

# Synchronous vs. Asynchronous Control for Large Robot Teams

Huadong Wang<sup>1</sup>, Andreas Kolling<sup>1</sup>, Nathan Brooks<sup>2</sup>,  
Michael Lewis<sup>1</sup>, and Katia Sycara<sup>2</sup>

<sup>1</sup> School of Information Sciences, University of Pittsburgh,  
Pittsburgh, PA 15260 U.S.A.

<sup>2</sup> Robotics Institute, Carnegie Mellon University,  
Pittsburgh, PA 15213 U.S.A.

huw16@pitt.edu, akolling@pitt.edu, nbb@andrew.cmu.edu,  
ml@sis.pitt.edu, katia@cs.cmu.edu

**Abstract.** In this paper, we discuss and investigate the advantages of an asynchronous display, called image queue, for foraging tasks with emphasis on Urban Search and Rescue. The image queue approach mines video data to present the operator with a relevant and comprehensive view of the environment, which helps the user to identify targets of interest such as injured victims. This approach allows operators to search through a large amount of data gathered by autonomous robot teams, and fills the gap for comprehensive and scalable displays to obtain a network-centric perspective for UGVs. It is found that the image queue reduces errors and operator's workload comparing with the traditional synchronous display. Furthermore, it disentangles target detection from concurrent system operations and enables a call center approach to target detection. With such an approach, it could scale up to a larger multi-robot systems gathering huge amounts of data with multiple operators.

**Keywords:** Human-robot interaction, metrics, evaluation, multi-robot system, interface design, simulation.

## 1 Introduction

The task of interacting with multi-robot systems (MrS) especially large robot team presents unique challenges for the user interface designer, which is very different from the task of interacting with single or limited number of robots. Traditional graphical user interfaces and infrastructures have difficulties to interact with MrS. The core issue is one of scale: in a system of  $n$  robots, any task that has to be done to one robot must be done to the remaining  $n - 1$  [1]. Interface for large robot team, should use an infrastructure that allows remote action management and centralized display of data. Many applications such as interplanetary construction, search and rescue in dangerous environments, or cooperating uninhabited aerial vehicles have been proposed for MrS. Controlling these robot teams has been a primary concern of many HRI researchers. These efforts have included theoretical and applied development of the Neglect Tolerance model and Fan-out model to characterize the control of independently operating robots [1, 2], predefined rules to coordinate

cooperating robots as in Playbook™ [3] and Machinetta [4], and techniques for influencing teams obeying biologically inspired control laws [5, 6, 7]. While our efforts to increase span of control over unmanned vehicle (UV) teams appear to be making progress, the asymmetry between what we can command and what we can comprehend is growing. Automation can reduce excessive demands for human input, but throttling the information being collected and returned is fraught with danger. A human is frequently included in the loop of a MrS expressly to monitor and interpret video being gathered by UVs. This can be a difficult task for even a single camera [8] and begins exceeding operator capability before reaching ten cameras [9, 10]. With increasing autonomy of robot teams and plans for biologically inspired swarms of much greater size the problem of absorbing and benefiting from their product seems even more pressing than learning how to command them.

Foraging tasks, carried out with a large robot team, require an exploration that needs to be more than simply moving each robot to different locations in the environment. Acquiring a specific viewpoint of targets of interest, e.g. victims in a disaster scenario, is of greater concern and increasing the explored area is merely a means to this end. While a great deal of progress has been made for autonomous, exploration the identification of targets is still done by human operators who ensure that the area covered by robots has in fact been thoroughly searched for the desired targets. Without means to combine the data gathered by all robots the human operator is required to synchronously monitor their output, such as a video feed for each robot. This requirement and load on the human operator directly conflicts with other tasks, especially navigation which requires the camera to be pointed in the direction of travel in order to detect and avoid objects. The need to switch attention among robots will further increase the likelihood that a view containing a target will be missed. Earlier studies [11, 12] confirmed that search performance on these tasks is directly related to the frequency with which the operator shifts attention between robots, possibly due to targets missed in the video stream while servicing other robots.

The problem addressed in this paper is the design of an asynchronous, scalable, and comprehensive display, without requiring a 3d reconstruction, to enable operators to detect relevant targets in environments that are being explored by large teams of UGVs. We will present one particular design for such a display and test it in the context of USAR with large robot teams with some degree of autonomy and supervised by a single operator.

## 2 Asynchronous Imagery

### 2.1 Pilot Studies

An asynchronous display method can alleviate the concurrent load put on the human operator and disentangle the dependency of tasks that require the video feed. Furthermore, it can avoid attentive sampling among cameras by integrating multiple data streams into a comprehensive display. This in turn allows the addition of new data streams without increasing the complexity of the display itself. A first approach for an asynchronous display is explored in [13]. The method therein is motivated by asynchronous control techniques previously used in extraterrestrial NASA

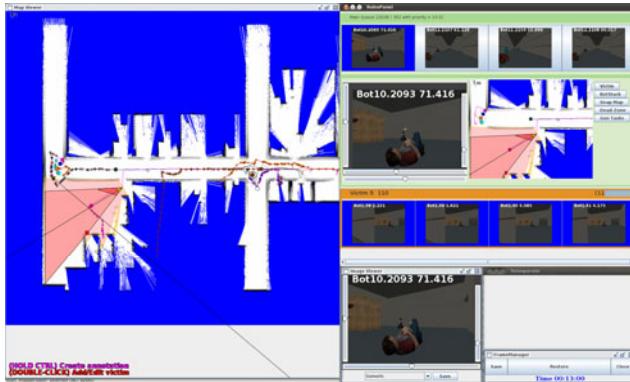
applications. These are faced with limited bandwidth and communication lags due to large interplanetary distances. Such lags make a direct operation and viewing of all outputs impossible. Instead, the robot team is instructed to gather information only at specific locations in the form of panorama images from an omnidirectional overhead camera. The operator then searches through all panorama images and determines the location of specific targets. The approach was tested in experiments which compared performance for operators controlling four robots in a team using streaming or asynchronous panorama displays. No significant improvement was found in terms of found victims, but the frequency of shifting focus between robots was correlated with performance for streaming video but not for asynchronous panoramas. As expected, the asynchronous display of information alleviates the need for excessive switching. It was conjectured that with larger robot teams the benefit of avoiding attention switching will increase and have an effect on performance in terms of victims. Further experiments in [14] scaling the team size to eight and twelve robots, however, found no further significant improvements. But this approach did not utilize all the available data from the video feeds that robots gather, so a huge amount of potentially useful information was discarded for the panorama condition. Furthermore, the operator needs to give the robots additional instructions where to sample panoramas.

In contrast to previous work our approach in this paper allows the use of autonomous exploration and we present an asynchronous display that mines all of the robot video feeds for relevant imagery. This imagery is then given to the operator for analysis. We coin this type of asynchronous display image queue and compare it to the traditional synchronous method of streaming live video from each robot (streaming video). In the next section we describe our test bed and followed by a detailed description of the image queue and a comparison with streaming video.

## 2.2 Image Queue Interface

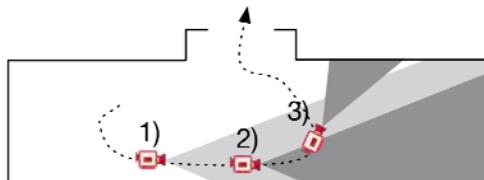
The goal of the image queue interface is to best utilize the advantages of an asynchronous display and to maximize the amount of time human operators can spend on tasks which human's performs better than robots. Currently, for USAR, this is the case for tasks such as victim identification and navigating robots out of dangerous areas in which they got stuck. As the number of robots in a system increases with improved autonomy the demands on operators for these tasks increase as well. Hence, another requirement for the interface is to provide the potential for scaling to larger numbers of robots and operators. The proposed image queue interface implements the idea of asynchronous monitoring via a priority queue of images that allows operators to identify victims requiring neither synchronicity nor any contextual information not directly provided by the image queue.

The image queue interface (Fig. 1) focuses on two tasks: (1) viewing imagery and (2) localizing victims. It consists of a filmstrip viewer designed to present the operator with a filtered view of what has passed before the team's cameras. A filtered view is beneficial because the video taken contains a high proportion of redundant images from sequential frames and overlapping coverage by multiple robots.



**Fig. 1.** GUI for the image queue condition

The filter attempts to reduce redundancy by only showing highly relevant images from the video stream. Relevance is scored by computing a utility for every image that determines its priority in the queue displayed in the filmstrip viewer. To achieve this we store every frame from all video streams in a database together with associated robot poses and laser scans taken at the time of capture. From this database we can retrieve any image and compute its utility. The computation of utility can be adapted to a particular application and for our experiment we computed utility via the area covered seen in an image. This visual coverage is computed by referencing the image in the map as seen in Fig. 2. Images with larger areas receive higher utility scores. Areas that have already been seen by other images in the filmstrip viewer do not count towards utility. In colloquial terms this kind of utility picks images that cover large areas with minimal overlap. Fig. 2 illustrates this concept of utility with a simple example.



**Fig. 2.** An illustration of the utility of individual frames from a video stream. The frame taken at 1) has the largest visual coverage and highest utility while the frame at 2) has no utility since it is entirely overlapped by 1). Frame 3) has some utility since it provides coverage in an area not covered by 1).

By aggregating imagery with the highest utility scores at regular intervals the image queue allows the operator to peruse a relatively small number of prioritized images that show most of the new area explored by the robots. Notice that exploration can continue while operators view the image queue so long as robots are sufficiently autonomous (or controlled by other operators). Operators can either click through or

scroll through a certain number of images in the queue. Once operators work through the first set of images the image queue marks the areas covered by these images as already seen and retrieves the next set of images with high utility. Tests of this system show that after 15 minutes of exploration an operator can view 70% of the area covered by viewing the 10 highest utility frames and 90% within the first 100 frames.

### 3 Methods

#### 3.1 USARSim and MrCS

The experiment reported in this paper was conducted using the USARSim robotic simulation with 12 simulated Pioneer P3-AT robots performing Urban Search and Rescue (USAR) foraging tasks. USARSim is a high-fidelity simulation of urban search and rescue (USAR) robots and environments developed as a research tool for the study of human-robot interaction (HRI) and multi-robot coordination. 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. USARSim also serves as the basis for the Virtual Robots Competition of the RoboCup Rescue League.



**Fig. 3.** GUI for the streaming video condition

MrCS (Multi-robot Control System), a multi-robot communications and control infrastructure with accompanying user interface, developed for experiments in multirobot control and RoboCup competition [15] was used in this experiment. MrCS provides facilities for starting and controlling robots in the simulation, displaying multiple camera, and laser output, as well as maps, and supporting inter-robot communication through Machinetta, a distributed multi-agent coordination infrastructure. Fig. 3 shows the elements of the conventional GUI for the streaming video condition. The operator selects the robot to be controlled from the colored thumbnails with live videos at the top right of the screen. The current locations and paths of the robots are shown on the Map Viewer (bottom left). Under manual

control, robots are tasked by assigning waypoints on a heading-up map on the Map Viewer or through the teleoperation widget (lower right).

An autonomy path planner was used in the current experiment to drive the robots unlike the panorama study [13] in which paths were generated manually by participants with specified panorama locations. As the previous study [14, 17], operators appeared to have little difficulty in following these algorithmically generated paths and identified approximately the same numbers of victims (per operator) as in following human generated paths. A new Segment Voronoi Diagram (SVD) path planner replaced the random tree planner used in earlier studies. The new planner generates paths that maintain a safe distance to nearby obstacles. Such paths are generally longer, smoother and more human-like.

### 3.2 Experimental Design

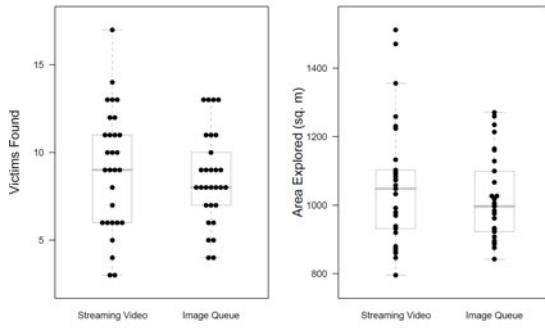
The experiment followed a two condition repeated measures designs comparing the conventional MrCS displays (streaming video) with MrCS augmented by the experimental image queue display counterbalancing conditions. Automated path planning to improve search performance and autonomous exploration was used in both conditions. The operators performed a supervisory control task in which the robots navigated autonomously with the operator allowed to override by directing them through new waypoints. When necessary, participants were able to teleoperate the in-focus robot to extricate it when it became stuck.

32 paid participants were recruited from the University of Pittsburgh community balanced among conditions for gender. None had prior experience with robot control although most were frequent computer users.

After providing demographic data and completing a perspective taking test participants read standard instructions on how to control robots via MrCS. In the following training session, participants practiced control operations for both streaming video and image queue condition for 10 min each. Participants were encouraged to find and mark at least one victim in the training environment under the guidance of the experimenter. After the training session, participants began the two 15 minute real task sessions in which they performed the search task controlling 12 robots in teams using either the streaming video or image queue display with a counterbalanced design. At the conclusion of each real task session, participants were asked to complete the NASA-TLX workload survey [16].

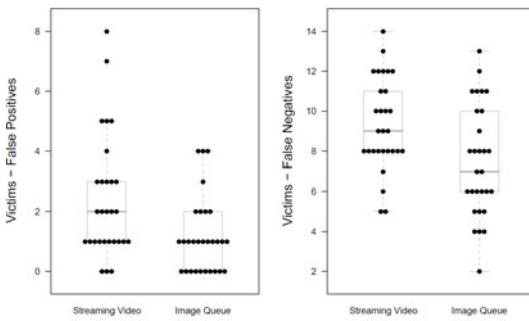
## 4 Result

Data were analyzed using a repeated measures ANOVA comparing streaming video with the image queue condition. Overall, in both conditions participants were successful in searching through the environment. On average participants in the streaming video condition found 9.10 victims while those in the image queue condition found 8.51 (Fig. 4) without a significant difference between conditions ( $F_{1,28} = .733$ ,  $p = .387$ ). In addition, as Fig. 4 shows, area explored by the 12 robots was also showed a none significant advantage for the streaming video condition, ( $F_{1,28} = 2.147$ ,  $p = .154$ ).



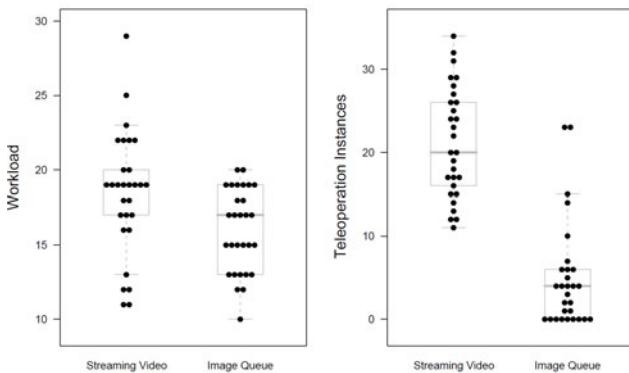
**Fig. 4.** Victims found and area explored

Victim marking errors have been classified into two types, false positive and false negative. A mark without any victim or multiple marks for one victim were counted as a false positive. The number of false positive shows a significant advantage for image queue condition than the streaming video condition ( $F_{1,28} = 13.032$ ,  $p = .001$ ). On the other hand, Victims that were missed, but present in the video feed, and not marked were calculated as a false negative. The image queue shows a similar significant improvement for the missing victim comparing with the streaming video condition( $F_{1,28} = 5.526$ ,  $p = .026$ ), with the average number of missing victims in the image queue condition dropping to 7.48 while participants in the streaming video condition missing 9.34 victims respectively (Fig. 5).



**Fig. 5.** Marking errors of victims

To more closely measure the subjects behavior linked to their monitoring and operation, teleoperation times (Fig. 6) were analyzed. The repeated measures ANOVA shows a significant difference for the count of teleoperation times between the streaming video and image queue condition, with participants in the streaming video condition, teleoperate average 21.24 times, which is significantly more than the participants in the streaming video condition, teleoperate 4.97 times respectively ( $F_{1,28} = 150.719$ ,  $p < .001$ ).



**Fig. 6.** Teleoperation and workload

The full scale NASA-TLX workload measure also revealed a significant difference, which is unlike the earlier studies [28] where no advantage was found for panorama pictures GUI. A significant advantage in workload ( $F_{1,28} = 7.347$ ,  $p = .001$ ) was observed favoring the image queue condition (Fig. 6). Examining individual dimension of workload comparing streaming video and image queue we found significant differences for temporal demand ( $F_{1,28} = 6.503$ ,  $p = .016$ ) and effort ( $F_{1,28} = 4.576$ ,  $p = .040$ ).

## 5 Discussion and Future Works

The purpose of this experiment was to examine the asynchronous image queue interaction method to control a large human robot teams. Our experiment result shows that under image queue conditions, which allowing interruption, relevance image retrieval, reviewable, location based image queue interface leads to somewhat equal search performance with lower errors and lower workload.

In the image queue condition humans were able to view the image queue and control 12 robots each, to mark the equal number of victim and to cover the equal area as the streaming video condition participants. This result supports the feasibility of our image queue interaction mode as a substitute for human operators (performance is no worse). The analysis of victim marking errors shows that in the streaming video condition participants marked 89.3% more victims at the wrong location and missed 23.1% more victims than image queue condition. This gain for image queue is particularly significant because avoiding missed targets is crucial to many foraging tasks, thoroughness and correctness is more important than other performance.

Similar improvements in reported workload suggest that substantial cognitive resources were required for streaming video condition. Based on the analysis result of teleoperation instance, participants in streaming video condition are more likely to intervene the robot for path planning and problem solving. Participants in streaming video condition have teleoperate 4.8 times more than those in image queue condition during the whole task.

Participants in the streaming video condition were confronted with a bank of videos (Fig. 3) much like a security guard monitoring too many surveillance cameras. Informal observation of participants suggested that due to the frequently distraction of robot operation requirement, victim running into the camera and task switch in between, operator puts great effort for task allocation and feels intensive time pressure. The significant difference of individual dimension of workload, temporal demand and effort, supports this observations.

Furthermore, 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 big team sizes with lower errors and workload. When people doing search tasks under stressed or heavily loaded conditions spontaneous self-synchronization is unlikely to arise and hence full shared information. Instead, image queue provided a benefit way of task division between operators, in which exploration and perceptual search (identifying targets) task could be neatly deconstructed to different operators. Framing the problem this way leads to the design conclusion that operators should be issuing task-centric commands. To realize this kind of control architecture we will propose a call center approach in which some operators address independent control needs for monitoring and exploration of UVs, whereas other operators address independent location based images in queue for victim marking and other perceived tasks.

Once we are facing the problem of managing 12 or more robot teams or in other words, the large scale systems, the call center design of the control architecture with image queue embedded will be proposed to solve these problems. Then information sharing problem with people's losing of situation awareness will be one of the priorities of our concern. Just imaging three operators controlling more than 30 UVs with some UV originated requests such as verifying/marketing targets or alerting operator on low fuel and other UV needs such as veering off path or becoming bogged down by terrain (trees in a forest for example) or similar that operators' must monitor to detect; as a result the image queue interaction mode may still be a good choice for people to share information or recall for their situation awareness.

According to these concerns, we want to explore the effect the combination of these two modes of control in large number of robots with multiple operators, in which under a call-center structures exploration task will be operated in the streaming video mode and perceptual search task will be performed in the image queue mode, and it could be useful in preparing to study larger scale systems in the future.

**Acknowledgments.** This research has been sponsored in part by AFOSR FA955008-10356 and ONR Grant N0001409-10680.

## References

1. Olsen, D., Wood, S.: Fan-out: Measuring Human Control of Multiple Robots. In: Human Factors in Computing Systems (CHI 2004). ACM Press, Vienna (2004)
2. Crandall, J., Goodrich, M., Olsen, D., Nielsen, C.: Validating human-robot interaction schemes in multitasking environments. IEEE Transactions on Systems, Man, and Cybernetics, Part A 35(4), 438–449 (2005)

3. Miller, C., Parasuraman, R.: Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors* 49(1), 57–75 (2007)
4. Scerri, P., Liao, E., Lai, L., Sycara, K., Xu, Y., Lewis, M.: Coordinating very large groups of wide area search munitions. *Theory and Algorithms for Cooperative Systems*
5. Kira, Z., Potter, M.: Exerting Human Control Over Decentralized Robot Swarms. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2010 (2010)
6. McLurkin, J., Smith, J., Frankel, J., Sotkowitz, D., Blau, D., Schmidt, B.: Speaking Swarmish: Human-Robot Interface Design for Large Swarms of Autonomous Mobile Robots. In: AAAI Spring Symposium (2006)
7. Ding, X., Powers, M., Egerstedt, M., Young, S., Balch, T.: Executive Decision Support: Single-Agent Control of Multiple UAVs. *IEEE Robotics & Automation Magazine* 16(2), 73–81 (2009)
8. Cooke, N., Pringle, H., Pedersen, H., Connor, O. (eds.): *Human Factors of Remotely Operated Vehicles*. Elsevier, Amsterdam (2006)
9. Humphrey, C., Henk, C., Sewell, G., Williams, B., Adams, J.: Assessing the Scalability of a Multiple Robot Interface. In: 2nd ACM/IEEE International Conference on Human-Robotic Interaction (2007)
10. Lewis, M., Wang, H., Chien, S., Velagapudi, P., Scerri, P., Sycara, K.: Choosing autonomy modes for multirobot search. *Human Factors* 52(2), 225–233 (2010)
11. Pepper, C., Balakirsky, S., Scrapper, C.: Robot Simulation Physics Validation. In: PerMIS 2007 (2007)
12. Taylor, B., Balakirsky, S., Messina, E., Quinn, R.: Design and Validation of a Whegs Robot in USARSim. In: PerMIS 2007 (2007)
13. Velagapudi, P., Wang, J., Wang, H., Scerri, P., Lewis, M., Sycara, K.: Synchronous vs. Asynchronous Video in Multi-Robot Search. In: First International Conference on Advances in Computer-Human Interaction, ACHI 2008 (2008)
14. Velagapudi, P., Wang, H., Scerri, P., Lewis, M., Sycara, K.: Scaling effects for streaming video vs. static panorama in multirobot search. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS (2009)
15. Robocup Rescue VR,  
[http://www.robocuprescue.org/wiki/index.php?title=VRCompetitions#Singapore\\_2010](http://www.robocuprescue.org/wiki/index.php?title=VRCompetitions#Singapore_2010)
16. Hart, S., Staveland, L.: Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 139–183. Elsevier, Amsterdam (1998)
17. Lewis, M., Wang, H., Chien, S., Velagapudi, P., Scerri, P., Sycara, K.: Choosing autonomy modes for multirobot search. *Human Factors* 52(2), 225–233 (2010)