

Camera Control and Decoupled Motion for Teleoperation

Stephen Hughes, Joseph Manojlovich, Michael Lewis, and Jeffrey Gennari

School of Information Science

University of Pittsburgh

Pittsburgh, PA USA

{shughes, josephm, ml, jgennari}@mail.sis.pitt.edu

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. This limitation on the perceptual links to the environment leaves the operator prone to 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 an experiment to 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.*

Keywords: Camera Controls, Teleoperation, Multiple Displays, Viewpoint control, Decoupled Viewpoints.

1 Introduction

Teleoperated robotic vehicles allow for safe exploration and inspection of hazardous environments with replaceable (albeit expensive) surrogates for humans. This promise has motivated research from several fields of study including: extra-planetary exploration [10], maintenance of nuclear facilities [3], military operations [6] or search-and-rescue efforts([9]). Although researchers have been flirting with granting various levels of autonomy to robotic explorers [3, 9], human supervision, judgment and intervention remain critical elements of these operations. Fong and Thorpe [6] characterize several types of interfaces used to keep humans involved in robotic activities, but observe that the most common interaction remains direct control while watching a video feed from vehicle mounted cameras. Part of the appeal for this technique is likely because of the similarity to control of traditionally piloted vehicles. In fact the same layout for the controls is often used, and the video feed can be projected in place of a windshield to maximize familiarity.

Despite the popularity of direct control, this approach is not without fault. Perception of the remote environment is limited to the visual channel and is often restricted to a very narrow field of view. A breakdown

of perceptual modalities, lack of vestibular and proprioceptive cues, and lack of direct interaction with the environment significantly add to the problems of remote perception [12]. Impairment at the perceptual level commonly leads to well-known, cognitive failures, including degradation of situational awareness, incorrect manipulation of the vehicle's attitude (pitch and roll), and failure to take precaution against obstacles [8]. Situational awareness is particularly critical and the problem of getting lost while relying on video feeds was most recently documented in [4].

Accepting that remote perception will never match direct perception, the objective of teleoperated systems should be *functional presence* [12]. This occurs when the operator receives enough cues to successfully conduct operations, without requiring the sense that they actually are situated at the remote location (as with the conventional understanding of the term presence). Designers who wish to balance the appeal of direct control with the complications of remote perception should consider the range of techniques that fall under the umbrella of direct control. While the video feed offers a common mechanism for supporting situational awareness, there is appreciable variability in the strategies for its generation. For example, McGovern provides accounts of systems that include single fixed camera, multiple fixed camera, and cameras that are dependent on the steering mechanism [8].

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

2 Considerations for Camera Control

2.1 A task driven approach

The video stream offers a window into the remote environment. The mechanisms provided to the controller 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 [11]. 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 viewpoint – or set of viewpoints – for a particular object. While both navigation and inspection may 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. For example, when navigating the operator may consider moving “over there”, but inspection is motivated by questions like “What does this object look like from the side?”

It is our belief that switching between these two subtasks is a major contributor to problems with situational awareness 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.

2.2 Camera Configurations

One of the most basic decisions about camera control is whether the operator has independent control over the orientation of the camera or if the orientation remains fixed with respect to the orientation of the vehicle.

Coupled Camera Controls – When the camera orientation is fixed, the direction of the gaze is directly dependent on the direction of travel. This approach is commonly known to the Virtual Reality (VR) community as *Gaze-directed steering* [2]. Such coupled camera controls offer a strong bias toward navigation tasks at the expense of all but the most trivial inspections. 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 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 the applications described in the introduction it is

unlikely that either of these are 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.

Decoupled Camera Controls – Allowing for an independently controlled camera decouples the direction of gaze from the direction of travel. Surprisingly, the VR community has largely shunned such decoupled camera controls. 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.[1]”. While decoupling the camera facilitates inspection, by allowing the controller to keep interesting objects in view, navigation suffers. For example, instructions to “Move Forward” may be ambiguous unless the viewer has a good understanding of which direction is forward relative to the direction that the camera is currently oriented. Furthermore, without a very good understanding of the environment, it would be ill advised to spend much time navigating while decoupled, for the simple reason that the operator may not be able to see where they are going. Fortunately, decoupled controllers have the option of realigning the direction of gaze and direction of motion when performing any extensive travel activities. However, this factors 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 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 [9].

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 would 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. 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”).

2.3 Ecological Cues vs. Instrumentation

Assuming that 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 peripheral fixed references may provide the operator with some insight to the degree of displacement.

Another area of interest for this experiment was to understand if these ecological cues are sufficient to understand the displacement induced by decoupling. Generally the body of the robot was visible to provide a fixed reference to the orientation of the viewing camera. The front, back and sides were distinct enough to provide a cue to the direction the camera was facing. However, as the pitch of the camera was significantly adjusted, the effectiveness of these cues may be diminished.

Numerous other studies have evaluated the effectiveness of various instruments to assist with spatial cognition including: you-are-here maps, compasses, trails, viewtracks, etc.[5, 12]. While these have met with varying degrees of success, they don't explicitly address the issues of resolving decoupled navigation. For this experiment, a two-handed compass was devised to reflect the orientations of both the vehicle and the independent camera. If the view were aligned with the front of the robot, a single line would be seen. As the independent camera pans to the side, a second line becomes visible creating an angle that corresponds to the magnitude of the displacement.

3 User Evaluation

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

- Coupled Motion, No Instrumentation, 1 Camera
- Decoupled Motion, No Instrumentation, 1 Camera
- Decoupled Motion, 2-handed Compass, 1 Camera
- Decoupled Motion, No Instrumentation, 2 Cameras
- Decoupled Motion, 2-handed Compass, 2 Cameras

Each of these conditions were implemented using the simulated USAR robot described by Lewis, Sycara, and Nourbakhsh [7]

3.1 Participants

65 men and women were paid \$15 to evaluate five camera control strategies (13 per condition). Participants were recruited from the general community of the University of Pittsburgh, with most subjects enrolled as undergraduates. One subject terminated the experiment prior to completing the outdoor condition, but data were still included for the indoor trial. Three additional subjects were excluded from the study based on lack of computer proficiency interfering with adequate completion of the task.

3.2 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 may 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 Fig. 1) 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.

The second environment resembled a more rugged outdoor environment with characteristics of a canyon or desert (Shown in Fig. 2). Unlike the first environment, target objects could be obscured by irregularities in the relief of the terrain; small craters or ridges may conceal a target unless it is 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.

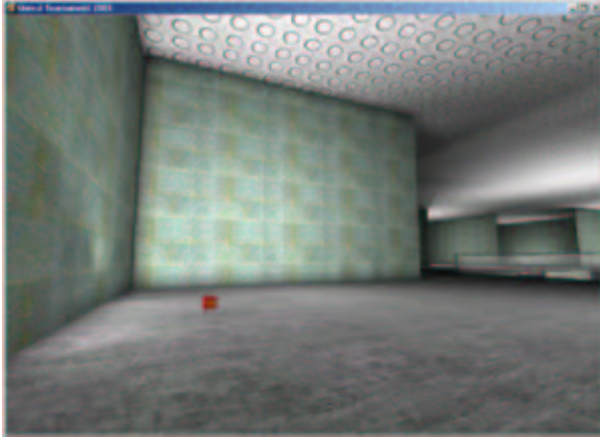


Figure 1: Screenshot of indoor environment.



Figure 2. 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.

3.3 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.

4 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 decoupled 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 3 shows that both decoupled conditions outperformed the coupled condition in terms of the number of markers identified. The statistical figures are presented in Table 1.

This result is further supported by an analysis of the operation of the decoupled techniques. 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 \approx 6:20, $\rho = 0.41$, 2 camera: N=52 mean disjoint time \approx 10:30, $\rho = 0.45$). Controllers who did not avail themselves of the decoupled option did not perform as well as those that exercised that option.

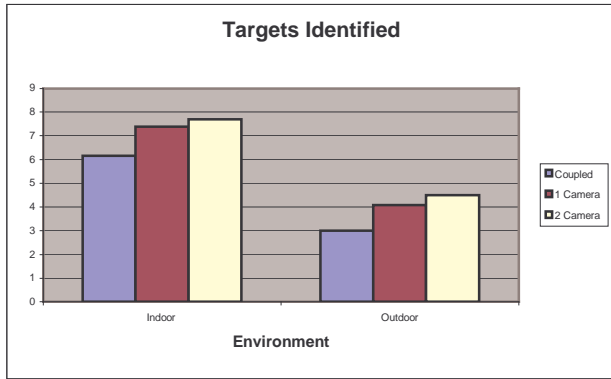


Figure 3

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	$t(50) = 0.48$	$t(49) = 0.76$

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.
- 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 differences between the indoor and outdoor conditions, suggesting that individuals essentially controlled the robot in a similar manner regardless of the environment.

Figure 4 shows that there were no significant differences between the time spent panning ($t(100) = .9$), but the 2-camera condition produced substantially more disjoint motion and idle disjoint times, $t(100) = 7.40, p < .01$ and $t(100) = 3.33, p < .01$.

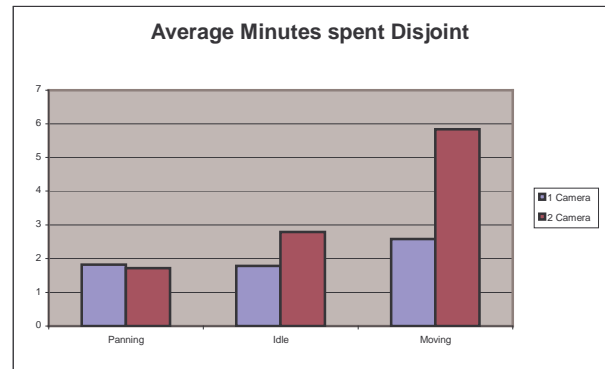


Figure 4.

This data suggests that while both groups did roughly the same amount of panning, the participants in the 2-camera condition were more likely to move with the camera disjoint from the orientation of the robot. This does not mean that these users were better able to resolve the ambiguity of decoupled motion. Instead this result probably reflects the operator shifting their attention to the view with the fixed camera, 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 recouple the independent camera with the orientation of the robot in order to comprehend their direction of travel (1-Cam mean: 87 recouples 2-Cam mean: 62 recouples, $t(100) = 3.98, p < .01$)

Given that the 1-camera condition is known to have engaged in additional panning activities – to realign the cameras – it is curious that the 2-camera condition seems to produce the same number of pans. A couple of theories might provide an explanation.

1. The effort to realign the cameras is trivial, and did not comprise a significant portion of the overall panning activity
2. Participants in the 2-camera condition could have actually searched the environment more thoroughly. However, without effective guidance on where to look for the targets, the extra panning did not translate into a higher number of targets identified.

3. After performing a navigation sub-task, memory for the orientation of the independent camera may have decayed. Additional panning may have been necessary to re-establish situational awareness when transitioning back to an inspection task.

5 Conclusions

The data collected from this experiment suggest that the use of a decoupled, 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 decoupled techniques that were examined did not show quantitative differences, 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 Collaborative Control of the robots. Advances in sensor technology may allow the robots to direct the operator to optimal viewing locations. Alternatively, the robot may need to direct the human to particular viewpoint to help it resolve some ambiguity [6]. 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, the 2-screen approach might allow for a more cooperative collaboration, where one screen represents human control, while the second screen is sensor-driven.

Acknowledgements

This research was supported by AFOSR contract number F49640-01-1-0542. The authors would also like to express their appreciation to the members of IS3954 who assisted with the logistics of conducting this study.

6 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] Bowman, D., D. Koller, and L. Hodges. *Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques*. in *Virtual Reality Annual International Symposium*. 1997.
- [3] Bruemmer, D.J., et al. *Mixed-Initiative Control for Remote Characterization of Hazardous Environments*. in *HICSS*. 2003. Waikoloa Village, HI.
- [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 IITSEC*. 2001.
- [5] Darken, R. and J.L. Siebert. *A Toolset for Navigation in Virtual Environments*. in *UIST '93*. 1993.
- [6] Fong, T. and C. Thorpe, *Vehicle Teleoperation Interfaces*. *Autonomous Robots*, 2001(11): p. 9-18.
- [7] 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 Interaction (HCI '03)*. 2003. Crete, Greece.
- [8] McGovern, D.E., *Experiences and Results in Teleoperation of Land Vehicles*. 1990, Sandia National Laboratories: Albuquerque, NM.
- [9] Murphy, R.R., et al. *Mixed-Initiative Control of Multiple Heterogeneous Robots for Urban Search and Rescue*.
www.csee.usf.edu/robotics/Publications/
- [10] Nguyen, L.A., et al., *Virtual Reality Interfaces for Visualization and Control of Remote Vehicles*. *Autonomous Robots*, 2001(11): p. 59-68.
- [11] 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.
- [12] 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.