USARSim and HRI: from Teleoperated Cars to High Fidelity Teams

Michael Lewis, *Member, IEEE*

***Abstract*— USARSim began as a human-robot interaction (HRI) research tool but has since found use in a much wider community and for purposes we had never envisioned. This paper describes a six year HRI research program at the University of Pittsburgh using the simulation. Our original work involved teleoperated control of single robots and primitive simulations. In the most recent experiment teams of operators were controlling 24 robot teams in a high fidelity environment. In between we developed and tested measures of coordination demand, tried out news ways for managing video generated by teams, and investigated scaling effects as operators controlled increasing numbers of robots. This paper provides a brief chronology of this research summarizing their designs and findings.**

I. Introduction

In 2000 in response to rapidly rising costs of academic virtual reality software, we began experimenting with game software as an alternative. After a review of the most suitable engines we chose Unreal Tournament [1] over Quake [2] because of its object-oriented design and convenient java-like scripting language. Our first game engine-based application, CaveUT [3], software for creating multi-projector cave-like displays, was completed in 2001 and reported in a special issue of Communications of the ACM [4] we organized to highlight research groups who had independently begun working with game engines. We developed UTSAF [5], software using the game engine as a stealth viewer (3D visualization) for the ModSAF [6] military simulation shortly thereafter.

Work on USARSim began in late 2002 under an NSF ITR grant to study Robot, Agent, Person (RAP) teams in Urban Search And Rescue (USAR). Because our primary research interest was in human-robot interaction and most USAR robots rely on teleoperation from camera video, accurate simulation of video was our primary concern. As work was beginning, Epic games released Unreal Tournament 2003 which included the Karma physics engine [7] relieving us from simulating behavior manually and dramatically strengthening the engine as a simulation tool.

Manuscript received August 15, 2009. This work was supported in part by AFOSR grants FA9550-07-1-0039, FA9620-01-0542 and ONR grant N000140910680

Michael Lewis is with the School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15260 phone: 412-624-9426; emal: ml@sis.pitt.edu.

II. Teleoperation with single robots

Our initial HRI research using USARSim explored two areas: situation awareness for attitude and camera and viewpoint control. A review of findings such as McGovern's [8] observation that all recorded robot rollovers at Sandia had involved teleoperation using an onboard camera led us to suspect that camera geometry was leading to what we called the "fixed camera" illusion. You will notice this effect if you drive a robot up a ramp. Because the camera is fixed to the robot chassis when the robot mounts an incline the camera will remain perpendicular to the surface making the ramp appear flat and level. Even if attitude data is displayed nearby on an artificial horizon or other analog display it remains difficult for the operator to integrate that data with the camera video being used to drive the robot. As a consequence operators controlling from a fixed camera are prone to driving robots onto dangerously slanted surfaces risking rollovers and other problems. One potential solution is to reference the camera to gravity rather than the robot's chassis. Now when the robot moves onto a slanted surface it looks slanted rather than flat. Because standard cameras do not come with gimbals for gravity referencing and delays associated with attitude sensing and servos might introduce even greater errors this potential solution would be expensive and difficult to test using real robots. In simulation by contrast it was easy to program the viewpoints to reference the chassis or true vertical. This first USARSim experiment reported in [9,10] compared 26 participants controlling a robot using either a fixed (FC) or gravity referenced camera (GRC). The robots were driven across irregular outdoor and indoor environments toward target beacons that could be seen from anywhere in the map. The GRC led to less extreme roll/distance traveled, lower times to completion, and less backing-up (needed to retreat from impassable terrain). An examination of reported cues highlighted the importance of including some part of the chassis within the GRC view to help gauge robot orientation relative to the scene.

The viewpoint control experiments were an outgrowth of earlier studies [11] of "attentive navigation", a technique for automating viewpoint control. As an actor moves through a (virtual) environment he/she may look straight ahead in the direction of travel or pan from side to side to capture a fuller understanding of what is visible from that location. Conventionally this process is automated by planning the agent's path so the viewer becomes a sight-seeing passenger able to look about freely. In attentive navigation, control of gaze is automated instead. Now the agent plans his own path through the world but control over where he looks is automated. This can make a lot of sense in virtual environments where the author may wish to direct a user's attention to some particular object or area but is less generally applicable to robotics. There are, however, similar issues related to coupling camera views to the direction of motion (straight ahead), pan-able views, and object tracking views. A long standing difficulty in mobile robotics involves moving cameras that are inadvertently left off axis when the robot is moved. The path appears clear in the side-pointing camera so the operator drives directly into an obstacle [12]. If the camera isn't movable, however, the operator may need to execute an elaborate dance often losing sight of the target in order to maneuver to obtain a desired viewpoint on an object.

An initial experiment [13] compared 65 operators in five conditions.

· Single Fixed Camera, No orientation indicator
· Single moving Camera , No orientation indicator
· Single moving Camera, orientation indicator
· Multiple moving Cameras, No orientation indicator
· Multiple moving Cameras, orientation indicator

For the first three conditions comparing single cameras the fixed camera eliminates the problem of off axis driving but at the cost of making it difficult to obtain different perspectives. The moving camera makes it easy to inspect the environment but at the risk of off axis driving. The orientation indicator provides an aid for restoring the moving camera to straight ahead before driving. In the multiple camera conditions the operator has the option of keeping one camera pointed in the direction of travel for driving while using the other to search the environment. The orientation indication provides assistance for returning a camera to straight ahead for driving if the operator chooses to use both cameras to search the environment.

Operators had a two step search task. First they needed to locate red cubes scattered throughout an indoor and an outdoor environment. After finding the cube they needed to maneuver closer in order to locate and read a letter on a side of the cube. The experiment found no advantage for the orientation indication. Both the moving camera and the two camera conditions led to identifying more targets suggesting that the ability to visually search the environment was more important to task performance than accurate driving.

A follow on experiment [14] investigated object tracking. Object tracking, called orbiting by [15] is a variant of attentive navigation in which the viewpoint remains fixed on an object as its platform moves through the environment. So, in moving around an object, that object remains in view without panning or other effort by the operator. This method is easy to write in simulation and could plausibly be implemented using laser or other ranging data to localize the object. The experiment compared two groups of 13 participants each performing the "lettered-cube" search task. In the control group operators used two moving cameras as in the earlier experiment. In the experimental group one of the cameras assisted operators by initiating object tracking for nearby cubes. Assisted operators identified more cubes and spend substantially less time maneuvering to read the letter. These viewpoint control experiments support the use of multiple cameras and show that for task relevant assistance such as object tracking, automated control of view point is accepted and benefits operators.

These first USARSim experiments addressed general issues in HRI and teleoperation that did not depend on the fidelity of the simulation. All three studies investigated the teleoperation-from-camera-video which required accurate reproduction of video but only approximate fidelity in other dimensions. Simple car models based on the vehicle class were used in these studies and little attention was given to scaling of the platform or environment.

III. MrCS and High Fidelity Simulation

A. MrCS

Our initial studies satisfied us that the game-based simulation could provide a sound research tool but our project's goal was to investigate much larger systems involving multiple robots and humans. Such systems require a high degree of automation and to model that accurately required paying more detailed attention to robot configuration, sensors, and environmental models. USARSim first took on a recognizable form in this revision, reported in Wang [16] which replaced the original agent-based [17] architecture with a more conventional organization, added conventional APIs ,and developed detailed models of existing platforms and sensors. Jijun Wang [18] developed the MrCS (Multirobot Control System) around the same time to integrate USARSim with Machinetta, a proxy-based coordination infrastructure, and a GUI for interacting with the system. The robot proxy provides low-level autonomy such as guarded motion, waypoint control and middle-level autonomy in path generation. It also communicates between the simulated robot and other proxies to enable the robot to execute the cooperative plan they have generated. The user interacts with the system through the user interface which sends

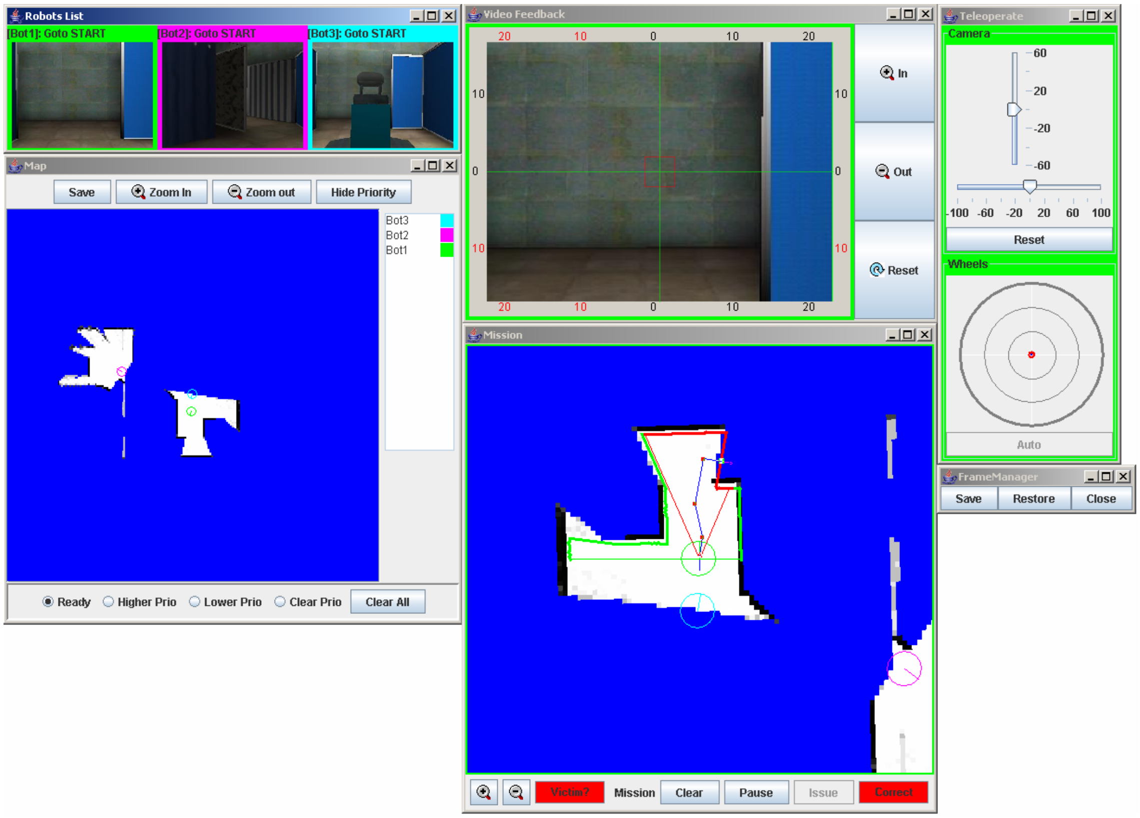

Figure 1. Multirobot Control System (MrCS) user interface

messages to robot proxies and reacts to their responses. Sensor outputs from the camera and laser go directly to the interface without passing through any proxy. A typical interface configuration is shown in Figure 1. On the left side are the global information components: the thumbnails of the individual's camera view (clicked to bring into focus); and the global Map (the bottom panel) that shows the explored areas and each robot's position. In the center are the individual robot control components. The upper component displays the video of the robot being controlled. The bottom component shows the controlled robot's local situation. The local map is camera up, always pointing in the camera's direction. Three increasingly sophisticated versions of MrCS can be downloaded from www.robocuprescue.org in the VR competition listings.

A. Coordinating Teams

The first experiment conducted using MrCS [18] compared manual and mixed-initiative control of 3 robots performing a USAR task followed shortly by an additional fully automated condition [19] to ensure that good mixed-initiative performance had not hiding superior automated performance. In the mixed-initiative condition operators could either teleoperate or assign waypoints. If a robot became idle it chose a waypoint at the nearest frontier and continued exploring.

In the experiment 14 participants searched for victims in both manual and mixed-initiative conditions in a counterbalanced repeated measures design. More area was explored and victims found in the mixed-initiative condition. Interestingly, switches in focus among robots was found to be correlated with good performance and operators switched attention more often in the mixed-initiative condition.

B. Coordination Demand

A theoretical controversy over the equivalence of computational complexity and human difficulty motivated the next set of experiments. Foraging tasks such as USAR allow robots to act more or less independently and we would expect increases in difficulty to be additive. Where robot actions are more interdependent, however, more frequent control might be needed making the task more difficult. Simply assigning robot roles in a plan, for instance, has been shown by Gerkey & Mataric [20] to be O(mn). If the complexity of choosing actions computationally approximates the difficulty of the task for a human, then it could be used to guide decisions about automation. Conversely, a human might be able to solve such problems heuristically making computational difficulty a bad estimator.

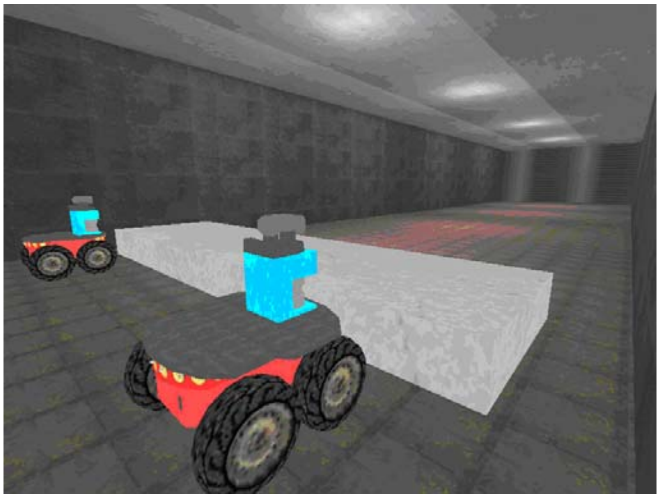

Figure 2. Tight coordination for box pushing

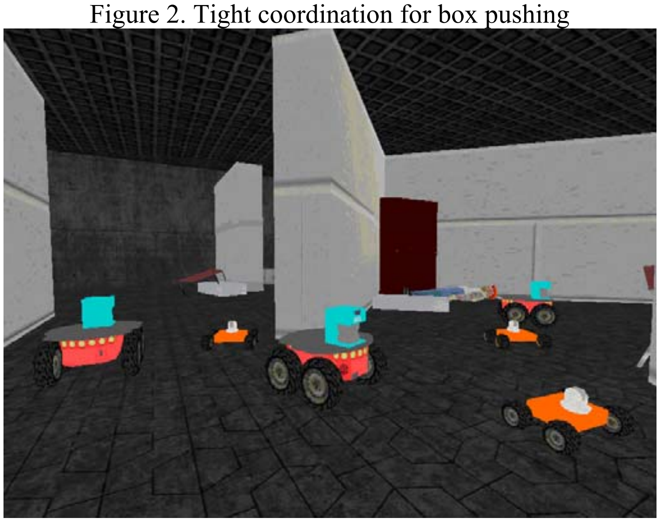

Figure 3. Explorer and Scout robots

These experiments were designed to evaluate coordination demand (CD) a proposed measure of the demand one robot's action(s) place on another. The measure is intended to extend Crandall's [21] neglect tolerance model to coordinating robot teams. The first experiment examined control of robots performing a box pushing task [22] (Figure 2). Fourteen participants controlled pairs of simulated P2-ATs using teleoperation or waypoint control and in the third condition a P2-AT paired with a P2-DX. As predicted CD=1 in the teleoperation condition as operators did not have time to do anything else. The heterogeneous pair showed higher CD, also as predicted. In a follow on experiment [23] seeking measures for less tightly constrained coordination a new definition of CD based on robot types was tested. The measure is based on the premise that CD involves marshalling the resources needed to perform a cooperative task. Since resources are held in common by robots of a particular type demand may be more accurately expressed and measured between types. Operators in this experiment controlled teams of robot pairs consisting of laser equipped explorer robots and camera carrying scouts (Figure 3). The operator needed to mark victims found using the scout's camera on the map generated by the explorer. Operators searched in three conditions with a 20 m explorer scan range (loosely coupled), a 5 m scan range (tightly coupled), or cooperative (explorer automatically follows the scout). Performance was as expected with the 20 m range leading to more victims and better performance with automated coordination.

C. *Scaling to Larger Teams*

While the question of fan-out (how many robots can an operator control?) is of general interest the question of how effects grow with team size offers greater promise for identifying bottlenecks and aspects of control best suited for automation. In a series of studies we have been investigating control of 4, 8, and 12 robot teams and in the most recent study teams of 24 robots. An experiment using a standard USAR task for 4, 8, and 12 robots [24] found a sharp decline

Figure 4. MrCS interface for 24 robots

in victims found and a slight decline in area explored between 8 and 12 robots. Two additional conditions [25] subdividing the operator's task into exploration (navigation) and perceptual search (scanning for victims) showed that effort involved in exploration accounted for most of the difficulty of the task and that victim finding performance was maintained by the perceptual search participants.

A similar investigation of scaling effects for use of static panoramas was less successful. In an earlier study [26] we compared use of streaming video from a team or 4 robots with still panoramas taken by robots at their terminal waypoints. The panoramas were marked on the map and could be accessed asynchronously as the operator found time to search them for victims. In the streaming video condition operators found slightly more victims and marked them with somewhat greater accuracy. We speculated that with more robots we might find an advantage for panoramas because of the greater moment-to-moment demand of monitoring streaming video. An experiment comparing these conditions for 4, 8, and 12 robot teams, however, replicated our earlier findings and showed a small but persistent advantage for streaming video.

In our most recent experiment [27] pairs of participants controlled 24 robots (Figure 4) in either a dedicated condition in which each was assigned control over 12 or a call center condition in which they were jointly assigned control over 24. Results showed roughly comparable performance with slightly more area explored and victims found by participants in the dedicated condition. This experiment was intended as a control for studies in which we predict increased automation will alter the relative advantages.

IV. Discussion

USARSim was developed and remains an excellent platform for conducting HRI research. In this paper we have described 12 experiments conducted over 6 years using the USARSim platform. We are currently contributing to the UE3 port which we hope will lead to an even more effective experimental platform. While the choice of game engines was relatively easy in 2000 there is now a much broader range to choose from including open or inexpensive source alternatives. The advantages USARSim brings have shifted from the engine to the community. The true value of the simulation now lies in the substantial collection of models and validation data and our ability to share and maintain this common infrastructure.

References

[1] J. Gerstmann, "Unreal tournament: Action game of the year, 1999. GameSpot; www.gamespot.com/features/1999/p3_01a.html

[2] M. Abrash, "Quake's game engine: The big picture", Dr. Dobb's Journal (Spring 1977)

[3] J. Jacobson and Z. Hwang, "Unreal tournament for immersive interactive theater," *Communications of the ACM*, 45(1), 39-42, 2002.

[4] M. Lewis and J. Jacobson, "Game Engines in Scientific Research," *Communications of the ACM*, 45(1), 27-31, 2002.

[5] P. Prasithsangaree, J. Manojlovich, S. Hughes and M. Lewis, "UTSAF: A Multi-Agent-Based Software Bridge for Interoperability between Distributed Military and Commercial Gaming Simulation, Simulation", 80(12), 647-657, 2004.

[6] R. Calder, J. Smith, J. Courtemarche, J. Mar, and A. Ceranowicz. "ModSAF behavior simulation and control," *Proceedings of the Second Conference on Computer Generated Forces and Behavioral Representation*, STRICOM-DMSO, July, 1993.

[7] Karma Mathengine Karma User Guide http://udn.epicgames.com/Two/rsrc/Two/KarmaReference/KarmaUserGuide.pdf Accessed: 2009, 2002.

[8] D. McGovern, "Teleoperation of Land Vehicles". In Stephen Ellis (Ed.) *Pictorial Communications in Virtual and Real Environments*. New York: Taylor and Francis, 182-195, 1991.

[9] M. Lewis, J. Wang, J. Manojlovich, S. Hughes and X. Liu, X. "Experiments with Attitude: Attitude Displays for Teleoperation,". *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, 1345-1349, 2003.

[10] M. Lewis. and J. Wang, J. "Gravity referenced attitude display for mobile robots: Making sense of what we see," *Transactions on Systems, Man and Cybernetics Part A*, 37(1), 94-105, 2007.

[11] S. Hughes and M. Lewis, "Partially guided viewpoint control in virtual environments," *Human Factors*, 47(3), 630-643, 2005.

[12] H. Yanco, J. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *J. Human-Comput. Interact.*, vol. 19, no. 1–2, pp. 117–144, 2004.

[13] S. Hughes and M. Lewis, M. "Robotic camera control for remote exploration," *Proceedings of the 2004 Conference on Human Factors in Computing Systems* (CHI 2004), Vienna, Austria, 511-517, 2004.

[14] S. Hughes and M. Lewis, "Assisted viewpoint control for telerobotic search," *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 2657-2661, 2004.

[15] S. Desney, G. Tan,G. Robertson and M. Czerwinski: "Exploring 3D navigation: combining speed-coupled flying with orbiting," *Proceedings of the 2004 Conference on Human Factors in Computing Systems* (CHI 2001): 418-425, 2001.

[16] J. Wang, "USARSim: A game-based simulation of the NIST reference arenas," unpublished, 2004.

[17] J. Wang, M.Lewis, and J. Gennari, "A Game Engine Based Simulation of the NIST USAR Arenas," *Proceedings of the 2003 Winter Simulation Conference*, New Orleans, LA, (1) 1039-1045, 2003.

[18] J. Wang, and M. Lewis, "Autonomy in human-robot control," *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 525-529, 2006.

[19] J. Wang and M. Lewis, "Human control of cooperating robot teams," 2007 Human-Robot Interaction Conference, ACM, 9-16, 2007.

[20] B. Gerkey and M. Mataric. A formal framework for the study of task allocation in multi-robot systems. International Journal of RoboticsResearch, 23(9):939–954, 2004.

[21] J. W. Crandall, M. A. Goodrich, D. R. Olsen, and C. W. Nielsen. Validating human-robot interaction schemes in multitasking environments. IEEE Transactions on Systems, Man, and Cybernetics, Part A, 35(4):438–449, 2005

[22] J. Wang and M. Lewis, "Assessing coordination overhead in control of robot teams," *Proceedings of 2007 IEEE International Conference on Systems, Man, and Cybernetics*, 2645-2649, 2007.

[23] J. Wang, H. Wang and M. Lewis,."Assessing Cooperation in Human Control of Heterogeneous Robots," *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction* (HRI'08), 9-16, 2008.

[24] P. Velagapudi, P. Scerri, K. Sycara, K., H. Wang, M. Lewis and J. Wang, "Scaling effects in multi-robot control," *2008 International Conference on Intelligent Robots and Systems* (IROS'08), 2121-2126, 2008.

[25] H. Wang, M. Lewis, P. Velagapudi, P. Scerri and K. Sycara K. "How Search and its Subtasks Scale in N Robots," *Proceedings of the Forth ACM/IEEE International Conference on Human-Robot Interaction* (HRI'09),141-148, 2009

[26] P. Velagapudi, J. Wang, H. Wang, P. Scerri, M. Lewis, M., and K. Sycara, "Synchronous vs. Asynchronous Video in Multi-Robot Search," *Proceedings of first International Conference on Advances in Computer-Human Interaction* (ACHI'08), 224-229, 2008.

[27] H. Wang, S. Chien, M. Lewis, P. Velagapudi, P. Scerri, and K. Sycara, "Human teams for large scale multirobot control, *IEEE International Conference on Systems, Man, and Cybernetics* (to appear), 2009.