Synchronous vs. Asynchronous Video in Multi-Robot Search

Prasanna Velagapudi[1], Jijun Wang[2], Huadong Wang[2], Paul Scerri[1], Michael Lewis[2], Katia Sycara[1]
[1]Robotics Institute, Carnegie Mellon University
[2]School of Information Sciences, University of Pittsburgh
Pittsburgh, PA
pkv@cmu.edu, {jiw, huw16}@pitt.edu, pscerri@cs.cmu.edu, ml@sis.pitt.edu, katia@cs.cmu.edu

Abstract

Camera guided teleoperation has long been the preferred mode for controlling remote robots, with other modes such as asynchronous control only used when unavoidable. In this experiment we evaluate the usefulness of asynchronous operation for a multirobot search task. Because controlling multiple robots places additional demands on the operator, removing the forced pace for reviewing camera video might reduce workload and improve performance. In the reported experiment participants operated four robot teams performing a simulated urban search and rescue (USAR) task using either conventional streaming video plus a 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 suggest that the asynchronous interface succeeded in reducing temporal demands for switching between robots.

1. 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 [7] 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.

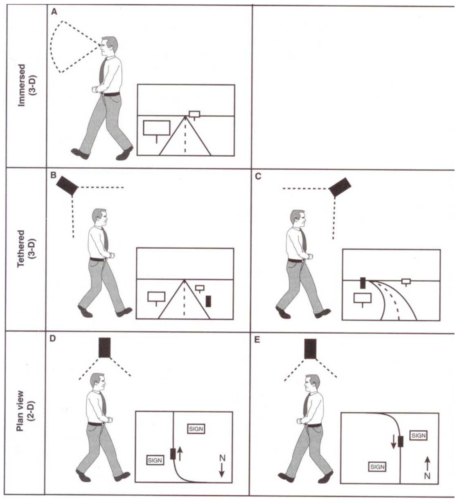

Figure 1. Viewpoints for control from Wickens and Hollands *Engineering Psychology and Human Performance*, 1999.

Control based on platform mounted cameras, however, is no panacea. Wickens and Hollands [23] identify 5 viewpoints used in control, depicted in Figure 1. Three of them, immersed, tethered, and "plan view" can be associated with the moving platform while 3rd person (tethered) and plan views require fixed cameras. In the immersed or egocentric view (A) the operator views the scene from a camera mounted on the platform. The field of view provided by the video feed is often much narrower than human vision, leading to the experience of viewing the world through a soda straw from a foot or so above the ground. This perceptual impairment leaves the operator prone to numerous, well-known operational errors, including disorientation, degradation of situation awareness, failure to recognize hazards, and simply overlooking relevant information [5,11]. A sloped surface, for example, gives the illusion of being flat when viewed from a camera mounted on a platform traversing that surface [9]. For fixed cameras, the operator's ability to survey a scene is limited by the mobility of the robot and his ability to retain viewed regions of the scene in memory as the robot is maneuvered to obtain adjacent views. A pan-tilt-zoom (ptz) camera resolves some of these problems but introduces new ones involving discrepancies between the robots heading and the camera view, which can frequently lead to operational mishaps [25]. A tethered "camera" (B,C) provides an oblique view of the scene showing both the platform and its 3D environment. The 3rd person fixed view (C) is akin to an operator's view controlling slot cars and has been shown effective in avoiding rollovers and other teleoperation accidents [11] but can't be used anywhere an operator's view might be obstructed such as within buildings or in rugged terrain. The tethered view (B) in which a camera "follows" an avatar (think Mario Brothers©) is widely favored in virtual environments [12,19] for its ability to show the object being controlled in relation to its environment by showing both the platform and an approximation of the scene that might be viewed from a camera mounted on it. This can be simulated for robotic platforms by mounting a camera on a flexible pole, giving the operator a partial view of his platform in the environment [24]. However, the restriction in field of view and the necessity of pointing the camera downward limit this strategy's ability to survey a scene, although it can provide a view of the robot's periphery and nearby obstacles that could not be seen otherwise. The exocentric views show a 2 dimensional version of the scene such as might be provided by an overhead camera. It cannot be directly obtained from an onboard camera, but for robots equipped with laser range finders, generating a map and localizing the robot provides a method for approximating an exocentric view of the platform. If this view rotates with the robot (heading up) it is a type D plan view. If it remains fixed (North up) it is of type E.

An early comparison at Sandia Laboratory between viewpoints for robot control [11] investigating accidents focused on the most common of these: (A) egocentric from onboard camera and (C) 3rd person. The finding was that all accidents involving rollover occurred under egocentric control while 3rd person control led to bumping and other events resulting from obstructed or distanced views. In current experimental work in remotely controlled robots for urban search and rescue (USAR), robots 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. 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 [4] 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 from video information alone, Yanco and Drury [24] found that first responders using a robot to find victims in a mock environment made little use of the generated map. One possible explanation is that video is more attention grabbing than other presentations [8], leading operators to control primarily from the camera while ignoring other available information. A number of recent studies conducted by Goodrich, Neilsen, and colleagues [3,14,16,26] have attempted to remedy this through an ecological interface that fuses information by embedding the video display within the map. The resulting interface takes the 2D map and extrudes the identified surfaces to derive a 3D version resembling a world filled with cubicles. The robot is located on this map, with the video window placed in front of it at the location being viewed. This strategy uses the egocentric camera view and the overhead view from the map to create a synthetic tethered view of the sort found most effective in virtual environments and games [12,19]. The anticipated advantages, however, have been difficult to demonstrate with ecological and conventional interfaces trading advantages across measures. Of particular interest have been comparisons between control based exclusively on maps or videos. In complex environments with little opportunity for preview, maps were superior in assisting operators to escape from a maze [14].

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 [19]. The escape task and the accidents reviewed at Sandia involved 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. While both navigation and search require the robot to move, they differ in the focus of the movement. Navigation occurs with respect to the environment at large, while search references a specific object or point within that environment. Switching between these two subtasks may play a major role in undermining situation awareness in teleoperated environments. For example, since search activities move the robot with respect to an object, viewers may lose track of their global position within the environment, possibly requiring additional maneuvering to reorient the operator before navigation can be effectively resumed. Because search relies on moving a viewpoint through the environment to find and 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 traversed areas have been thoroughly searched for targets. This conflicts with the navigation task which requires the robot's camera to view the direction of travel in order to detect and avoid obstacles and steer toward its goal. If the operator attempts to compromise by choosing a path to traverse and then panning the camera to search as the robot moves, he runs both the risk of hitting objects while he is looking away and missing targets as he attends to navigation. For multirobot control these difficulties are accentuated by the need to switch attention among robots, multiplying the likelihood that a view containing a target will be missed. In earlier studies [21,22] we have demonstrated that success in search is directly related to the frequency with which the operator shifts attention between robots over a variety of conditions. An additional issue is the operator's confidence that an area has been effectively searched. In our natural environment we move and glance about, using planning and proprioception to knit the resulting views into a representation of our environment. In controlling a robot we are deprived of these natural bridging cues and have difficulty recognizing as we pan and tilt whether we are resampling old views or missing new ones. The extent of this effect was demonstrated by Pausch [15] who found that participants searching for an object in a virtual room using a headmounted display were twice as fast as when they used a simulated handheld camera. Since even the handheld camera provides many ecological cues we should expect viewing from a moving platform through a ptz camera to be substantially worse.

2. Experiment

2.1. 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 robotic 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 are either taken with an omnidirectional overhead camera (camera facing upward to a convex mirror reflecting 360°) and dewarped [13,18] or stitched together from multiple pictures from a ptz camera [20] to 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 [6].

In our USAR application, which requires finding victims and locating them on a map, we merge map and camera views as in [16]. The operator directs navigation from the map by assigning waypoints to robots, 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 [3] are yet another advantage offered by this approach.

2.2. USARSim and MrCS

The experiment was conducted in the high fidelity USARSim robotic simulation environment [10] we 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. USAR-

Sim is freely available and can be downloaded from *http://www.sourceforge.net/projects/usarsim*. It uses Epic Games' UnrealEngine2 to provide a high fidelity simulator at low cost. USARSim supports HRI by accurately rendering user interface elements (particularly camera video), accurately representing robot automation and behavior, and accurately representing the remote environment that links the operator's awareness with the robot's behaviors. MrCS (Multi-robot Control System), a multirobot communications and control infrastructure with accompanying user interface developed for experiments in multirobot control and RoboCup competition [21] 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 [17], a distributed mutiagent system. The distributed control enables us to scale robot teams from small to large.

Figures 2 and 3 show the elements of the MrCS involved in this experiment. In the standard MrCS (Fig. 2) the operator selects the robot to be controlled from the colored thumbnails at the top of the screen that show a slowly updating view from the robot's camera. Streaming video from the in focus robot which the operator now controls is displayed on the Image Viewer. To view more of the scene the operator uses pan/tilt sliders (not shown) to control the camera. Robots are tasked by assigning waypoints on a heading-up map on the Mission Panel (not shown) or through a teleoperation widget (not shown). The current locations and paths of the robots are shown on the Map Data Viewer. Although the experimental panoramic interface (Fig. 3) looks much the same it behaves quite differently. Robots are again selected for control from the colored thumbnails which now lack images. Panoramic images are acquired at the terminal point of waypoint sequences. Icons conveying the robot's location and orientation at these points are placed on the map for accessing the panoramas. The operator can then view stored panoramas by selecting an icon and dragging a mouse over the Image Viewer to move the image around or using the mouse's scroll wheel to zoom in and out of the image. The associated icon on the Map Data Viewer changes orientation in accordance with the part of the scene being viewed.

2.3. Method

Two equivalent search environments previously used in the 2006 RoboCup Rescue Virtual Robots competition [1] were selected for use in the experiment. Each environment was a maze like hall with many rooms and obstacles, such as chairs, desks, cabinets, and bricks. Victims were evenly distributed within the environments. A third simpler environment was used for training. The experiment followed a repeated measures design with participants searching for victims using both panorama and streaming video modes. Presentation orders for mode were counterbalanced. Test environments were presented in a fixed order confounding differences between the environments with learning effects. Because the environments were closely matched we will discuss these differences as transfer of training effects.

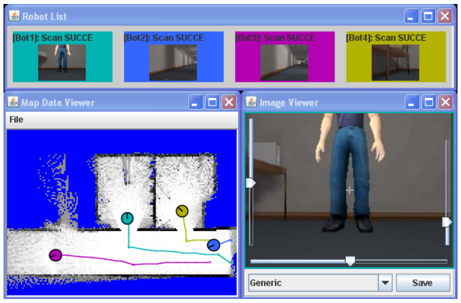

Figure 2. MrCS components for Streaming Video mode

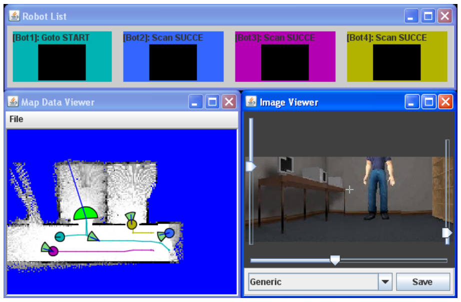

Figure 3. MrCS components for Asynchronous Panorama mode

2.3.1. Participants. 21 paid participants, 9 male and 12 female old recruited from the University of Pittsburgh community. None had prior experience with robot control although most (15) were frequent computer users. Six of the participants (28%) reported playing computer games for more than one hour per week.

2.3.2. Procedure. After collecting demographic data the participant read standard instructions on how to control robots via MrCS. In the following 15 $\sim$ 20 minute training session, the participant practiced control operations for panorama and streaming video modes (both were enabled) and tried to find at least one victim in the training environment under the guidance of the experimenter. Participants then began two testing ses-

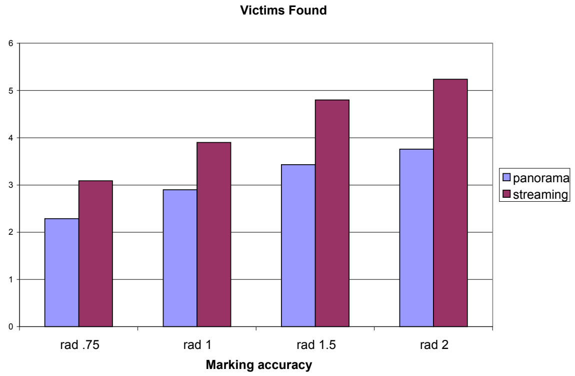

Figure 4. Effects of display modes

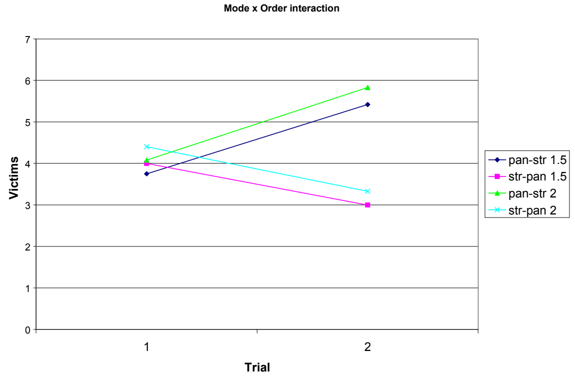

Figure 5. Mode by Order interaction

sions in which they performed the search task using the panorama and streaming video modes.

3. Results

Only one participant failed to find any victims under the most lenient criterion of marking the victim within 2m of the actual location. This occurred in the panorama mode on the initial trial. Overall, participants were successful in searching the environment in either mode, finding as many as 9 in a trial. The average across conditions using the 2m radius was 4.5, falling to 4.1 for a 1.5m radius, 3.4 at 1m and 2.7 when they were required to mark victims within .75m. Repeated measures ANOVAs found differences in victim detection favoring the streaming video mode at the 1.5m radius F(1,19) = 8.038, p=.01, and 2.0m radius F(1,19)=9.54, p=.006. Figure 4 shows these differences.

Although no significant order effect (learning) was observed, a significant interaction was found between video mode and presentation order for victims marked within a 1.5m, F(1,19)=7.34, p=.014 or a 2m, F(1,19)=8.77, p=.008, range. Figure 5 illustrates the substantial differences between the presentation order groups. In contrast to overall trends, the group receiving streaming video on the first trial performed no better than those initially using the panorama. Whether this was due to failure of randomization to provide equivalent groups or asymmetric transfer of training between these conditions cannot be determined from these data.

As in earlier studies, we found a positive relation, F(1,19)=3.86, p=.064, between the number of times the operator switched between robots and the number of victims found, seen in Figure 6. In accord with our hypothesis that this is due to the forced pace of performing the task using streaming video, no relation was found between the frequency of switching and victims for the panorama mode.

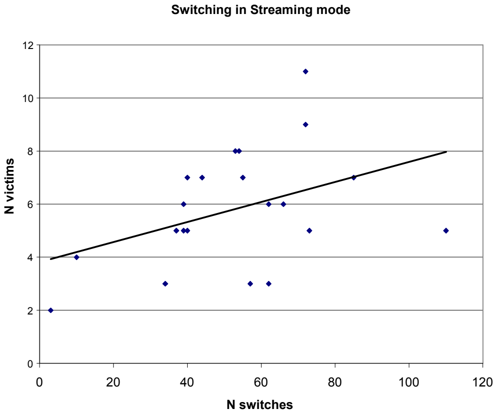

Figure 6. Finding victims was related to switches in Streaming mode.

4. Discussion

Our original motivation for developing a panorama mode for MrCS was to address restrictions posed by a communications server added to this year's RoboCup Rescue competition to simulate bandwidth limitations and drop-outs due to attenuation from distance and obstacles. Although the panorama mode was designed to drastically reduce bandwidth and allow operation despite intermittent communications our system was so effective we decided to test it under conditions most favorable to a conventional interface. Our experiment shows that under such conditions allowing uninterrupted, noise free, streaming video a conventional interface leads to somewhat better (5 vs. 4 victims) search performance. The switching results, however, suggest that asynchronous panoramas do overcome the forced pace switching needed to avoid missing unattended targets in realtime interfaces. We would expect this advantage to grow as the number of robots increases with performance surpassing streaming video at some point. Just as [14] have demonstrated that maps may be better than cameras for navigation we hope that asynchronous video and related strategies may play a role in improving multirobot search capabilities. Coupled with the ability to control robots under poor communication conditions such as are expected in USAR and other field work we believe that interface innovations of this sort have an important role to play in making control of robot teams a reality.

5. Acknowledgements

This research has been sponsored in part by AFOSR FA9550-07-1-0039, AFOSR FA9620-01-1-0542, L3-Communications (4500257512) and NSF ITR IIS-0205526. We are indebted to the anonymous reviewers for their helpful comments.

References

[1] S. Balakirsky, S. Carpin, A. Kleiner, M. Lewis, A. Visser, J. Wang, and V. Zipara, "Toward hetereogeneous robot teams for disaster mitigation: Results and performance metrics from RoboCup Rescue," *Journal of Field Robotics*, in press.

[2] D. J. Bruemmer, J. L. Marble, D. A. Few, R. L. Boring, M. C. Walton, and C. W. Nielsen, "Let rover take over: A study of mixed-initiative control for remote robotic search and detection," *IEEE Trans. Sys., Man, Cyber.-Part A*, 35(4):494-504, July 2005.

[3] D. J. Bruemmer, D. A. Few, M. C. Walton, R. L. Boring, J. L. Marble, C. W. Nielsen, and J. Garner, "Turn off the television: Real-world robotic exploration experiments with a virtual 3-D display," *Proc. HICSS*, Kona, HI, 2005.

[4] J. Casper and R. R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center," *IEEE Trans. Sys., Man, Cyber.-Part B*, 33(3):367-385, June, 2003.

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

[6] M. Fiala, "Pano-presence for teleoperation," *Proc. IROS*, 3798-3802, 2005.

[7] T. Fong and C. Thorpe, "Vehicle teleoperation interfaces," *Autonomous Robots*, no. 11, pp. 9-18, 2001.

[8] R. Kubey and M. Csikszentmihalyi. "Television addiction is no mere metaphor," *Scientific American*, 286(2):62-68, 2002.

[9] M. Lewis and J. Wang, "Gravity referenced attitude display for mobile robots : Making sense of what we see," *IEEE Trans. Sys., Man, Cyber.-Part A*, 37(1), 94-105, 2007.

[10] M. Lewis, J. Wang, and S. Hughes, "USARsim : Simulation for the Study of Human-Robot Interaction," *Journal of Cognitive Engineering and Decision Making*, 1(1), 98-120, 2007.

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

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

[13] J. R. Murphy, "Application of Panospheric Imaging to a Teleoperated Lunar Rover," *Proc. Int. Conf. Sys., Man, Cyber.*, 1995.

[14] C. Nielsen and M. Goodrich, "Comparing the usefulness of video and map information in navigation tasks," *Proc. Human-Robot Interaction Conf.*, Salt Lake City, Utah, March 2006.

[15] R. Pausch, M. A. Shackelford, and D. Proffitt, "A user study comparing head-mounted and stationary displays," *IEEE Symp. on Research Frontiers in Virtual Reality*, October 25-26, 1993, San Jose, CA.

[16] B. W. Ricks, C. W. Nielsen, and M. A. Goodrich, "Ecological displays for robot interaction: A new perspective," *Int. Conf. on Intelligent Robots and Systems IEEE/RSJ*, Sendai, Japan, 2004.

[17] P. Scerri, Y. Xu, E. Liao, G. Lai, M. Lewis, and K. Sycara, "Coordinating large groups of wide area search munitions," D. Grundel, R. Murphey, and P. Pandalos (Ed.), *Recent Developments in Cooperative Control and Optimization*, Singapore: World Scientific, 451-480.

[18] N. Shiroma, N. Sato, Y. Chiu, and F. Matsuno, "Study on effective camera images for mobile robot teleoperation," *Proc. IEEE Int. Workshop on Robot and Human Interactive Communications*, Kurashiki, Okayama Japan, 2004.

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

[20] R. Volpe, "Navigation results from desert field tests of the Rocky 7 Mars rover prototype," in *The International Journal of Robotics Research*, 18, pp.669-683, 1999.

[21] J. Wang, and M. Lewis, "Human control of cooperating robot teams," *2007 Human-Robot Interaction Conference*, ACM, 2007a.

[22] J. Wang, and M. Lewis, "Assessing coordination overhead in control of robot teams," *Proc. Int. Conf. Sys., Man, Cyber.*, 2007b

[23] C. Wickens and J. Hollands, *Engineering Psychology and Human Performance*, Prentice Hall, 1999.

[24] H. A. Yanco and J. L. Drury, "'Where am I?' Acquiring situation awareness using a remote robot platform". *Proc. Int. Conf. Sys., Man, Cyber.*, October 2004.

[25] H. A. Yanco, J. L. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Journal of Human-Computer Interaction*, 19(1 and 2):117-149, 2004.

[26] H. Yanco, M. Baker, R. Casey, B. Keyes, P. Thoren, J. Drury, D. Few, C. Nielsen, and D. Bruemmer, "Analysis of human-robot interaction for urban search and rescue," *Proc. PERMIS*, 2006.