Robotic Camera Control for Remote Exploration

Stephen Hughes and Michael Lewis
School of Information Science
University of Pittsburgh
135 N. Bellefield Ave
Pittsburgh, PA 15260 USA
+1 412-624-9426
shughes@mail.sis.pitt.edu ml@sis.pitt.edu

ABSTRACT

A video stream from a single camera is often the foundation for situational awareness in teleoperation activities. 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. This paper provides a brief overview of viewpoint control research for 3D virtual environments (VE) to motivate a user study that evaluates the effectiveness of viewpoint controls on a simulated robotic vehicle. Findings suggest that providing a camera that is controlled independently from the orientation of the vehicle may facilitate wayfinding tasks. Moreover, there is evidence to support the use of separate cameras and interfaces for different navigational subtasks

Categories & Subject Descriptors: H5.m. Information interfaces and presentation : Miscellaneous.

General Terms: Design, Experimentation, Human Factors.

Keywords
Robotics, Human-Robot interaction, robotics aided search-and-rescue, remote exploration, viewpoint control

INTRODUCTION

Robotic navigation allows an expendable surrogate 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. The strongest perceptual link to the remote environment often comes through a video feed supplied from a camera mounted on the robot [8]. Poor camera placement, restricted fields-of-view and other camera properties can degrade this link and leave the operator open to a collection of well known operational errors, including disorientation, failure to recognize hazards and simply overlooking relevant information [4, 11].

This research explores camera mounting and control opportunities as design points to promote a meaningful video feed that can mute some of these perceptual obstacles. Specifically, it is hypothesized that multiple cameras, with the option of independent control can mitigate some of problems associated with maintaining situational awareness and can increase the effectiveness of search tasks.

RELATED WORK

Milgram has observed strong parallels between the interaction required to navigate remote and artificial environments [12, 13]. 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 technology [10]. Second, design of interfaces for robotic exploration can draw on the extensive literature of viewpoint control from virtual environments.

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 navigation. Certainly robotic operators can't alter the laws of physics, but by granting the robot autonomy for local navigation, minor obstacles could be avoided, offering a similar interaction experience to the operator.

Navigation in Virtual Environments

Navigation and adjusting the viewpoint in virtual environments have been identified as having a profound impact on all other user tasks [9], and have been the focus of several classification attempts. Tan et. al. observe that most work falls into two categories: efforts to use virtual

CHI 2004, April 24–29, 2004, Vienna, Austria.

environments to understand the cognitive principles of navigation and the design of techniques to facilitate interaction within virtual environments [16]. Adopting Bowman's terminology these segments can be labeled as "wayfinding" and "travel" respectively [3].

Wayfinding represents tactical and strategic elements of defining a path through an environment [5], and is frequently expressed in terms of three subtasks: Exploration, Search and Inspection. The goal of exploration is to advance spatial knowledge of the environment without any specific target. Searches seek to determine and track to the location of a certain object or class of objects. A primed search occurs when the location is known, in contrast to naïve searches where the operator has no a priori knowledge of the target's location. Finally, inspection occurs when a specific viewpoint or series of viewpoints is required for a target object. While these tasks are considered mutually exclusive, they are frequently compounded into sequences [6]. For example, operators engaged in a search may need to transition into inspection to identify a discriminating feature of a potential target. While these wayfinding concepts were presented in virtual environment literature, they extend readily any interaction that requires acquisition and application of spatial knowledge, including robotic control.

Travel is the execution of the wayfinding objectives by manipulating the position (X,Y,Z) and orientation (Yaw, Pitch and Roll) of the viewpoint. Mine provides a taxonomy that divides travel techniques, based on the type of input device, into direct – gestures and body movements; independent – 6 degree of freedom (6DOF) devices; and mapped – controllers with < 6DOF [14]. Each class of controllers has benefits and drawbacks that are beyond the scope of this discussion. Presently mapped controls are the most pervasive, and most likely to be used by robotic operators, so they will capture the focus of our discussion.

MAPPED CONTROLS

Mapped control techniques can rely on input devices that are readily available and familiar, such as joysticks and mice. However, these interfaces must attempt to manipulate 6DOF using a device that has inherently fewer control options. Reviewing the literature, we identify four techniques for mapping controls that are frequently used in virtual environments.

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 CAD and modeling software, the increased cognitive burden of remembering the current state can negatively impact performance [2].

Constraining – Movement of the viewpoint is limited to certain operations; manipulations of other attributes are simply discarded. The most common example of constraining is to restrict motion to a ground plane, eliminating the need for vertical translation [18]. 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 known as gaze-directed steering, in which the viewer's motion is determined by the direction they are looking [3].

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

These four techniques are not exclusive; in fact often some combination is required to bring the control space from 6DOF to 2DOF.

The preceding discussion provides an organizational overview for VE travel techniques. It is our belief that the potential for mode errors outweighs the benefit of extra degrees of freedom offered by overloaded techniques. We also hold that constraining viewpoint control may be useful as a function of the environment and the robot's physiology, but should not be employed as an arbitrary design decision. For example, as noted previously, an argument can be made for constraining control of a ground vehicle to 4DOF, which could be managed by two independent controllers. Coupling also has practical connections to robotic control interfaces and will be analyzed as part of a user study described below. Offloading also holds great promise for robotic exploration, especially in light of the push for the design of more supervisory and autonomous systems. This topic will be revisited as part of the future directions section.

TECHNIQUES CONSIDERED

McGovern provides accounts of robotic systems that include cameras that are dependent on the steering mechanism, a independently controlled camera, and multiple fixed cameras [11]. Each of these techniques was considered for evaluation and is described in more detail below.

Coupled Camera Controls

Mounting a fixed camera on the front of a robot yields the equivalent of the popular gaze-directed steering interface described above. This approach has become one of the most pervasive forms of control for virtual environments, perhaps because of its intuitive nature. To navigate a ground vehicle, 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), much like driving a car. However, the ease of travel may come at the expense of all but the most trivial inspections. 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 a good view of the object, and then pivot back to resume motion. Knowing when to turn to face the object requires that the controller have a good sense of the both the overall configuration and scale of the environment. For robotic exploration applications, it is unlikely that either of these will be the case. Moreover, there is no guarantee that the object of interest even remains in the field of view, further increasing the chances that useful viewpoints may be overlooked or missed. In addition to the cognitive burdens that the coupled approach will likely introduce, there is also the problem of making repeated 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.

Independent control

Allowing for an independently controlled camera with constraints on elevation and roll reduces the control space to 4DOF. 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 virtual environments shun this technique for just that reason. Baker and Wickens [1] 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". Simple travel operations such as "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. Furthermore, without a very good understanding of the environment, it would be ill advised to spend much time navigating moving in a direction other than where the camera is facing. Fortunately, independent controllers have the option of realigning the direction of gaze and direction of motion when performing any extensive travel activities. However, this may factor into the "complexity of the control problem", referenced above by Baker.

The ability to assess the angular displacement of the camera is critical to situational awareness. One way to achieve this is by mounting the camera in a position where the body of the robot is visible in the periphery of the viewpoint. The orientation of the camera may be discerned by identifying unique features associated with the front, sides or rear of the robot. However, It is unclear if these ecological cues provide the operator with enough insight to the degree of displacement. Numerous other studies have evaluated the effectiveness of various instruments to assist with spatial cognition including: you-are-here maps, compasses, trails, viewtracks, etc. [6, 17]. To track displacement between the orientation of the robot and an independently controlled camera, a two-handed compass was developed. Pictured in Figure 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).

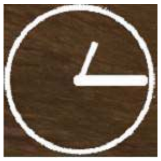

Figure 1: Two Handed Compass

Multiple Cameras

The prospect of equipping teleoperated robots with multiple cameras is frequently raised to support stereopsis. In these scenarios, two cameras are focused in 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 [15].

Two cameras, mounted on the same robot may also be used to align with the subtasks of inspection and search/exploration to further reduce the disruption of task-switching. A fixed screen, coupled with the orientation of the robot would be used for searching, while the controllable camera could be manipulated for inspection. Switching tasks would simply be a matter of selecting which feed requires attention. The cognitive demand could be reduced from understanding the robot in the context of an unfamiliar environment to simply remembering the state of the robot (i.e "the inspection screen is set to look off about 30° to the right".)

USER STUDY

A user evaluation was conducted to assess the impact of these three camera control variations on wayfinding tasks in a simulated teleoperation environment, resulting in five conditions:

- Single Fixed Camera, No Instrumentation
- Single Independent Camera , No Instrumentation
- Single Independent Camera, 2-handed Compass
- Multiple Cameras, No Instrumentation
- Multiple Cameras, 2-handed Compass.

Each of these conditions were implemented using the simulated four-wheeled Urban Search and Rescue robot described by Lewis, Sycara, and Nourbakhsh [10]. Figure 2 shows a schematic of the simulation architecture. The bulk of the simulation is handled by Epic Games' Unreal Tournament (UT) Game Engine [7], including structural modeling for the robot and the environment and the physics of their interaction. Modifications are made to the UT interface through the GameBots API which allows programmatic control of the UT actors. Finally, attaching a UT spectator to the robot enabled the second, independently controlled camera.

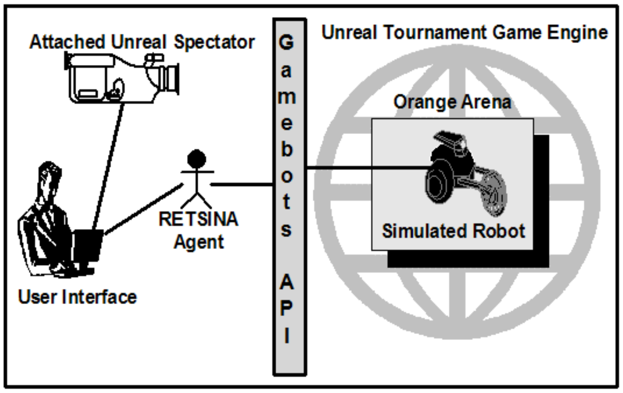

Figure 2: Architecture of Simulation

Participants

65 men and women were paid $15 to evaluate five camera control strategies (13 per condition). Participants were recruited from the University of Pittsburgh community, with most subjects enrolled as undergraduates. 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 based on lack of computer proficiency interfering with adequate completion of the task.

Design and Procedure

Subjects were asked to navigate a non-trivial environment for fifteen minutes 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 design forces the explorers to:

- Locate an object from a distance
- Position the robot nearer the potential target
- 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. All subjects were required to confirm an understanding of the task and the controls by identifying at least one target object in a training environment.

The experiment was a repeated-measures design and two separate environments were used to counterbalance the effects of the technique. The first environment (shown in Figure 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. Upon entering, rooms could be inspected with a quick survey to determine if they contained a target that required more attention.

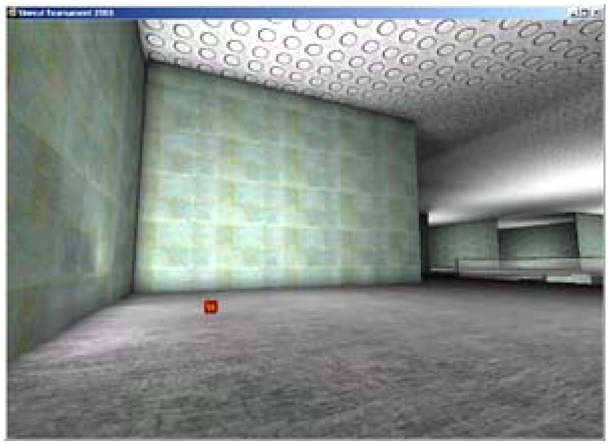

Figure 3: Screenshot of indoor environment.

The second environment resembled a more rugged outdoor environment with characteristics of a canyon or desert (Shown in Figure 4). 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. Participants were advised that a good strategy might be to survey the scene from a high elevation. 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. Additionally, the second environment was much more expansive than the first (about 4 times the land area). Success in this environment required coverage of more terrain rather than intricate navigation.

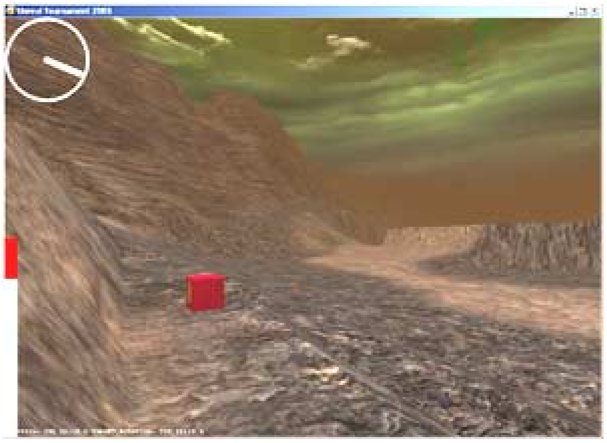

Figure 4: Screenshot of outdoor environment

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). Entries were written to the log file nineteen times per second, allowing for a complete reconstruction of each session.

Apparatus

The robot was controlled using a Logitech Extreme digital 3D 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 800x600 resolution. For the 2-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.

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.

- Across all conditions, significantly more objects were found in the indoor environment (mean 7.2) than the outdoor environment (mean 4.0, t(127) = 8.78). 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.
- The two-handed compass did not produce a significant difference in any of the independent trials.

As a result of these findings, the data was pooled for the following analysis: comparisons were made between coupled, 1-camera, and 2-camera conditions and within the indoor and outdoor trials. Figure 5shows that both independent conditions outperformed the coupled condition in terms of the number of markers identified. The statistical figures are presented in Table 1.

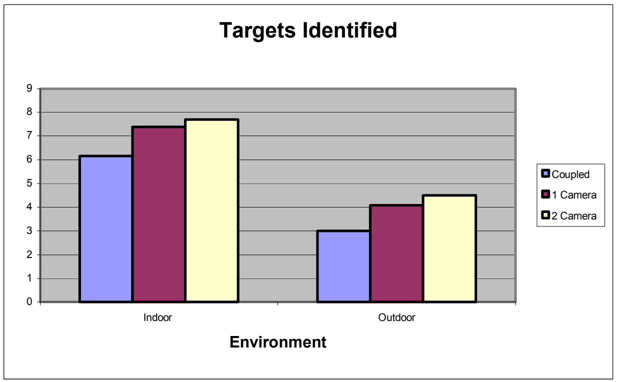

Figure 5

Table 1
Differences in number of targets identified

	Indoor	Outdoor
Coupled, 1-Camera	T(37) = 1.75, p <.05	t(36) = 2.00, p < .05
Coupled, 2-Camera	T(37) = 1.98, p <.05	t(37) = 2.39, p < .05
1-Camera, 2-Camera	T(50) = 0.48	t(49) = 0.76

This result is further supported by an analysis of the uses of the independent cameras. Recall that panning the camera is left to the discretion of the viewer; if the controller opts to not exercise the option of independently panning the camera, the control effectively degenerates into the coupled 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 strong correlation was found between the amount of time that the controller was disjoint and the number of markers found (1 camera: N=50, mean disjoint time ≈ 6:20, ρ = 0.41, 2 camera: N=52 mean disjoint time ≈ 10:30, ρ = 0.45). Controllers who did not avail themselves of the independent camera control did not perform as well as those that exercised that option.

Although there were no differences detected in the effectiveness of the 1-camera and 2-camera 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:

- Pan Time – The number of ticks that recorded a differential yaw value for the independent camera.
- Disjoint Time – The number of ticks where the orientation of the camera varied from the orientation of the robot in excess of 10°.
- Disjoint Motion – Disjoint time when the robot was also moving.
- Idle Disjoint time – Disjoint time where the robot is neither panning the camera nor moving the robot.
- Re-coupling – 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 differences between the indoor and outdoor conditions, suggesting that individuals essentially controlled the robot in a similar manner regardless of the environment.

Figure 6 shows that the 2-camera condition spent almost twice as much time disjoint than the 1-camera conditions. This result was significant for both disjoint motion and idle disjoint times, t(100) = 7.40, p <.01 and t(100) = 3.33, p <.01.

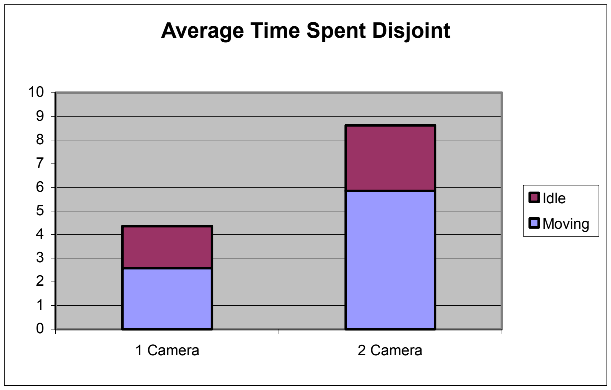

Figure 6

This does not mean that these users 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 1-camera robots were not afforded this luxury and were therefore more likely to re-couple the independent camera with the orientation of the robot in order to comprehend their direction of travel for large scale movements (1-Cam mean: 87 re-couples 2-Cam mean: 62 re-couples, t(100) = 3.98 p < .01).

Finally, we analyze the effect of dividing attention across the two video feeds. The 2-Camera +Compass condition recorded fewer pans than the 1-camera condition (t(74) = 1.67, p<.05), or the 2-Camera, No Instrumentation condition (t(50) = 2.11, p <.05) as shown in Figure **7**. This result suggests that the operators of the 2-camera conditions were not maintaining the state of the independent camera when they were not attending to it. The operators with the compass used it to reorient themselves when they returned their attention to the independent camera, while the operators who had no instrumentation could have been using additional panning motions to reestablish their situational awareness.

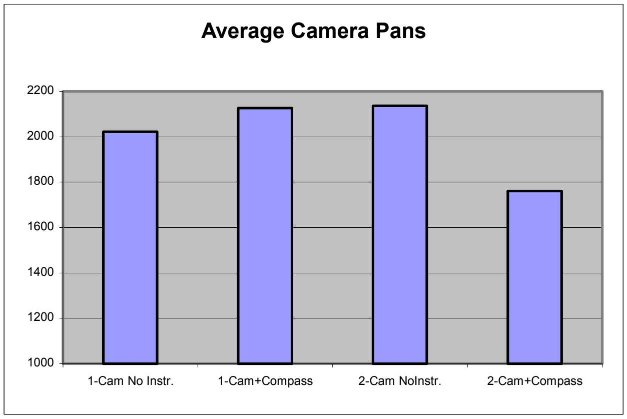

Figure 7

IMPLICATIONS AND FUTURE DIRECTIONS

The data collected from this experiment suggest that the use of an independent, controllable camera increase 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. In the long run 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 2-camera display may find that one of the screens is more dominant, suggesting that a screen-in-screen technique may be appropriate.

Finally, the parity of the 2-camera display offers some interesting opportunities for offloading control of the camera to an autonomous agent. Having both the human and the robot battle for control of the cameras would likely be disruptive to the point that neither would accomplish much. However, we speculate that the 2-screen approach might allow for a more cooperative collaboration, where one screen may represent human control, while the second reflects the offloaded actions of an autonomous agent.

ACKNOWLEDGMENTS

This research was supported by AFOSR contract number F49640-01-1-0542. The authors would also like to express their appreciation to Jijun Wang, Joeseph Manojlovich and the members the teleoperation seminar who assisted with the logistics of conducting this study.

REFERENCES

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

[2] Beaten, R., et al. *An Evaluation of Input Devices for 3-D Computer Display Workstations*. in *Proc. of SPIE-The International Society for Optical Engineering*. 1987.

[3] Bowman, D., D. Koller, and L. Hodges, *A Methodology for the Evaluation of Travel Techniques for Immersive Virtual Environments.* Virtual Reality: Research, Development and Applications, 1998. **3**(2): p. 120-131.

[4] Darken, R., K. Kempster, and B. Peterson. *Effects of Streaming Video Quality of Service on Spatial Comprehension in a Reconnaissance Task*. in *Proceedings of the meeting of I/ITSEC*. 2001.

[5] Darken, R. and B. Peterson, *Spatial Orientation, Wayfinding and Representation.*, in *Handbook of Virtual Environment Technology*, K. Stanney, Editor. 2001, Lawrence Erlbaum Associates: Mahway, NJ.

[6] Darken, R. and J.L. Siebert. *Wayfinding Strategies and Behaviors in Large Virtual Worlds*. in *ACM SIGCHI 96*. 1996.

[7] Epic Games, I. *Unreal Tournament 2003 Game*. http://www.unrealtournament2003.com

[8] Fong, T. and C. Thorpe, *Vehicle Teleoperation Interfaces.* Autonomous Robots, 2001(11): p. 9-18.

[9] Hix, D., et al. *User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment*. in *IEEE Virtual Reality '99*. 1999.

[10] Lewis, M., K. Sycara, and I. Nourbakhsh. *Developing a Testbed for Studying Human-Robot Interaction in Urban Search and Rescue*. in *10th International Conference on Human Computer Interation (HCII '03)*. 2003. Crete, Greece.

[11] McGovern, D.E., *Experiences and Results in Teleoperation of Land Vehicles*. 1990, Sandia National Laboratories: Albuquerque, NM.

[12] Milgram, P. and J. Ballantyne. *Real World Teleoperation via Virtual Environment Modelling*. in *International Conference on Artificial Reality & Tele-existence*. 1997. Tokyo.

[13] Milgram, P. and H. Colquhoun, *A Taxonomy of Real and Virtual World Display Integration*, in *Mixed Reality - Merging Real and Virtual Worlds.*, Y.O.a.H. Tamura, Editor. 1999, Springer Verlag: Berlin. p. 1-16.

[14] Mine, M., *Virtual Environment Interaction Techniques*. 1995, UNC Chapel Hill Computer Science Technical Report TR95-018.

[15] Murphy, R.R., et al. *Mixed-Initiative Control of Multiple Heterogeneous Robots for Urban Search and Rescue*. www.csee.usf.edu/robotics/Publications/

[16] Tan, D.S., G.G. Robertson, and M. Czerwinski. *Exploring 3D Navigation: Combining Speed-coupled flying with orbiting*. in *CHI 2001 Conference on Human Factors in Computing Systems*. 2001. Seattle, WA.

[17] Tittle, J.S., A. Roesler, and D.D. Woods. *The Remote Perception Problem*. in *Human Factors and Ergonomics Society 46th annual meeting*. 2002. Baltimore, MD.

[18] Ware, C. and S. Osborne. *Exploration and Virtual Camera Control in Three Dimensional Environments*. in *Proceedings of the 1990 Symposium on Interactive 3D Graphics*. 1990.