being able to look in one direction while moving in another, however, designers of VEs shun this technique for just that reason. Baker and Wickens offer a representative statement: "Travel-gaze decoupling…makes a certain amount of 'ecological' sense, since we can easily look to the side while we move forward. This is probably too difficult to implement and the added degrees of freedom probably add to the complexity of the user's control problem" [3]. While decoupling the camera facilitates inspection by allowing the controller to keep interesting objects in view, navigation suffers. The simple travel command "Move Forward" may meet with unexpected results unless the viewer has a good understanding of how the camera is oriented relative to the front of the vehicle [13]. Fortunately, independent controllers have the option of realigning the direction of gaze and direction of motion when performing any extensive navigation activities. However, this may factor into the "complexity of the control problem," referenced above by Baker.

Multiple Cameras: The prospect of equipping teleoperated robots with multiple cameras is frequently raised to support stereopsis. In these scenarios, two cameras are focused on the same point. The disparity in the placement of the cameras allows computer vision algorithms to resolve topological ambiguities. Using multiple video streams has also been considered for so-called marsupial teams of robots, where a second robot provides a supplementary, exocentric view of the first robot. This exocentric view can be useful in disambiguating obstacles that may have immobilized the primary robot, allowing recovery from otherwise fatal mistakes [32].

Two cameras, mounted on the same robot may also be used to align with the subtasks of inspection and navigation to further reduce the disruption of task-switching. A fixed screen, coupled with the orientation of the robot could be used for navigation, while the controllable camera could be manipulated for inspection. Switching tasks would simply be a matter of selecting which feed requires attention.

C. Ecological Cues Versus Instrumentation

Assuming that gaze-travel decoupling is permitted, situational awareness may degrade if the operator cannot quickly assess the angular magnitude of displacement. Ecological cues, such as visual flow or fixed, peripheral references may provide the operator with some insight to the camera orientation relative to the robot's heading. However, it is unclear if these cues are sufficient. Numerous other studies have evaluated the effectiveness of various instruments to assist with spatial cognition including: you-are-here maps, compasses, trails, viewtracks, etc. [10], [37]. To track displacement between the orientation of the robot and an independently controlled camera, a two-handed compass was developed. Pictured in Fig. 1, the viewer can use this instrument to instantly detect misalignment between the orientation of the robot (the short hand) and the orientation of the camera (the long hand).

Fig. 1. Two-Handed compass.

A user study, described in the following section, explores the impact of various control mappings and instrumentation options.

D. Automated Tasks

Successful navigation is also frequently dependent on adopting sophisticated strategies such as acquiring survey views, or moving is structured patterns [5]. Even with these strategies, there is the chance that the effort applied to manipulating the viewpoint will distract from extracting the desired information.

Offloading some of the viewpoint controls to an automated system can mandate effective navigation strategies while simultaneously reducing the control burden. The hallmark of automation is that the machine takes responsibility for the completion of certain tasks. While the notion of a fully autonomous entity replacing human presence is appealing, human observation and supervision remain critical elements of robotic activity. Collaborative control systems have shown great promise, allowing the human operators to focus their efforts on perceptual judgments and decision making that exceed the current capability of automation [15].

Studies of how people interact with 6DOF devices reveal that there is a division between interaction with translation and orientation controls [20], [28]. People tend to issue clusters of commands, toggling back and forth between sequences of translations and sequences of rotations. This natural boundary suggests a division of labor between manual and automated viewpoint control, yielding two paradigms: guided positioning systems and guided orientation systems.

In a guided positioning system, assistance is provided with moving the viewpoint through the environment. There are multiple ways to establish the route, depending on the amount of environmental knowledge afforded to the system. At the most basic level, the route may be a preprogrammed sequence of steps through the environment, offering a generic tour. Generalizing this approach, the viewer may be able to specify a set of interests, generating a more personalized tour [12]. When the system has very little foreknowledge of the environment, it will likely adopt a naïve search strategy that systematically moves the viewpoint through the environment. Examples include the lawnmower method, which moves along narrow, adjacent strips, and radial search where exploration progresses in increasing concentric circles or contour following [10].

Guided positioning may also occur on a more localized level. Kay's STRIPE system (Fig. 2), provides automatic positioning, constrained by an explicit set of operator defined waypoints [25]. The automation attempts to reconcile the known ground-plane with the next waypoint to determine which direction to

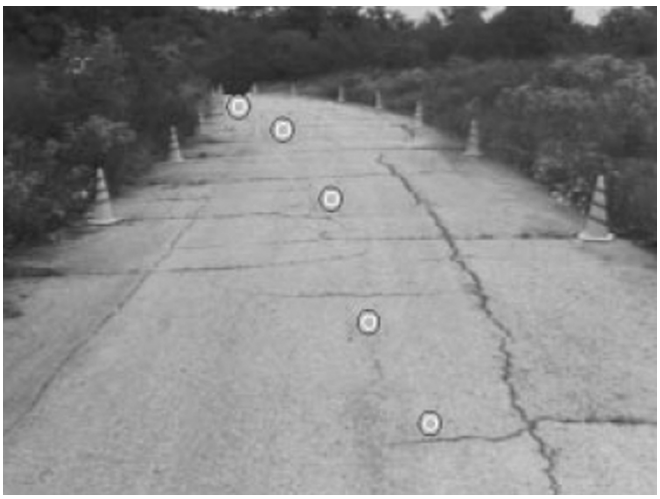


Fig. 2. Kay's stripe system [25].

move the vehicle. The system may factor in obstacle avoidance or other intelligent behavior to automatically position the robot according to the operator's wishes.

Guided orientation systems offer the inverse automation: The operator moves the position of the viewpoint, but the camera orientation is automatically adjusted. Bajscy outlines two important tasks that must be supported to effectively implement automatic gaze redirection: shifting and holding [2]. Gaze shifting involves transitioning the focus of the camera from one point of interest in the environment to another, while gaze holding describes the activity of keeping an interest point in focus despite camera movement or other environmental changes. Guided orientation can also be used to provide cues as to how the camera is moving. By predictively panning the camera when nearing a turn or tilting when approaching a staircase, a more natural interaction can be achieved [33].

Previous mixed-initiative robotic systems have emphasized positioning operations and path-planning for the robot [15]. Bruemmer, for instance, promotes granting the robot the ability to "veto dangerous human commands to avoid running into obstacles or tipping itself over" [8]. While poor positioning (either by the human or the robot) can clearly jeopardize the safety of the robot, proper orientation is just as critical to the success of the robot's mission. If the robot is looking in the wrong direction, relevant information can easily be overlooked.

While guided orientation has not played a prominent role in the robotics literature, the VEs literature offers some insight into this issue. Constrained Navigation and the Attentive Camera are two approaches that promote a supportive, yet unscripted explorations [18]. Using these techniques, the orientation of the camera can be systematically redirected to relevant features based on a viewer-determined location in the environment. User evaluations of these techniques have revealed significant benefits including: increased the likelihood that key viewpoints are utilized [19]; better understanding the presence and configuration of key elements [22]; and reduced search time to locate key elements [23].

Despite the clinical successes of the Attentive Camera, anecdotal feedback indicated strong dissatisfaction with automatic reorientation of the camera. Frustration likely stemmed from the lack of coordination between the operator and the autonomous agent. The system may have been trying to show a critical feature to the operator who was otherwise engaged in piloting to a new location. The perception that the system was working contrary to the viewer's immediate task led to frequent stops to "correct" the system, potentially overriding the some of the benefits of the automation. At a minimum, this "wrestling for control" had a negative impact on the overall complexity of the interaction, which diverted valuable time from the primary objective [23]. A second user study implements a variation of the Attentive Camera in an attempt to address these issues and may have implications for robotic exploration.

IV. User Study: Camera Configurations

A. Design and Procedure

A user evaluation was conducted to assess the impact of these three camera configurations and the role of instrumentation on exploration tasks in a simulated teleoperation environment, resulting in five conditions:

1) single fixed camera, no instrumentation;
2) single independent camera, no instrumentation;
3) single independent camera, two-handed compass;
4) multiple (fixed + controllable) cameras, no instrumentation;
5) multiple (fixed + controllable) cameras, two-handed compass.

Subjects were asked to explore a nontrivial environment with the task of locating as many target objects as possible. Targets were identified on two levels of specificity. Objects were to be initially identified by class and then confirmed by a discriminating feature. For example, a target might be described to the searcher as a red cube with a "J" on one face. This task forces the explorers to:

1) locate an object from a distance;
2) position the robot nearer the potential target;
3) inspect the object more closely to identify the discriminating feature.

Prior to starting the task, participants were given verbal instructions on the objectives, and a demonstration of the controls. Participants were advised that the robot had a fixed amount of energy and that they should continue to explore until the robot stopped responding to their commands. In fact, all trials were timed to last exactly fifteen minutes. A training period allowed the subjects to familiarize themselves with the robot's capabilities. All participants were required to confirm an understanding of the task and the controls by identifying at least one target object in a training environment.

Two separate environments were used to counterbalance the effects of individual strategies. The first environment (shown in Fig. 3) loosely resembled a warehouse structure, with two levels connected by a ramp. The warehouse was comprised of a series of rooms that were arranged such that there was no obvious or continuous path. This closed layout meant that targets were generally not visible from a distance; navigation to each room was necessary to verify its contents.

The second environment resembled a more rugged outdoor environment with characteristics of a canyon or desert (shown in

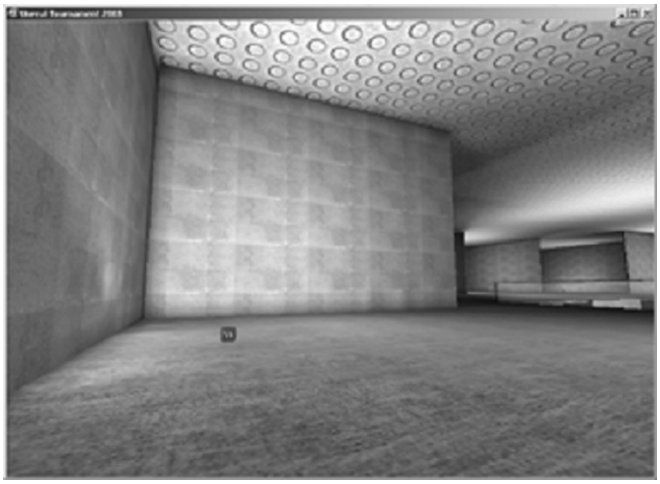


Fig. 3. Screen shot of indoor environment.

Fig. 4. Screen shot of outdoor environment.

Fig. 4). Generally, the second environment was more open than the first, although several mountainous structures prevented the entire scene from being surveyed from a single vantage point. Unlike the first environment, target objects could be obscured by irregularities in the terrain; small craters or ridges might conceal a target unless it was viewed from precisely the right viewing position. Additionally, the second environment was much more expansive than the first (about four times the land area). Success in this environment required coverage of more terrain rather than intricate navigation. Aside from the target objects, both environments are void of nonarchitectural features.

Twelve targets were evenly distributed throughout both environments. Targets consisted of a red cube marked on one side with a yellow letter. Participants were advised that not all letters of the alphabet would be represented, nor were they in any particular sequence. Placement of the targets ensured that it was always possible to acquire a view of the letter (i.e., the letter was never face down). However, the identifying side was occasionally placed in close proximity to a wall or other obstruction. This limited the conspicuity of the letter and forced the controller to explicitly maneuver to acquire a useful point of view.

Data were recorded in the form of a written list of all targets identified, as well as in an automatically recorded log file that tracked the position, velocity and orientation (for both the robot and camera). These time-stamped entries were written to a log file whenever the viewer issued a command, allowing for a complete reconstruction of each session.

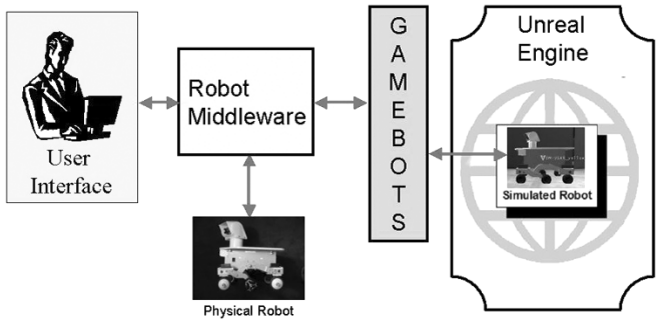


Fig. 5. Architecture of simulation.

B. Participants

Sixty-five men and women were paid to evaluate five camera control strategies (13 per condition). Participants were recruited from the University of Pittsburgh community, with most subjects enrolled as undergraduates. Given the dependence on vision and identification on color objects, it was stipulated that participation required a self-report of normal or corrected-to-normal color vision. One participant terminated the experiment prior to completing the full experiment, but data were still included for the completed portions. Three additional participants were excluded from the study due to corruption of the log files that prohibited analysis.

C. Apparatus

Each of these conditions were implemented using the simulated four-wheeled Urban Search and Rescue robot described by Lewis et al. [27]. Fig. 5 shows a schematic of the simulation architecture. The bulk of the simulation is handled by Epic Games' Unreal Tournament (UT) Game Engine [14], including structural modeling for the robot and the environment and the physics of their interaction. Programmatic control of the robot was achieved through the use of the GameBots API, which relays simple text commands through a TCP/IP socket [24]. The GameBots commands were, in turn, issued from the robot middleware package that connects to the user interface. This means that the simulated robot was directed from the same control panel that is used to control physical robots.

The robot was controlled using a Logitech Extreme digital 3-D joystick. The main stick control was used to direct the position of the robot (forward and backward motion incrementally influenced the velocity of the robot, while side-to-side motion caused the robot to pivot). In the appropriate conditions, the orientation of the camera was controlled using the hat-switch on the top of the joystick (Yaw was controlled by lateral movement, Pitch was adjusted by moving the hat switch forward and backward). The display was presented on a $21''$ monitor using 800×600 resolution. For the two-camera conditions, a second $21''$ monitor was added: one monitor displayed the video feed from the fixed camera, while the second displayed the feed from the independent camera.

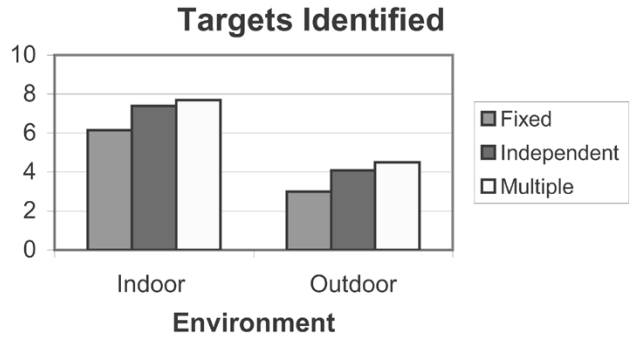


Fig. 6. Targets identified.

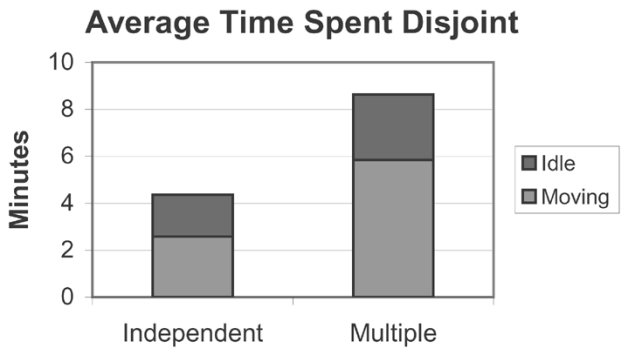


Fig. 7. Disjoint times.

TABLE I
DIFFERENT IN NUMBER OF TARGETS IDENTIFIED

	Indoor	Outdoor
Fixed, Independent	t(37) = 1.75, p <.05	t(36) = 2.00, p < .05
Fixed, Multiple	t(37) = 1.98, p <.05	t(37) = 2.39, p < .05
Independent, Multiple	t(50) = 0.48 p > .05	t(49) = 0.76 p > .05

D. Results

Data were first analyzed to determine if there were differences in effectively completing the task. With respect to the number of markers found, there were two findings in the initial investigation that will impact the way that the analysis proceeds.

1) Across all conditions, significantly more objects were found in the indoor environment (mean $= 7.2$, $sd = 2.3$) than the outdoor environment (mean $= 4.0$, $sd = 1.8 t(127) = 8.78$, $p < 0.01$). This can probably be attributed to the increase in space and corresponding sparseness of the targets. However, it may also be caused by the absence of well-defined places to search for the targets.
2) The two-handed compass did not produce a significant difference in number of targets identified for any of the independent trials.

As a result of these findings, the data was pooled for the following analysis: Comparisons were made between one-camera fixed (Fixed), one-camera independent (Independent), and two-camera conditions (Multiple) and within the indoor and outdoor trials.

Fig. 6 shows that both the Independent and Multiple conditions outperformed the Fixed condition in terms of the number of markers identified. The statistical figures are presented in Table I.

This result is further supported by an analysis of the uses of the controllable cameras. Recall that panning the camera is left to the discretion of the viewer; if the controller opts to not exercise the option of panning the camera, the control effectively degenerates into the fixed condition. With this in mind, movement logs were analyzed to extract the amount of time that the camera orientation was disjoint (greater than 10° from the vehicle orientation in either direction). A correlation was found between the amount of time that the controller was disjoint and the number of markers found (Independent: $N = 50$, mean disjoint time $= 6 : 11$, $sd = 1 : 42$, $r = 0.41$, Multiple:$N = 52$ mean disjoint time $= 10 : 20$, $sd = 2 : 45$, $r = 0.45$). Operators who did not avail themselves of the camera orientation controls did not seem to perform as well as those that exercised that option.

Although there were no differences detected in the effectiveness of the Independent and Multiple conditions, an analysis of the movement logs reveals that strategies used to manipulate the robot were fundamentally different. Specifically, the following measures were extracted from the log files:

1) Pan Motions—The commands issued to adjust the yaw of the independent camera.
2) Disjoint Time—The number of ticks where the orientation of the camera varied from the orientation of the robot in excess of 10°.
3) Disjoint Motion—Disjoint Time when the robot was also moving.
4) Idle Disjoint time—Disjoint time where the robot is neither panning the camera nor moving.
5) Recoupling—the number of times where the angular displacement between the independent camera and the orientation of the robot was reduced, and the magnitude of the displacement was within 10°.

For each of these measures, there were no pair-wise differences between the indoor and outdoor conditions, suggesting that individuals essentially controlled the robot in a similar manner regardless of the environment.

Fig. 7 shows that the Multiple condition spent almost twice as much time disjoint than the Independent conditions. This result was significant for both disjoint motion and idle disjoint times, $t(100) = 7.40$, $p < 0.01$ and $t(100) = 3.33$, $p < 0.01$. This does not mean that users in the Multiple condition were better able to deal with the ambiguity of decoupled motion. Instead this result probably reflects the operator shifting their attention to the view with the fixed camera screen, leaving the camera in the disjoint position until it was needed again. Participants controlling the one-camera robots were not afforded this luxury and were therefore more likely to recouple the camera with the orientation of the robot in order to comprehend their direction of travel for large scale movements (Independent mean: 87 recouples Multiple mean: 62 recouples, $t(100) = 3.98$, $p < 0.01$).

Given that operators in the Independent conditions were continuously realigning their gaze with travel, one might expect

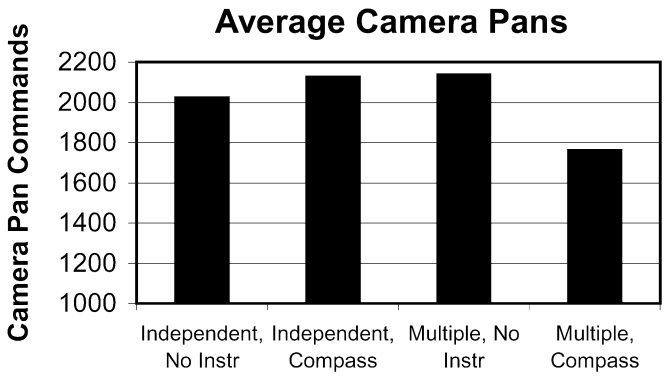


Fig. 8. Multiple camera interaction.

these actions to be reflected by an increased number of panning motions. However, this was not always the case, as shown in Fig. 8. The two-camera with compass condition had significantly fewer pans ($t(74) = 1.81$, $p < 0.05$), but there was no difference between the one-camera conditions and the two-camera, no instrumentation condition ($t(74) = 0.39, p > 0.05$). This result suggests that the operators of the Multiple conditions were dividing their attention across the two video feeds and were not maintaining the state of the controllable camera when they were not using it. The compass allowed operators to quickly reorient themselves when they returned their attention to the controllable camera, while the operators who had no instrumentation may have been using additional panning motions to reestablish their situational awareness. Similar results are found in the surveillance literature where traditional interfaces require the operator to switch between multiple video streams projected on a small number of displays [34]. Operators need a brief time to reassimilate their view and visually interpret the orientation in which the camera was last used.

E. Discussion

The data collected from this study suggests that the use of an independent, controllable camera increases the overall functional presence, as witnessed by improved search performance. The major shortcoming of the coupled camera seems to be its inability to efficiently perform inspection activities—acquiring a useful point of view within a limited range.

While the two independent techniques that were examined did not show quantitative differences in terms of search performance, they both offered qualitatively different experiences. Understanding these differences, we may be able to exploit them for better still performance. At a minimum, the two techniques offer variety—designers can cater to preferences or individual differences. Optimizations might produce more tangible improvements. For example, knowing that there is a need to realign the view with the orientation of the robot may standardize a control that automates that process. Likewise, further study of the two-camera display may find that one of the screens is more dominant, suggesting the use of a higher-resolution camera, or a screen-in-screen approach.

Finally, the parity of the two-camera display offers some interesting opportunities for off-loading control of the camera to an autonomous agent. In terms of guided positioning, the two-camera approach may prove beneficial to route-drawing systems like Stripe, affording a constant substrate to specify travel orders and monitor progress. Alternatively, this configuration may also prove useful in guided orientation. Having both the human and the robot battle for control of a single camera has been reported as exceedingly disruptive. However, the two-camera approach might allow for a more cooperative collaboration, where one screen represents human control, while the second screen is sensor-driven.

V. Follow-Up Condition: Guided Orientation

Based on the analysis of the camera configuration study, a follow-up condition was added to the user study in order to assess the viability of using the two-camera configuration for guided orientation.

A. Procedure

Participants were assigned to the same task as the previous experiment: a timed exploration with the goal of finding and identifying target objects. For this evaluation, the indoor, warehouse-like environment was used. Since there were no noticeable differences in operator behavior in the first experiment, the outdoor environment was not evaluated.

The experimental treatment varied according to whether or not orientation assistance was provided, yielding two conditions: sensor-driven orientation and user-controlled orientation.

Sensor-Driven Orientation (Assisted): The viewer supervised two monitors: one fixed-orientation, one independent-orientation. In addition to the pan-tilt commands issued from the viewer, the second monitor also reflected the recommendations of a guided-orientation system. Designed to simulate the effects of a line-of-sight proximity sensor, the second camera would shift the camera to fixate on the closest visible cube. If no cubes were detected, the camera would be rotated to align with the heading of the robot. The operator still had the ability to pan and tilt the second camera, temporarily overriding the recommendations, but automation would resume when the robot was moved forward or backward.

User-Controlled Orientation (Unassisted): To control for the effects of assistance, this experiment compares the results of the Sensor-driven orientation to the two-camara + compass condition from the previous experiment. Again, this treatment simulated two cameras mounted on the robot, each displayed on a separate monitor. One monitor represented the video feed from a camera mounted in a fixed, forward-facing position, allowing the operator to understand the heading of the robot. The second screen reflected pan and tilt alterations to the orientation of the camera.

B. Participants

This evaluation recruited 13 new undergraduate students from the University of Pittsburgh to experience the sensor-driven orientation condition. This was designed to balance the 13 subjects who previously had experienced the user-controlled orientation (two-camera + compass) condition. Unfortunately, one of the assisted participants had to be excluded due to a corrupt data file. Upon completion, participants were compensated for their

involvement in this study. All participants self-reported normal or corrected-to-normal color vision.

C. Results

Data were first analyzed with regard to the number of targets successfully identified. The sensor-driven orientation condition consistently identified more targets ($M = 9.1$, $\mathrm{sd} = 1.1$) than the user-controlled condition (mean $= 7.5$, $\mathrm{sd} = 2.5$) $t(23) = 1.93$, $p < 0.05$, indicating that at a broad level that the operator was benefiting from the assistance. Examining the results a bit deeper, however, reveals some interesting nuances that explain this difference.

The assistance provided by the sensor-driven condition did not make the operator more sensitive to the presence of target objects. Given that the robot was close enough to the target to activate the sensors, there was no difference in the number of targets overlooked between the sensor-based and user-controlled conditions; each condition averaged around one overlooked target per trial. Even though the sensor-based condition adjusted the view of the second camera, the operators either weren't paying attention or failed to notice the shift in gaze. Recall that target identification was a two-step process: 1) locate the target and 2) identify the letter on the target. While the assistance did not seem to help with the first stage, it made a strong impact on the second. A pair-wise analysis reveals that the time spent inspecting the targets was nearly 20 s less under the sensor-driven condition: $t(11) = 3.40$, $p < 0.01$. This difference can be directly attributed to the way the viewing parameters were manipulated. Consistent with previous research, the user-controlled treatment primarily toggled between position and orientation adjustments; the two were simultaneously adjusted less than 2% of the time. In contrast, simultaneous movement and panning occurred on the order of 60% in the assisted condition. The individual manipulation of the parameters thus resulted in longer time to identify the target object. The benefit of having a shorter identification time is that it leaves more time to search the rest of the environment, potentially exposing the operator to more targets. This inference is bolstered by an analysis of the movement logs which indicate that the assisted condition moved the robot nearly 13% more than the user-controlled condition ($t(23) = 2.26$, $p < 0.05$), despite issuing roughly the same number of commands overall.

D. Discussion

These results show that automatic gaze redirection in the two-camera paradigm can help with identifying objects in a search task. While the assistance was intended to help with both shifting the gaze to attract the operator's attention and holding the object in view for inspection purposes, the benefits seemed to be derived largely from the later operation. It was disappointing that the system was not better at assisting with target location, however noticing targets on the screen was left entirely to the to the operator. It may be possible for the robot to shoulder some of this burden by taking a more active role in alerting the viewer (e.g., with an auditory cue) that it has found something interesting, and would like the operator to take a look [16].

Unlike previous studies in guided orientation, the operators did not seem to struggle to maintain control of the viewpoint. There is currently not a direct comparison to a one-camera guided orientation system. However, analysis of the movement logs does not show excessive "homing" of the viewing orientation witnessed in previous studies. Anecdotal responses at the conclusion of the experience also seem to confirm the lack of intrusiveness. Instead of overriding the system recommendations, the viewer could opt to temporarily disregard the assistance if they were engaged in another attention-demanding task.

From the perspective of the autonomous agent, little to no effort had to be devoted to coordinating its actions with the viewer. Traditionally, if the viewer overrides the agent, there are a host of problems associated with attempting to intuit why the recommendation was overridden and how that should impact future recommendations. Relegating the agent's assistance to a secondary screen means that its actions are less disruptive and therefore suggests that errors caused by a lack of coordination might be mitigated.

VI. Implications and Future Directions

An important question to ask with any simulation-based experiment is, "How well do these results generalize to the real world?" In this case, should we expect similar results given the actual dynamics of interacting with real robots?

From an engineering perspective, simulations are frequently measured in terms of the completeness of the model. The structure of environment is deemed accurate if all the components are present and proportionally scaled. Likewise, the physical properties of the environment and the robots must behave consistently with the real world. The design specifications of USARSim [38] address these concerns. However, building robots and environments to a common scale and modeling values from specification sheets are only the first steps; they cannot guarantee a simulation's fidelity. To reap the benefits of simulation, the synthetic experience must elicit the same operator behavior as the physical environment.

We have informally observed that tasks that caused difficulties in the real environment also caused problems in the simulation. For example, negotiating corners without getting stuck. While this kind of anecdotal feedback is encouraging, more rigorous assessments are necessary to identify the degree of operator correspondence between USARSim and actual robots. To this end, validation studies directly comparing operator behavior between real and simulated robots are currently underway.

Even without these studies, the results of this research can still inform the design of robotic control interfaces. The first major result of this work concretely demonstrates the ineffectiveness of the fixed-camera strategy with regard to performing inspection-based tasks. In addition to the cognitive burdens that this approach will introduce, there is the problem of constantly making physical adjustments to the orientation of the robot. Not only is the probability that the robot will get stuck or be obstructed increased, but designers should also be concerned about

the amount of energy that is required to repeatedly pivot the entire robot back and forth.

The follow-up exploration into guided orientation provides some good initial data regarding the potential of the two-camera approach to resolve some of the intrusiveness that frequently characterizes this type of system. While the lack of a direct comparison to a single-camera control condition precludes any definitive answers, we have gained a strong insight that this approach could be useful. Furthermore, the results do show that gaze-holding can be a powerful form of assistance, and might even be a useful task to cede to automation in a mixed-initiative control structure.

Despite the success of the two-camera paradigm demonstrated by this work, there are still some practical concerns with adopting this approach. The largest of which is the bandwidth consumed by transmitting two video feeds from the remote location. Bandwidth is already the most precious resource in teleoperation activities and there is often difficulty sending one quality video feed, let alone two. Hopefully, technological advances will eventually obviate this problem, but in the meantime, the results of this study can still inform design of recommendation systems.

A second practical concern with this research is that it did not factor in the impact of imperfect information and trust in the recommendation system. In this sterile experiment, the recommendation system was afforded a perfect understanding of the environment and always offered meaningful, relevant assistance. Further work needs to be done to assess whether or not the benefits recorded in this study will hold up in the face of occasional bad advice.

REFERENCES

[1] E. Arthur, P. Hancock, and S. Chrysler, "Spatial orientation in real and virtual worlds," in *Proc. Human Factors Ergon. Soc. 37th Annu. Meeting*, Seattle, WA, 1993.

[2] R. Bajscy, J. Kosecka, and H. Christiansen, "Discrete event modeling of navigation and gaze control," *Int. J. Comput. Vision*, vol. 14, no. 2, pp. 179–191, 1995.

[3] M. P. Baker and C. D. Wickens, "Human Factors in Virtual Environments for the Visual Analysis of Scientific Data,", NCSA-TR032 and Inst. of Aviation Rep. ARL-95-8/PNL-95-2, 1995.

[4] R. Beaten, R. DeHoff, N. Weiman, and P. Hildebrandt, "An evaluation of input devices for 3-D computer display workstations," in *Proc. SPIE-Int. Soc. Opt. Eng.*, 1987.

[5] D. Bowman, "Interaction techniques for common tasks in immersive virtual environments: Design, evaluation and application," in *Doctoral. Comput. Sci.*. Atlanta: Georgia Inst. Technol., 1999.

[6] D. Bowman, D. Koller, and L. Hodges, "A methodology for the evaluation of travel techniques for immersive virtual environments," *Virtual Reality: Res., Develop., Applicat.*, vol. 3, no. 2, pp. 120–131, 1998.

[7] D. Bowman, E. Kruijff, and J. La Viola, "An introduction to 3-D user interface design," *Presence: Teleoper. Virtual Environ.*, vol. 10, no. 1, pp. 96–108, 2001.

[8] D. J. Bruemmer, J. L. Marble, D. D. Dudenhoeffer, M. O. Anderson, and M. D. McKay, "Mixed-initiative control for remote characterization of hazardous environments," in *HICSS*, Waikoloa Village, HI, 2003.

[9] R. Darken, K. Kempster, and B. Peterson, "Effects of streaming video quality of service on spatial comprehension in a reconnaissance task," in *Proc. Meeting I/ITSEC*, Orlando, FL, 2001.

[10] R. Darken and J. L. Siebert, "Wayfinding strategies and behaviors in large virtual worlds," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, Seattle, WA, Nov. 1996.

[11] A. Datey, "Experiments in the use of immersion for information visualization," in *Masters. Computer Science*. Blacksburg: Viginia Polytechnic Univ., 2002.

[12] S. M. Drucker and D. Zeltzer, "Intelligent camera control in a virtual environment," in *Proc. Graphics Interface*, Banff, Alberta, Canada, 1994.

[13] J. Drury, J. Scholtz, and H. Yanco, "Awareness in human–robot interactions," in *Proc. IEEE Conf. Syst., Man, Cybern.*, Washington, DC, 2003.

[14] (2005) Epic Games, Unreal Tournament 2003 Game. [Online]. Available: http://www.unrealtournament2003.com

[15] T. Fong and C. Thorpe, "Vehicle teleoperation interfaces," *Auton. Robots*, no. 11, pp. 9–18, 2001.

[16] T. Fong, C. Thorpe, and C. Baur, "Robot, Asker of questions," *Robot. Auton. Syst.*, no. 42, pp. 235–243, 2003.

[17] J. Gabbard and D. Hix, "A taxonomy of usability characteristics in virtual environments," *ONR*, 1997.

[18] A. Hanson and E. Wernert, "Constrained 3-D navigation with 2-D controllers," in *Visualization '97*. Los Alamitos, CA: IEEE Comput. Soc. Press, 1997.

[19] A. Hanson, E. Wernert, and S. Hughes, "Constrained navigation environments," in *Scientific Visualization Dagstuhl '97 Proceedings*, H. Hagen, G. Nielson, and F. Post, Eds: IEEE Computer Society Press, 1999, pp. 95–104.

[20] K. Hinkley, R. Pausch, J. Goble, and N. Kassell, "A survey of design issues in spatial input," in *Proc. ACM Symp. User Interface Software Technol.*, Marina Del Rey, CA, 1994.

[21] D. Hix, E. Swan, J. Gabbard, M. McGee, J. Durbin, and T. King, "User-centered design and evaluation of a real-time battlefield visualization virtual environment," in *Proc. IEEE Virtual Reality*, 1999, pp. 96–103.

[22] S. Hughes and M. Lewis, "Attentive camera navigation in virtual environments," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Nashville, TN, 2000.

[23] ——, "Attentive interaction techniques for searching virtual environments," in *Proc. Human Factors Ergon. Soc. 46th Annu. Meeting*, Baltimore, MD, 2002.

[24] G. Kaminka, M. Veloso, S. Schaffer, C. Sollitto, R. Adobbati, A. Marshall, A. Scholer, and S. Tejada, "Gamebots: A flexible test bed for multiagent team research," *Commun. ACM*, vol. 45, no. 1, pp. 43–45, Jan. 2002.

[25] J. Kay, "STRIPE: Remote driving using limited image data," Ph.D. thesis, Computer Sci. Dept., Carnegie Mellon Univ., 1997.

[26] M. Lewis and J. Jacobson, "Game engines in research," *Commun. ACM*, vol. 45, no. 1, pp. 27–48, 2002.

[27] M. Lewis, K. Sycara, and I. Nourbakhsh, "Developing a testbed for studying human–robot interaction in urban search and rescue," in *Proc. 10th Int. Conf. Human–Comput. Interation*, Crete, Greece, 2003.

[28] M. Masliah and P. Milgram, "Measuring the allocation of control in a 6 degree-of-freedom docking experiment," in *Proc. Conf. Human Factors Comput. Syst.*, The Hague, The Netherlands, 2000.

[29] D. E. McGovern, "Experiences and Results in Teleoperation of Land Vehicles," Sandia Nat. Labs., Albuquerque, NM, Tech. Rep. SAND 90-0299, 1990.

[30] P. Milgram and J. Ballantyne, "Real world teleoperation via virtual environment modeling," in *Proc. Int. Conf. Artif. Reality Tele-Existence*, Tokyo, 1997.

[31] M. Mine, "Virtual Environment Interaction Techniques," Comput. Sci., Univ. North Carolina, Chapel Hill, Rep. TR95-018, 1995.

[32] R. R. Murphy, J. L. Casper, M. J. Micire, and J. Hyams, "Mixed-Initiative Control of Multiple Heterogeneous Robots for Urban Search and Rescue," Univ. Central Florida, Orlando, CRASAR-TR2000-1, 2000.

[33] D. Nieuwenhuisen and M. H. Overmars, *Motion Planning for Camera Movements in Virtual Environments*. Utrecht, The Netherlands: Inform. Comput. Sci. Dept., Utrecht Univ., 2002.

[34] S. Ou, D. R. Karuppiah, A. H. Fagg, and E. Riseman, "An augmented virtual reality interface for assistive monitoring of smart spaces," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2004, pp. 33–42.

[35] T.-M Rhyne, "Computer games' influence on scientific and information visualization," *Computer*, vol. 33, no. 12, pp. 154–159, 2000.

[36] D. S. Tan, G. G. Robertson, and M. Czerwinski, "Exploring 3D navigation: Combining speed-coupled flying with orbiting," in *CHI 2001 Conf. Human Factors Comput. Syst.*, Seattle, WA, 2001.

[37] J. S. Tittle, A. Roesler, and D. D. Woods, "The remote perception problem," in *Human Factors Ergon. Soc. 46th Annu. Meeting*, Baltimore, MD, 2002.

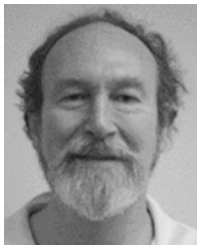
[38] (2005) ULab, USARSim. [Online]. Available: http://usl.sis.pitt.edu/ulab/usarsim_download_page.htm

[39] C. Ware and S. Osborne, "Exploration and virtual camera control in three dimensional environments," in *Proc. Symp. Interactive 3-D Graphics*, 1990, pp. 175–183.

[40] D. D. Woods, J. S. Tittle, M. Feil, and A. Roesler, "Envisioning human–robot coordination in future operations," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 34, no. 2, pp. 210–218, May 2004.

Stephen B. Hughes received the M.S. degree in computer science from Indiana University, Bloomington, in 1997 and is currently pursuing the Ph.D. degree at the University of Pittsburgh, Pittsburgh, PA.

His primary research areas include visual information systems, virtual environments, visual cognition, mixed-initiative systems, and information intermediaries.



Michael Lewis received the Ph.D. degree in engineering psychology from the Georgia Institute of Technology, Atlanta, in 1986.

He is an Associate Professor in the Department of Information Science and Telecommunications, School of Information Sciences, University of Pittsburgh, Pittsburgh, PA. His research has investigated the design of analogical representations, the effectiveness of visual information retrieval interfaces (VIRIs), human–agent interaction, and virtual environments. His current research involves information fusion and human control of mixed-initiative systems.