Assisted Viewpoint Control for Tele-Robotic Search

Stephen Hughes and Michael Lewis
University of Pittsburgh
Pittsburgh, PA 15260
{shughes, ml}@mail.sis.pitt.edu

Mixed-Initiative robotic systems have been successful with providing assistance for maneuvering a robot's position through a hazardous environment. However, providing assistance to resolve viewing parameters such as camera orientation have met with mixed results. Previous attempts, while effective at increasing the likelihood that key viewpoints are utilized, have met with strong dissatisfaction and struggles for control, overriding the some of the potential benefits of the automation. This study evaluates the effects of offloading the assistance to a separate, co-located video feed. The findings suggest that it is possible to provide assistance to viewpoint control without encroaching on the operator's threshold for desired control.

INTRODUCTION

Maneuvering a remote viewpoint has become a common activity, spanning multiple disciplines including extra-planetary exploration, maintenance of nuclear facilities, military operations or search-and-rescue efforts (Murphy, Casper et al.). In fact it can be argued that the same skill set applies to most teleoperation, scientific visualization and virtual reality interfaces (Milgram and Colquhoun 1999).

Human operators are capable adjusting the flow of visual information by simultaneously manipulating multiple viewing parameters. However, success is often dependent on adopting sophisticated strategies such as acquiring survey views, or moving is structured patterns (Bowman 1999). Even with these strategies, there is the chance that the effort applied to manipulating the viewpoint will distract from extracting the desired information.

One way to facilitate viewpoint control is to develop and refine methods for mapping control metaphors to operational tasks. Previous work has shown that gaze-directed steering (binding the orientation of the view to the direction of travel) can effectively reduce the cognitive load of navigation. However, if motion needs to occur relative to a particular object, rather than a specified direction, gaze-directed steering is not appropraite (Bowman 1999). Unfortunately, viewers frequently engage in both kinds of motion over the course of exploring an environment (Darken and Siebert 1996). For example, operators engaged in a direction-based search may need to transition into object-based inspection to identify a discriminating feature of a potential target.

Hughes (2004) proposed that these navigational subtasks could be successfully split across video feeds from two separate cameras. A fixed, forward-facing camera effectively implements gaze-directed steering and allows the operator to understand the heading of the robot. The second screen allows the viewer to manipulate pan and tilt controls to alter the orientation of a different, co-located camera. This view affords a rapid inspection of the environment, allowing the gaze direction to be independent of robot heading. If these views are simultaneously presented, the operator needs only to choose to attend to the screen that contains the information relevant to his current subtask. While this approach offers seamless transition between navigation subtasks, the operator may still be overwhelmed by the numerous parameters that need to be maintained.

Automation of some of the viewpoint controls 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 in offloading some of the viewpoint parameters to the automated system, allowing the human operators to focus their efforts on perceptual judgments and decision making that exceed the current capability of automation (Fong and Thorpe 2001).

Automated Tasks

Studies of how people interact with 6 Degree-of-Freedom devices reveal that there is a division between interaction with translation and orientation controls (Hinkley, Pausch et al. 1994; Masliah and Milgram 2000). 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 pre-programmed sequence of steps through the environment, offering a generic tour. Generalizing this approach, the viewer may be able to specify a set of interests. Based on this input and other constraints such as obstacles in the environment, it is possible to generate a more personalized tour (Drucker and Zeltzer 1994). 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, radial search, where exploration progresses in increasing concentric circles or contour following (Darken and Siebert 1996).

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 must be supported to effectively implement automatic gaze redirection: shifting and holding (Bajscy, Kosecka et al. 1995). 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 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 (Nieuwenhuisen and Overmars 2002)

Previous mixed-initiative robotic systems have emphasized positioning operations and path-planning for the robot (Fong and Thorpe 2001). Bruemmer, for instance, promotes granting the robot the ability to "veto dangerous human commands to avoid running into obstacles or tipping itself over" (2003). 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. Poor camera orientation can leave the operator open to a collection of well known operational errors, including disorientation, failure to recognize hazards and simply overlooking relevant information (McGovern 1990; Darken, Kempster et al. 2001).

While guided orientation has not played a prominent role in the robotics literature, the virtual environments literature offers some insight to this issue. Constrained Navigation and the Attentive Camera are two approaches that promote a supportive, yet unscripted explorations (Hanson and Wernert 1997). Using these techniques, the orientation of the camera can be systematically redirected to relevant features based on a viewer-determined location in the environment. Subsequent studies of these techniques have revealed significant benefits including: increased the likelihood that key viewpoints are utilized (Hanson, Wernert et al. 1999); better understanding the presence and configuration of key elements (Hughes and Lewis 2000); and reduced search time to locate key elements (Hughes and Lewis 2002).

Despite the clinical successes of the Attentive Camera, anecdotal feedback indicated strong dissatisfaction with automatic re-orientation 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 (Hughes and Lewis 2002)

This research proposes that adopting the two-camera approach can be beneficial in resolving some of the coordination issues associated with guided orientation systems. Specifically, we hypothesize that the major obstacle to guided orientation systems is that the gaze redirection interferes with the operator's ability to understand the heading of the robot. By maintaining the heading in a separate camera view, operators will be able to position the robot with greater ease and they will be less likely to override the automation. These two factors will allow the operator to more successfully extract information from the environment.

EXPERIMENTAL EVALUATION

Procedure

Participants were asked to navigate a non-trivial environment for fifteen minutes with the task of locating and identifying as many target objects as possible. Twelve targets were evenly distributed throughout the environment and were identified on two levels of detail. Objects were to be initially identified by class and then confirmed by a discriminating feature. Specifically, targets consisted of a red cube marked on one side with a yellow, capital letter. Identification of a target was considered complete when the viewer could accurately name the letter marked on the cube. Participants were not aware of how many target objects were in the environment, however, they were advised that not all letters of the alphabet would be represented or repeated, 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 or completely abutted to a wall).

The 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 that would cover the entire space. The closed layout meant that targets were generally not visible from a distance; navigation to each room was necessary to verify its contents. Aside from the target objects, the environment was void of non-architectural features.

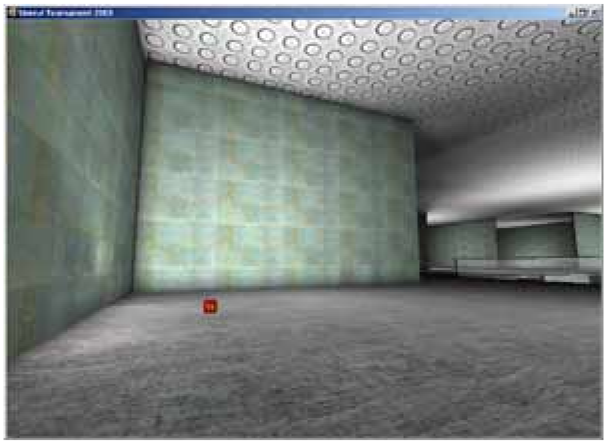

Figure 1: Exploration environment

The experimental treatment varied according to whether or not orientation assistance was provided, yielding two conditions.

User-controlled orientation (Unassisted): 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. This view affords a rapid inspection of the environment without altering the heading of the robot.

Sensor-driven orientation (Assisted): Again 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.

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. 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 whenever the operator issued a command, allowing for a complete reconstruction of the participants movement.

Participants

26 undergraduate students were recruited from the University of Pittsburgh and assigned to either the assisted or unassisted condition. One subject from the assisted condition had to be excluded due to a corrupt data file. Upon completion, participants were compensated $15 for their involvement in this study. All participants self-reported normal or corrected-to-normal color vision.

Apparatus

The robot simulation was implemented using the architecture described by Lewis, Sycara, and Nourbakhsh (Lewis, Sycara et al.). Figure 2 shows a schematic of the simulation architecture. The bulk of the simulation is handled by Epic Games' Unreal Tournament (UT) Game Engine (2003), 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 (2002) which allows programmatic control of the UT actors. Finally, attaching a UT spectator to the robot via a TCP/IP network enabled the second, independently controlled camera.

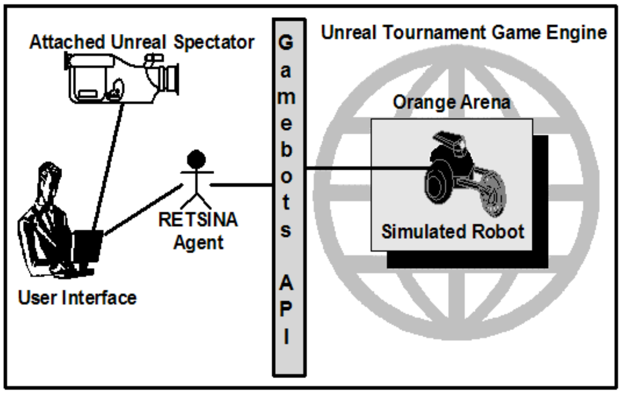

Figure 2: Architecture of Simulation

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. 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 two 21" monitors using 800x600 resolution.

RESULTS

Data were first analyzed with regard to the number of targets successfully identified. The sensor-driven orientation condition consistently identified more targets (mean 9.1) than the user-controlled condition (mean 7.5) $t(23) = 1.93, p < .05$, 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. Recall that target identification was a two-step process: 1) locate the target 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 seconds less under the sensor-driven condition: $t(11) = 3.40, p < .01$. 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 ($t(23) = 2.26, p <.05$), despite issuing roughly the same number of commands overall.

CONCLUSION

The results of this study 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 that it has found something interesting, and would like the operator to take a look (Fong, Thorpe et al. 2003).

Unlike previous studies, the operators did not seem to struggle to maintain control of the viewpoint. This was supported by the analysis of the movement logs as well as anecdotal responses at the conclusion of the experience. 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 intrusive and therefore errors caused by a lack of coordination can be mitigated.

Finally, there are 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. Since the primary benefit observed stemmed from tracking an object during inspection, it may be possible to achieve similar benefits from a user-activated tracking assistance system with a single camera. However, further testing would be required to validate this approach.

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.

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, Joseph Manojlovich who assisted with the logistics of conducting this study.

REFERENCES

Bajscy, R., J. Kosecka and H. Christiansen (1995). "Discrete Event Modeling of Navigation and Gaze Control." International Journal of Computer Vision, Special Issue on Qualitative Vision **14**(2): 179-191.

Bowman, D. (1999). Interaction Techniques for Common Tasks in Immersive Virtual Environments: Design, Evaluation and Application. Computer Science, Georgia Institute of Technology.

Bruemmer, D. J., J. L. Marble, D. D. Dudenhoeffer, M. O. Anderson and M. D. McKay (2003). Mixed-Initiative

Control for Remote Characterization of Hazardous Environments. HICSS, Waikoloa Village, HI.

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

Darken, R. and J. L. Siebert (1996). Wayfinding Strategies and Behaviors in Large Virtual Worlds. ACM CHI 1996 Conference on Human Factors in Computing Systems.

Drucker, S. M. and D. Zeltzer (1994). Intelligent Camera Control in a Virtual Environment. Proceedings of Graphics Interface '94, Banff, Alberta, Canada.

Epic Games (2003) Unreal Tournament 2003 Game http://www.unrealtournament2003.com Accessed: 2003

Fong, T. and C. Thorpe (2001). "Vehicle Teleoperation Interfaces." Autonomous Robots(11): 9-18.

Fong, T., C. Thorpe and C. Baur (2003). "Robot, Asker of Questions." Robotics and Autonomous Systems(42): 235-243.

Gamebots (2002) http://www.cs.cmu.edu/~galk/GameBots/WEB/docapi.html Accessed: 2003

Hanson, A. and E. Wernert (1997). Constrained 3D Navigation with 2D controllers. Visualization '97, IEEE Computer Society Press.

Hanson, A., E. Wernert and S. Hughes (1999). Constrained Navigation Environments. Scientific Visualization: Dagstuhl '97 Proceedings. H. Hagen, G. M. Nielson and F. Post, IEEE Computer Society Press**:** 95-104.

Hinkley, K., R. Pausch, J. Goble and N. Kassell (1994). A survey of design issues in spatial input. Proceedings of ACM Symposium on User Interface Software and Technology (UIST), Marina Del Rey, CA.

Hughes, S. and M. Lewis (2000). Attentive Camera Navigation in Virtual Environments. IEEE International Conference on Systems, Man & Cybernetics, Nashville, TN.

Hughes, S. and M. Lewis (2002). Attentive Interaction Techniques for Searching Virtual Environments. Human Factors and Ergonomics Society's 46th Annual Meeting, Baltimore, MD.

Hughes, S. and M. Lewis (2004). Robotic Camera Control for Remote Exploration. ACM CHI 2004 Conference on Human Factors in Computing Systems, Vienna, Austria.

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

Masliah, M. and P. Milgram (2000). Measuring the allocation of control in a 6 degree-of-freedom docking experiment. Conference on Human Factors in Computing Systems, The Hague, The Netherlands.

McGovern, D. E. (1990) Experiences and Results in Teleoperation of Land Vehicles, Sandia National Laboratories: SAND 90-0299

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

Murphy, R. R., J. L. Casper, M. J. Micire and J. Hyams (2000) Mixed-Initiative Control of Multiple Heterogeneous Robots for Urban Search and Rescue, University of Central Florida: CRASAR-TR2000-1

Nieuwenhuisen, D. and M. H. Overmars (2002) Motion Planning for Camera Movements in Virtual Environments, Utrecht University Information and Computing Sciences: