

Scaling Effects for Synchronous vs. Asynchronous Video in Multi-robot Search

Huadong Wang, Michael Lewis, & Shih-Hsiang Chien
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260

Prasanna Velagapudi
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Camera guided teleoperation has long been the preferred mode for controlling remote robots with other modes such as asynchronous control only used when unavoidable. Because controlling multiple robots places additional demands on the operator we hypothesized that removing the forced pace for reviewing camera video might reduce workload and improve performance. In an earlier experiment participants operated four teams performing a simulated urban search and rescue (USAR) task using a conventional streaming video plus map interface or an experimental interface without streaming video but with the ability to store panoramic images on the map to be viewed at leisure. Search performance was somewhat better using the conventional interface; however, ancillary measures suggested that the asynchronous interface succeeded in reducing temporal demands for switching between robots. This raised the possibility that the asynchronous interface might perform better if teams were larger. In this experiment we evaluate the usefulness of asynchronous video for teams of 4, 8, or 12 robots. As in our earlier study we found a slight advantage in accuracy in marking victim locations for streaming video but overall performance was very similar.

INTRODUCTION

Practical applications of robotics can be classified by two distinct modes of operation. Terrestrial robotics in tasks such as surveillance, bomb disposal, or pipe inspection has used synchronous realtime control relying on intensive operator interaction usually through some form of teleoperation. Interplanetary and other long distance robotics subject to lags and intermittency in communications have used asynchronous control relying on labor intensive planning of waypoints and activities that are subsequently executed by the robot. In both cases planning and decision making are performed primarily by humans with robots exercising reactive control through obstacle avoidance and safeguards. The near universal choice of synchronous control for situations with reliable, low latency communication suggests a commonly held belief that experientially direct control is more efficient and less error prone. When this implicit position is rarely discussed it is usually justified in terms of “naturalness” or “presence” afforded by control relying on teleoperation. Fong and Thorpe (2001) 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 appears to heavily influence the appeal for this interaction style.

Remotely controlled robots for urban search and rescue (USAR), are typically equipped with both a ptz video camera for viewing the environment and a laser range finder for building a map and localizing the robot on that map. The video feed and map are usually presented in separate windows on the user interface and intended to be used in conjunction. While Casper and Murphy

(2003) reporting on experiences in searching for victims at the World Trade

Center observed that it was very difficult for an operator to handle both navigation and exploration of the environment from video information alone, Yanco and Drury (2004) found that first responders using a robot to find victims in a mock environment made little use of the generated map. Nielsen and Goodrich (2006), by contrast, found search generated maps to be superior in assisting operators to escape from a maze.

When considering such potential advantages and disadvantages of viewpoints it is important to realize that there are two, not one, important subtasks that are likely to engage operators (Tan, Robertson, & Czerwinski, 2001). The escape task was limited to *navigation*, the act of explicitly moving the robot to different locations in the environment. In many applications search, the process of acquiring a specific viewpoint or set of viewpoints containing a particular object may be of greater concern. Because search relies on moving a viewpoint through the environment to find and better view target objects, it is an inherently egocentric task. This is not necessarily the case for navigation which does not need to identify objects but only to avoid them.

Search, particularly multi-robot search, presents the additional problem of assuring that areas the robot has traversed have been thoroughly searched for targets. This requirement directly conflicts with the navigation task which requires the camera to be pointed in the direction of travel in order to detect and avoid objects and steer toward its goal. These difficulties are accentuated by the need to switch attention among robots which may increase the likelihood that a view containing a target will be missed. In earlier studies (Wang & Lewis, 2007a, Wang & Lewis

2007b) we have demonstrated that success in search is directly related to the frequency with which the operator shifts attention between robots and hypothesized that this might be due to victims missed while servicing other robots. Recent data (Wang et al., 2009), however, suggests that other effects involving situation awareness may be involved.

Asynchronous Imagery

To combat these problems of attentive sampling among cameras, incomplete coverage of searched areas, and difficulties in associating camera views with map locations we are investigating the potential of asynchronous control techniques previously used out of necessity in NASA applications as a solution to multi-robot search problems. Due to limited bandwidth and communication lags in interplanetary robotics, camera views are closely planned and executed. Rather than transmitting live video and moving the camera about the scene, photographs are taken from a single spot with plans to capture as much of the surrounding scene as possible. These photographs taken with either an omnidirectional overhead camera (camera faces upward to a convex mirror reflecting 360°) and dewarped (Murphy, 1995, Shiroma et al., 2004) or stitched together from multiple pictures from a ptz camera (Volpe, 1999) provide a panorama guaranteeing complete coverage of the scene from a particular point. If these points are well chosen, a collection of panoramas can cover an area to be searched with greater certainty than imagery captured with a ptz camera during navigation. For the operator searching within a saved panorama the experience is similar to controlling a ptz camera in the actual scene, a property that has been used to improve teleoperation in a low bandwidth high latency application (Fiala, 2005).

In our USAR application which requires finding victims and locating them on a map we merge map and camera views as in (Ricks, Nielsen, & Goodrich, 2004). The operator directs navigation from the map being generated with panoramas being taken at the last waypoint of a series. The panoramas are stored and accessed through icons showing their locations on the map. The operator can find victims by asynchronously panning through these stored panoramas as time becomes available. When a victim is spotted the operator uses landmarks from the image and corresponding points on the map to record the victim's location. By changing the task from a forced paced one with camera views that must be controlled and searched on multiple robots continuously to a self paced task in which only navigation needs to be controlled in realtime we hoped to provide a control interface that would allow more thorough search with lowered mental workload. The reductions in bandwidth and communications requirements (Bruemmer et al., 2005) are yet another advantage offered by this approach.

Pilot Experiment

In a recent experiment reported in (Velagapudi et al., 2008) we compared performance for operators controlling 4 robot teams at a simulated USAR task using either streaming or asynchronous video displays. Search performance was somewhat better using the conventional interface with operators marking slightly more victims closer to their actual location at each degree of relaxation.

This superiority, however, might have occurred simply because streaming video users had the opportunity to move closer to victims thereby improving their estimates of distance in marking the map. A contrasting observation was that frequency of shifting focus between robots, a practice we have previously found related to search performance (Scerri et al., 2004) was correlated with performance for streaming video participants but not for participants using asynchronous panoramas. Because operators using asynchronous video did not need to constantly switch between camera views to avoid missing victims we hypothesized that for larger team sizes where forced pace search might exceed the operator's attentional capacity asynchronous video might offer an advantage. The present experiment tests this hypothesis.

EXPERIMENT

USARSim and MrCS

The experiment was conducted in the high fidelity USARSim robotic simulation environment (Lewis, Wang, & Hughes, 2007) developed as a simulation of urban search and rescue (USAR) robots and environments intended as a research tool for the study of human-robot interaction (HRI) and multi-robot coordination. The MrCS (Multi-robot Control System), a multirobot communications and control infrastructure with accompanying user interface developed for experiments in multirobot control and RoboCup

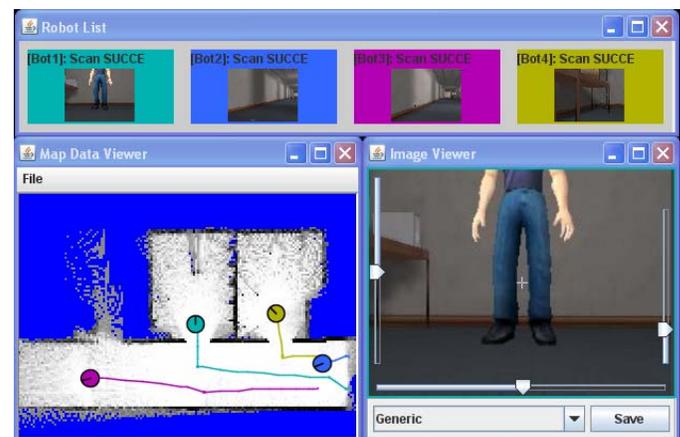


Figure 1. MrCS components for Streaming Video mode

competition (Wang & Lewis, 2007a) was used with appropriate modifications in both experimental conditions. MrCS provides facilities for starting and controlling robots in the simulation, displaying camera and laser output, and supporting inter-robot communication through Machinetta (Scerri et al., 2004) a distributed multiagent system. The distributed control enables us to scale robot teams from small to large.

Figures 1 and 2 show the elements of the MrCS involved in

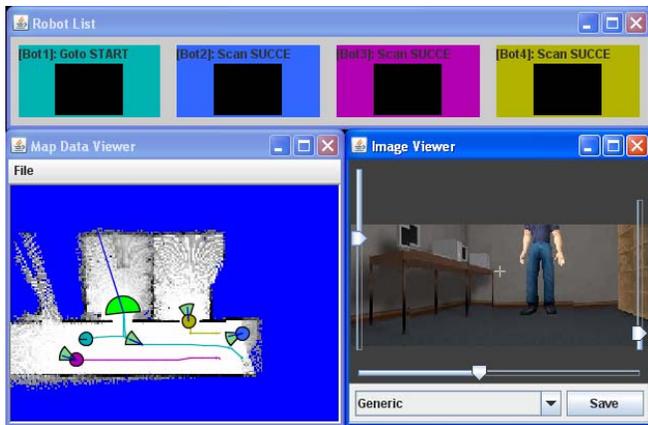


Figure 2. MrCS Asynchronous Panorama mode

this experiment. In the standard MrCS (Fig. 1) the operator selects the robot to be controlled from the colored thumbnails at the top of the screen. Robots are tasked by assigning waypoints on a heading-up map through a teleoperation widget. The current locations and paths of the robots are shown on the *Map Data Viewer*. In the Panorama interface thumbnails are blanked out and images are acquired at the terminal point of waypoint sequences.

Method

A large search environments previously used in the 2006 RoboCup Rescue Virtual Robots competition (Balakirsky et al., 2007) was selected for use in the experiment. The environment consisted of maze like halls with many rooms and obstacles, such as chairs, desks, cabinets, and bricks. Victims were evenly distributed throughout the environments. Robots were started at different locations leading to exploration of different but equivalent areas of the environment. A third simpler environment was used for training. The experiment followed a between groups design with participants searching for victims using either panorama or streaming video modes. Participants searched over three trials beginning with 4 robots, then searching with 8, and finally 12. Robots were started from different locations within a large environment making learning from previous trials unlikely.

Participants and Procedure

29 paid participants were recruited from the University of Pittsburgh community. None had prior experience with robot control although most were frequent computer users. Approximately a quarter of the participants reported playing computer games for more than one hour per week.

Procedure

After collecting demographic data the participant read standard instructions on how to control robots via MrCS. In the following 15~20 minute training session, the participant practiced control operations for either the panorama or streaming video mode and tried to find at least one victim in the training environment under the guidance of the experimenter. Participants then began three testing sessions in which they performed the search task controlling 4, 8, and 12 robots.

RESULTS

Data were analyzed using a repeated measures ANOVA comparing streaming video performance with that of

asynchronous panoramas. On the performance measures, victims found and area covered, the groups showed nearly identical performance with victim identification peaking sharply at 8 robots accompanied by a slightly less dramatic maximum for search coverage. The differences in precision for marking victims observed in the pilot study were found again. For victims marked within 2m, the average number of victims found in the panorama condition was 5.36 using 4

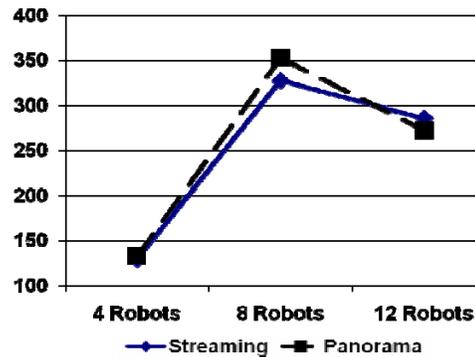


Figure 3. Area Covered

robots, 5.50 for 8 robots, but dropping back to 4.71 when using 12 robots. Participants in the Streaming condition were significantly more successful at this range, $F_{1,29} = 3.563, p < .028$, finding 4.8, 7.07 and 4.73 victims respectively.

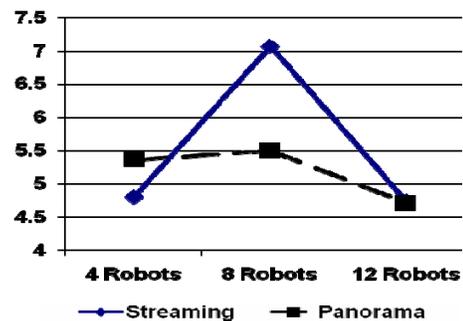


Figure 4. Victims Found as a function of N robots (2 m)

A similar advantage was found for victims marked within 1.5m, with the average number of victims found in the panorama condition dropping to 3.64, 3.27 and 2.93 while participants in the streaming condition were more successful, $F_{1,29} = 6.255, p < .0025$, finding 4.067, 5.667 and 4.133 victims respectively.

Fan-out (Olsen & Wood 2004) is a model-based estimate of the number of robots an operator can control. While Fan-out was conceived as an invariant measure, operators are noticed to adjust their criteria for adequate performance to accommodate the available robots (Wang et al., 2009 , Humphrey et al., 2007). We interpret Fan-out as a measure of attentional reserves. If Fan-out is greater than the number of robots there are remaining reserves. If Fan-out is less than the number of robots, capacity has already been exceeded.

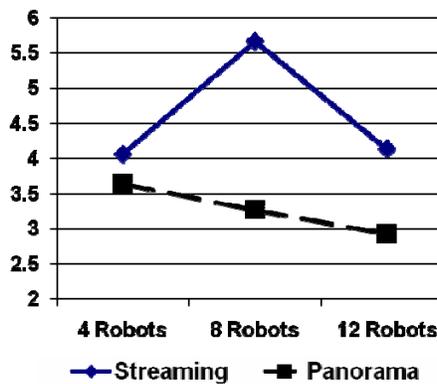


Figure 5. Victims Found as a function of N robots (1.5m)

Fan-out for the panorama conditions increased from 4.1, 7.6 and 11.1 for 4 to 12 robots. Fan-out, however, was uniformly higher in the streaming video condition, $F_{1,29} = 3.355$, $p < .034$, with 4.4, 9.12 and 13.46 victims respectively.

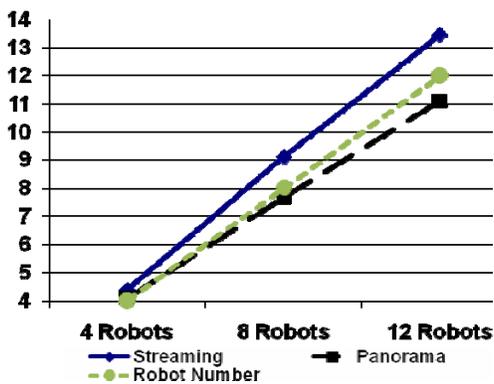


Figure 6. Fan-out as a function of N robots

Number of robots had a significant effect on every dependent measure collected except waypoints per mission (next lowest N switches in focus robot, $F_{2,54} = 16.74$, $p < .0001$). The streaming and panorama conditions were easily distinguished by some process measures. Both streaming and panorama operators followed the same pattern issuing the fewest waypoints per mission to command 8 robots, however, panorama participants in the 8 robot condition issued observably fewer (2.96 vs. 3.16) waypoints.

The closely related pathlength/mission measure follows a similar pattern with no interaction but significantly shorter paths (5.07 m vs. 6.19 m) for panorama participants, $F_{2,54} = 3.695$, $p = .065$.

The number of missions and switches in focus by contrast were nearly identical for the two groups showing only the recurring significant effect for N robots. A similar closeness is found for NASA-TLX workload ratings which rise together monotonically for N robots (figure 9).

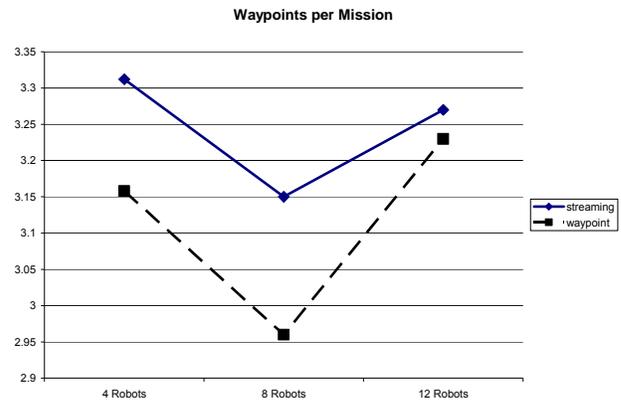


Figure 7. Waypoints per mission

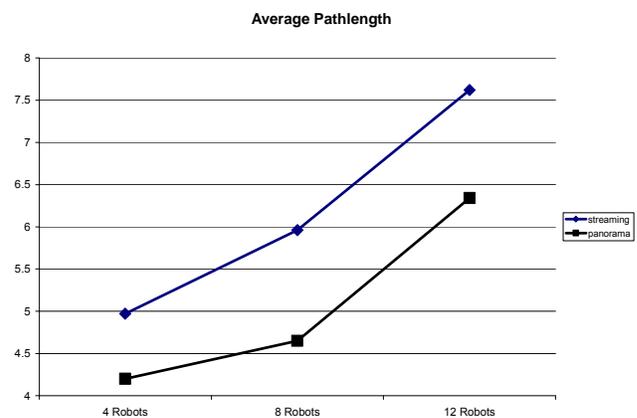


Figure 8. Average pathlength per mission

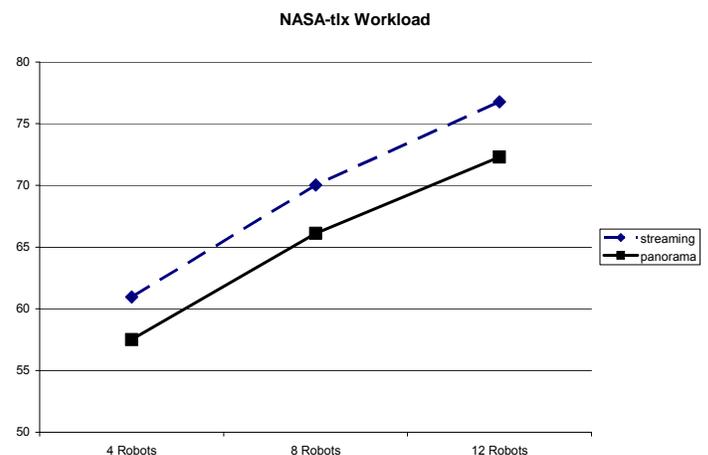


Figure 9. NASA-TLX Workload

DISCUSSION

The most unexpected thing about these data is how similar the performance of streaming and asynchronous panorama participants was. The tasks themselves appear quite dissimilar. In the panorama condition participants direct their robots by adding waypoints to a map without getting to see the robots' environment directly. Typically they tasked robots sequentially and then went back to look at the panoramas that had been taken. Because

panorama participants were unable to see the robot's surrounding except at terminal waypoints, paths needed to be shorter and contain fewer waypoints in order to maintain situation awareness and avoid missing potential victims. Despite fewer waypoints and shorter paths, panorama participants managed to cover the same area as streaming video participants within the same number of missions. Ironically, this greater efficiency may have resulted from the absence of distraction from streaming video (Yanco & Drury, 2004) and is consistent with (Nielsen & Goodrich, 2006) in finding maps especially useful for navigating complex environments.

Examination of pauses in the streaming video condition failed to support our hypothesis that these participants would execute additional maneuvers to examine victims. Instead, streaming video participants seemed to follow the same strategy as panorama participants of directing robots to an area just inside the door of each room. This leaves panorama participants' inaccuracy in marking victims unexplained other than through a general loss of situation awareness. This explanation would hold that lacking imagery leading up to the panorama, these participants have less context for judging victim location within the image and must rely on memory and mental transformations.

Panorama participants also showed lower Fan-out perhaps as a result of issuing fewer waypoints for shorter paths leading to more frequent interactions.

While differences in switching focus among robots were found in our earlier study (Wang & Lewis 2007b) the present data (figure 6) show performance to be almost identical.

While we undertook this study to determine whether asynchronous video might prove beneficial to larger teams we found performance to be essentially equivalent to the use of streaming video at all team sizes with a small sacrifice of accuracy in marking victims. This surprising finding suggests that in applications that may be too bandwidth limited to support streaming video or involve substantial lags, map-based displays with stored panoramas may provide a useful display alternative without seriously compromising performance.

ACKNOWLEDGMENTS

This research has been sponsored in part by AFOSR FA9550-07-1-0039 and AFOSR FA9550-08-1-0356.

REFERENCES

- Balakirsky, S., Carpin, S., Kleiner, A., Lewis, M., Visser, A., Wang, J., & Zipara, V. (2007). "Toward heterogeneous robot teams for disaster mitigation: Results and performance metrics from RoboCup Rescue," *Journal of Field Robotics*, 24(11-12), 943-967.
- Bruemmer, D., Few, A., Walton, M., Boring, R., Marble, L., Nielsen, C., & Garner, J. (2005). "Turn off the television: Real-world robotic exploration experiments with a virtual 3-D display." *Proc. HICSS, Kona, HI*.
- Casper, J. & Murphy, R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 33(3):367-385.
- Fiala, M. (2005). "Pano-presence for teleoperation," *Proc. Intelligent Robots and Systems (IROS 2005)*, 3798-3802.
- Fong, T. & Thorpe, C. (1999). "Vehicle teleoperation interfaces," *Autonomous Robots*, no. 11, pp. 9-18.
- Humphrey, C. Henk, G. Sewell, B. Williams, B. & Adams, J. (2007) "Assessing the Scalability of a Multiple Robot Interface," *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robotic Interaction*.
- Lewis, M., Wang, J., & Hughes, S. (2007), "USARsim : Simulation for the Study of Human-Robot Interaction," *Journal of Cognitive Engineering and Decision Making*, 1(1), 98-120.
- Murphy, J. (1995). "Application of Panoramic Imaging to a Teleoperated Lunar Rover," in *Proceedings of the 1995 International Conference on Systems, Man, and Cybernetics*.
- Nielsen, C. & Goodrich, M. (2006). Comparing the usefulness of video and map information in navigation tasks. In *Proceedings of the 2006 Human-Robot Interaction Conference*, Salt Lake City, Utah.
- Olsen, D. & Wood, S. (2004). Fan-out: measuring human control of multiple robots, in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, ACM Press: Vienna, Austria. p. 231-238.
- Ricks, B., Nielsen, C., and Goodrich, M. (2004). Ecological displays for robot interaction: A new perspective. In *International Conference on Intelligent Robots and Systems IEEE/RSJ*, Sendai, Japan.
- Scerri, P., Xu, Y., Liao, E., Lai, G., Lewis, M., & Sycara, K. (2004). "Coordinating large groups of wide area search munitions," In D. Grundel, R. Murphey, and P. Pandalos (Ed.), *Recent Developments in Cooperative Control and Optimization*, Singapore: World Scientific, 451-480.
- Shiroma, N., Sato, N., Chiu, Y. & Matsuno, F. (2004). "Study on effective camera images for mobile robot teleoperation," In *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, Kurashiki, Okayama Japan.
- Tan, D., Robertson, G. & Czerwinski, M. (2001). "Exploring 3D navigation: Combining speed-coupled flying with orbiting," in *CHI 2001 Conf. Human Factors Comput. Syst.*, Seattle, WA.
- Velagapudi, P., Wang, J., Wang, H., Scerri, P., Lewis, M., & Sycara, K. (2008). Synchronous vs. Asynchronous Video in Multi-Robot Search, *Proceedings of first International Conference on Advances in Computer-Human Interaction (ACHI'08)*.
- Volpe, R. (1999). "Navigation results from desert field tests of the Rocky 7 Mars rover prototype," in *The International Journal of Robotics Research*, 18, pp.669-683.
- Wang, H., Lewis, M., Velagapudi, P., Scerri, P., & Sycara, K. (2009). How search and its subtasks scale in N robots, *2009 Conference on Human-Robot Interaction (HRI'09)*, ACM.
- Wang, J. & Lewis, M. (2007a) "Human control of cooperating robot teams", *2007 Human-Robot Interaction Conference*, ACM.
- Wang, J. & Lewis, M. (2007b). "Assessing coordination overhead in control of robot teams," *Proceedings of the 2007 International Conference on Systems, Man, and Cybernetics*.
- Yanco, H. & Drury, J. (2004). "Where am I?" Acquiring situation awareness using a remote robot platform. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*.
- Yanco, H., Baker, M., Casey, R., Keyes, B., Thoren, P., Drury, J., Few, D., Nielsen, C., & Bruemmer, D. (2006). "Analysis of human-robot interaction for urban search and rescue," *Proceedings of PERMIS*.