Asynchronous Visualization of Spatiotemporal Information for Multiple Moving Targets

by

Huadong Wang

B.S. in Measurement, Control technology and Instruments, Tsinghua University, P.R.C., 2004

M.S. in Management Science and Engineering, Tsinghua University, P.R.C., 2007

Submitted to the Graduate Faculty of

The School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Huadong Wang

2013

and approved by

Hassan Karimi, Professor, School of Information Sciences

Illah R. Nourbakhsh, Professor, The Robotics Institute, Carnegie Mellon University

Stephen C. Hirtle, Professor, School of Information Sciences

Vladimir Zadorozhny, Associate Professor, School of Information Sciences

Dissertation Advisor: Michael Lewis, Professor, School of Information Sciences

Asynchronous Visualization of Spatiotemporal Information

for Multiple Moving Targets

Huadong Wang, [PhD]

University of Pittsburgh, 2013

In the modern information age, the quantity and complexity of spatiotemporal data is increasing both rapidly and continuously. Sensor systems with multiple feeds that gather multidimensional spatiotemporal data will result in information clusters and overload, as well as a high cognitive load for users of these systems.

To meet future safety-critical situations and enhance time-critical decision-making missions in dynamic environments, and to support the easy and effective managing, browsing, and searching of spatiotemporal data in a dynamic environment, we propose an asynchronous, scalable, and comprehensive spatiotemporal data organization, display, and interaction method that allows operators to navigate through spatiotemporal information rather than through the environments being examined, and to maintain all necessary global and local situation awareness.

To empirically prove the viability of our approach, we developed the Event-Lens system, which generates asynchronous prioritized images to provide the operator with a manageable, comprehensive view of the information that is collected by multiple sensors. The user study and interaction mode experiments were designed and conducted. The Event-Lens system was discovered to have a consistent advantage in multiple moving-target marking-task performance measures. It was also found that participants' attentional control, spatial ability, and action video gaming experience affected their overall performance.

TABLE OF CONTENTS

PREFACE XII

1.0 INTRODUCTION 1

1.1 PROBLEM STATEMENT 2

1.1.1 Data Overload Problem for Multiple Sensor Systems 3

1.1.2 Strategies with Moving Targets in Dynamic Environments 4

1.1.3 Situation Awareness of Spatiotemporal Objects 5

1.2 APPROACH OUTLINE 6

1.3 THESIS STRUCTURE 7

2.0 RELATED WORK 8

2.1 MULTIPLE FEEDS SENSOR SYSTEM 8

2.2 MULTI-ROBOT SYSTEM 10

2.2.1 Multiple Feeds Problem in Multi-robot Systems 10

2.2.2 Interaction with multi-robot systems 12

2.2.3 Information Fusion for Multi-robot Systems 15

2.3 SPATIOTEMPORAL INFORMATION FUSION 17

2.3.1 Spatial and Temporal Data Visualization and Integration 17

2.3.2 Spatiotemporal Data Organization 24

2.4 ASYNCHRONOUS DISPLAY 30

2.4.1 Panorama Pictures 30

2.4.2 Image Queue 32

2.5 ANALYZING MULTIPLE MOVING TARGETS 36

2.5.1 Spatiotemporal Objects 37

2.5.2 Movement Analysis Task 39

2.5.3 Visual Analytics of Multiple Moving Targets 40

3.0 DESIGN OF THE SYSTEM 43

3.1 SCENARIOS AND MOTIVATION 43

3.2 SYSTEM ARCHITECTURE AND OVERVIEW 46

3.3 DATABASE DESIGN 49

3.4 PRIORITY QUEUE 51

3.5 SPATIOTEMPORAL INFORMATION INTERACTION 52

3.5.1 Overall Assessment 52

3.5.2 Temporal Navigation 53

3.5.3 Spatial Navigation 54

3.6 MARKING OF MULTIPLE MOVING TARGETS 56

4.0 METHODS 59

4.1 USARSIM 59

4.2 MRCS – THE MULTI-ROBOT CONTROL SYSTEM 62

4.3 SYSTEM AUTONOMY 64

4.4 IMAGE PROCESSING IN IMAGE FILTER 68

4.4.1 Image Utility Calculation 68

4.4.2 Spatial and Temporal Uniqueness in Sub-queues 70

4.4.3 Pre-Index of the Images ... 72
4.5 INDIVIDUAL DIFFERENCES ... 74
4.5.1 Attentional Control ... 74
4.5.2 Spatial Ability ... 76
4.5.3 Action Video Gaming Experience ... 79
5.0 EVALUATION ... 82
5.1 APPARATUS ... 83
5.2 BEHAVIOR ANALYSIS USER STUDIES ... 84
5.2.1 Situation Awareness Survey Question Design ... 84
5.2.2 Participants ... 86
5.2.3 Procedure ... 86
5.2.4 Results ... 88
5.2.5 Discussion ... 93
5.2.5.1 Strategy of Searching and Tracking Moving Target ... 93
5.2.5.2 Impact of Individual Differences ... 97
5.3 INTERACTION MODE EXPERIMENT ... 100
5.3.1 Hypothesis ... 100
5.3.2 Experiment Design ... 101
5.3.3 Participants ... 102
5.3.4 Procedure ... 103
5.3.5 Performance Metrics ... 104
5.3.6 Results ... 104
5.3.7 Discussion ... 108

5.3.7.1 Synchronous Control vs. Asynchronous Control 108
5.3.7.2 Identification of the Moving Target 111
5.4 GENERAL DISCUSSION 113
6.0 CONCLUSIONS 115
6.1 THESIS CONTRIBUTION 115
6.2 FUTURE WORK 116
APPENDIX A 119
APPENDIX B 121
APPENDIX C 123
APPENDIX D 125
BIBLIOGRAPHY 126

LIST OF TABLES

Table 1. Task type and example SA questions 85

Table 2. Categories of Moving Target Search and Tracking Strategies 94

Table 3. Experiment Design 102

Table 4. Target marking errors and effects of target identifiability 106

LIST OF FIGURES

Figure 1. An example of the interface for multi-robot system 10

Figure 2. Space–time cube 18

Figure 3. 2012 Olympic Torch Relay Route analyzed by GeoTime 19

Figure 4. Demo of Microsoft’s Photosynth Software 20

Figure 5. Major components of STV 21

Figure 6. Overview of PhotoScope’s main screen. 22

Figure 7. User interface of the MERL forensic surveillance system. 23

Figure 8. Differently scaled time view of the Timeline browser 25

Figure 9. The Time Quilt layout “weaves” a linear timeline into columns of photo clusters 26

Figure 10. The reconstruction of 3D points and viewpoints to browsing the photos 27

Figure 11. The UI layout of map-based storyboard with the location controller 28

Figure 12. Advanced View Modes: pre-process the input video 28

Figure 13. An example of the explored map for panorama picture experiment 31

Figure 14. GUI for Image Queue interface 33

Figure 15. Image Queue GUI with automatic target recognition 34

Figure 16. Example for ATR group map 35

Figure 17. Characteristics of objects, locations, and times 38

Figure 18. Visual analytics for the detection of anomalous maritime behavior 41

Figure 19. Time graph display of the temporal variations of the crime rates 42

Figure 20. Pick and Drop interaction in FromDaDy (Hurter et al., 2008) 42

Figure 21. An illustration of the system architecture for Event-lens 47

Figure 22. Prototype of GUI for the Event-Lens interface 48

Figure 23. P3AT robot (left) with laser scan area (light pink) and camera viewed area (dark pink) showed on the map right 50

Figure 24. Example database with an empty total shape containing images 1, 2 and 3 51

Figure 25. Top utility image list with visual coverage 53

Figure 26. The toolbar of timeline 54

Figure 27. An illustration showing the sub-queue of a user's selected area 55

Figure 28. Target identification and marking interaction 57

Figure 29. An illustration of target trajectory analysis function 58

Figure 30. USARSim architecture 61

Figure 31. MrCS architecture 63

Figure 32. The conventional MrCS user interface for the streaming video condition 64

Figure 33. A sample map with the structures used and computed by the autonomy 66

Figure 34. FOV of the laser scanner (180 degree) and RGB camera (45 degree) 67

Figure 35. Example database with a blue total shape after selecting image 1. 68

Figure 36. An illustration of how to calculate redundant imagery 69

Figure 37. Visualization of images collected in a region 70

Figure 38. Image filtering process in sub-queue 71

Figure 39. Visualization of images collected in a region 73

Figure 40. Example of the spatial orientation test (Hegarty & Waller, 2004) 78

Figure 41. Target marking performance and effects of operator's spatial ability (SpA) 88

Figure 42. Target marking errors and effects of operator's spatial ability (SpA) 89

Figure 43. Situation awareness queries and effects of operator's spatial ability (SpA) 90

Figure 44. Situation awareness queries and effects of operator's video game experience 91

Figure 45. Perceived workload and effects of perceived attentional control (PAC) 92

Figure 46. Effects of interaction methods (left) and target identifiability (right) on target marking performance 105

Figure 47. Effects of interaction methods (left) and target identifiability (right) on perceived workload 106

Figure 48. Effects of interaction methods (left) and target identifiability (right) on average marking time 107

Figure 49. Effects of target identifiability on number of target marked per image viewed 108

Figure 50. Target Marking Procedure Analysis for Operator with streaming video and Event-Lens conditions 109

PREFACE

First, I would like to express my deepest gratitude to my advisor and my mentor, Dr. Michael Lewis, for his intelligence, support and guidance in my career as a graduate student, for his encouragement and patience when my confidence faltered, for his suggestions on handling seemingly impossible situations, and, most of all, for his friendship and advice on life. Achieving one's PhD is a challenging journey physically, mentally and emotionally. Without him, I will never reach the other side.

Thanks to my colleagues in School of Information Sciences and Computer Science Department at the University of Pittsburgh and in the Robotics Institute at Carnegie Mellon University for their hard work, creativity, support, and encouragement. This work would not have been possible without you: Boyou Sun, Nathan Brooks, Shih-Yi James Chien, Dr. Prasanna Velagapudi, Huichao Xue, Dr. Andreas Kolling, and Pei-Ju Lee. Special thanks to all my friends from the Computer Science Department at University of Pittsburgh; their friendship made this experience enjoyable.

Thanks also to the members of my committee, for their effort and time in guiding me. Thanks to Dr. Karimi for his seminar class in my research area; thanks to Dr. Nourbakhsh for his expertise and for offering different perspectives; thanks to Dr. Hirtle, for all his inspirational classes and books; and thanks to Dr. Zadorozhny, for his professional insights on databases, which has helped me sharpen my research problem from beginning. I am immensely grateful and deeply indebted to them.

Finally, I would like to thank my family for their emotional support and understanding throughout my graduate studies. I am forever grateful to my parents and grandparents for their love, encouragement, unwavering confidence in me, and, foremost, for giving me the opportunity to explore a new world. Most importantly, special thanks to my wife, Xiaohua Hu, for her faith in me and giving me unending encouragement and support, throughout our six-year separation in two states.

1.0 INTRODUCTION

In the modern information age, the quantity and complexity of spatiotemporal data is rapidly and continuously increasing. The ability to specify and visualize spatiotemporal events is fundamental to reasoning and to making critical decisions in many scenarios. For example, when there is a disaster, such as an earthquake or a fire, the most urgent task is that of searching and rescuing (SAR) victims. Urban search and rescue (USAR), however, is a prime example of a class of mass information, safety-critical situations, which are situations in which a run-time error or failure could result in death, injury, loss of property, or environmental harm (Leveson, 1986). Safety-critical situations, which are usually also time-critical, provide one of the bigger challenges for robot designers, due to the requirement that robots must perform exactly as intended and support humans in efficient and error-free operations (Yanco et al., 2004).

Due to the extensiveness of the USAR domain, there may be many spatiotemporal information items presented in an interface visualization, which may lead to clutter and cognitive overload. As a result, spatiotemporal information items need to be abstracted or managed in order to be presented in a manner that minimizes cognitive load. It is essential to identify how spatiotemporal information items relate to each other both spatially and temporally, and how to transform or abstract the information items into a single, more coherent information item.

Second, in most real-life situations, the required information is dynamic and contains two levels. First, the environment is dynamic. For example, in a fire scenario, parts of a structure's wall may collapse, which may change the floor plan and affect the planned route of the robots that are providing assistance. Second, the target of rescue itself is dynamic. For example, a victim's status may change, which would be a temporal dynamic, while some of the victims may move to other places, which would be a spatial dynamic. More types of situations may have information in both of these categories. As a result, providing visual cues for this type of dynamic information is not only required, but vital to these efforts, and as a result, it is necessary to introduce a multi-dimensional visualization method to USAR operations.

The main problem of designing such systems is finding a way to present this highly multidimensional data, which includes both spatial and temporal components, to a human operator in order to allow browsing and searching of recorded data in an efficient and intuitive way. To meet future safety-critical situations and enhance time-critical decision-making missions in dynamic environments, such as those found in USAR; to support the easy and effective managing, browsing, and searching of spatiotemporal data in a dynamic environment; to help operators to navigate through the information rather than through the environments; and to maintain necessary global and local situation awareness, we want to design and evaluate an asynchronous interaction approach and relative interface that visualizes the spatiotemporal data for multiple moving targets.

1.1 PROBLEM STATEMENT

Numerous research efforts have focused on helping users to extract, analyze, and understand spatiotemporal data. Visualization is an external mental aid that enhances cognitive abilities and is the most commonly used technique. When information is presented visually, efficient innate human capabilities can be used to perceive and process data. A greater amount of information can be seen and understood in a few minutes by a human as compared to a robot (Becker, Cleveland & Wilks, 1987). Information visualization techniques amplify cognition by increasing human mental resources, reducing search times, improving recognition of patterns, increasing inference making, and increasing monitoring scope (Ware, 2000). These benefits translate into system- and task-related performance factors for individuals and groups, which both affect the completion of analysis, decision-making, and communication tasks. The time, effort, and number of work products required to do these types of tasks are all reduced (Wright & Kapler, 2003). Visualization is also a great aid in data analysis processes, which are often designed to efficiently map data to visual primitives and present them in an effective way so that users can easily discern patterns (Jin, 2009).

However, simply “seeing” the data, or even “seeing” the pattern within data, is far from enough. Even though many visualization methods have built-in navigation tools to help users explore spatiotemporal data, such as zooming or scrolling, testing and analyzing are done through visual inspection, which is error-prone, inefficient, and not reusable. The task becomes more cumbersome when data relationships become more complex, such as when users must incorporate multiple moving sensor networks. As a result, our design and the exploratory study are centered on the following three categories of research questions:

1.1.1 Data Overload Problem for Multiple Sensor Systems

Woods et al. (2002) have characterized three basic data overload problems: clutter, workload bottleneck, and finding the significant data. Unfortunately, our research faces all of these problems:

- Clutter problem: We must deal with multidimensional data from multiple sensor systems such as videos, images, geographical information, time stamps, and others, which may lead to too much information on the screen. As a result, the research questions remain: What kind of visualization method should be used for viewing complex spatiotemporal information? Which of the displayed components could be reduced?
- Workload bottleneck: We face a difficult scheduling problem, as there are too many tasks to perform in the time available when users control the multiple sensor system in a synchronous way. Instead, we want to explore the asynchronous display method, which can alleviate the concurrent load put on human operators and can disentangle the dependency of tasks that require direct attention to multiple video feeds. In other words, the asynchronous display method will turn force-paced tasks into self-paced tasks and will also lower the user's mental workload.
- Finding the significance of data: A large scale of data not only means the prior data is not known, but it also means that a huge amount of redundant information can hinder efficient interactions. Consequently, how to reduce this data overload and these redundancies by providing spatially and temporally unique information that is of maximum utility will be the following research question. An information filter will be designed for the multiple sensor system.

1.1.2 Strategies with Moving Targets in Dynamic Environments

We want to explore and understand the operators' cognitive process of spatiotemporal data. The questions "What are strategies that users use in dealing with spatiotemporal data? And how did they locate the relevant events?" are critical to the design of these types of interaction procedure and interface displays. Without a deep understanding of the information-seeking tasks and goals of the consumer of the interface, it is impossible to develop a system that is truly useful and informative. A careful understanding of the perceptual and cognitive capabilities of people is necessary to develop effective information visualizations (Spence, 2007). We also expect that this knowledge would contribute to usability requirements for future systems.

1.1.3 Situation Awareness of Spatiotemporal Objects

How can we create effective abstract and intuitive spatiotemporal representations that help users maintain awareness of key information and actions about an evolving crisis situation that occurs in geographic space over time? Furthermore, how can we help users to freely manipulate spatial and temporal data and smoothly navigate through these scenarios?

A great deal of research was focused on the development of visualization and analysis tools for spatiotemporal data. However, the way in which a user interacts with spatiotemporal data regarding the moving target is rarely considered. How can specific visual representations and tools be used to support the user in navigating, exploring, and even manipulating the spatiotemporal data in dynamic environments?

Ivanov et. al (2007) have explained that most interaction approaches are "sensor-centric: they present the world as it is seen through a sensor—say, a camera." However, we lack experience with such unfamiliar views that involve a remote control camera. Consequently, in computer vision, before an event can be analyzed and understood, "it is often a challenge to normalize pose, scale, or illumination, or to find a video that contains some event in its entirety."

We want to facilitate usability requirements gathering for the design of future spatiotemporal interfaces. Therefore, we propose possible design guidelines and general implications for future visualization technologies to support tasks that involve the use of spatiotemporal data.

1.2 APPROACH OUTLINE

This thesis presents an asynchronous, scalable, and comprehensive spatiotemporal data organization, display, and interaction method, which allows operators to navigate through the spatiotemporal information rather than the environments, as well as to maintain all necessary global and local situation awareness (SA). The research will focus on the design considerations for the entire procedure and will seamlessly bring together all components with an overall consideration of the interface, visualization, and interaction.

To empirically prove the viability of our approach, we developed the Event-Lens system, which generates asynchronous prioritized images to provide the operator with a manageable but comprehensive view of the information collected by multiple sensors. The Event-Lens system uses a computational geometry algorithm to calculate the utility of images, and solves the data overload and redundancy problems by providing spatial and temporal unique information with top utilities.

A system evaluation was conducted, as we want to show that the Event-Lens system represents a significant improvement over the state of the art towards solving the multiple-feeds problem of multiple-sensor systems for dynamic environments.

The Event-Lens system also provides multiple interaction methods for users to manipulate spatiotemporal data. A user study was conducted that included open-question tasks, which encouraged exploration through the Event-Lens interface and the environment. By doing this, we learned more about users' spatial and temporal search strategies and explored possible requirements and design guidelines for the future. The experiment was also designed to compare our new approach with conventional synchronous approaches to tasks that involved different types of multiple moving targets.

1.3 THESIS STRUCTURE

This remainder of the thesis is structured as follows: related work is first described in Chapter 2, showing our asynchronous interaction procedure that assimilates existing systems and technologies in multiple dimensions. Chapter 3 describes how each component in this asynchronous interaction procedure works, followed by implementation detail for the multi-robot systems (MrS) domain in Chapter 4. Chapter 5 shows user studies and experiment design, setups, results, and discussions. A more general conclusion and possible avenues for future work is summarized in Chapter 6.

2.0 RELATED WORK

There is a large body of related work in this field, which will be divided into five general areas: multiple feeds sensor systems, multi-robot systems, spatiotemporal information fusion, asynchronous display, and analyzing multiple moving targets.

2.1 MULTIPLE FEEDS SENSOR SYSTEM

A sensor system is an umbrella term used to describe any combination of detectors configured as a single platform, such as unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), remote operated vehicles (ROVs) and other robots; or those that are configured as a network, such as wide-area video surveillance networks. Any detector that measures a property of the environment or sensor system itself and transforms that value into a digital signal meets the requirements of a sensor (Morison, 2010).

The sensor systems that have multiple feeds provide more capability and opportunity to discover needed information; however, there is a fundamental disconnect between the system's capabilities and the realization of that opportunity. Sensor systems do not presently allow an operator to act as if he or she were present in the environment, and they do not provide "persistent observation without significant coordination, attention, and cognitive costs, or guarantee that the right sensor data will be available at the correct place and time" (Woods &

Watts, 1997; Woods et. al, 2002). In addition, there is a misleading trend in sensor system design, in that the system is designed as intermediary that only captures and transmits data from the distant environment. As a result, the system will face the inherent question of its inability to integrate the diverse sensor feeds, which is called the multiple feeds problem. This may result in information overload and may also overload the human operator's perceptual abilities.

The multiple feeds problem was described by Robin Murphy in September 2009: "The developers acknowledge that sense making is a problem—cameras are sprouting like mushrooms from every conceivable mounting space in the hopes of giving the operator that magic missing viewpoint. However, it is all ad hoc; there is no understanding of what viewpoints are needed for which tasks or how to transition between them, so-called 'Perceptual interfaces' remain Neolithic."

Murphy's multiple feeds problem stems from the lack of coordination of multiple sensor perspectives. For human-sensor systems, "coordinating perspectives means coordinating distinct spatial points-of-view" (Morison, 2010). From another perspective, this multiple feeds problem pointed out the direction for our interface design; that is, our design must extend human perceptual ability by allowing operators to control their point of view not only spatially, but also temporally. In other words, information also exists in the fusion of the point-of-view, and our interface should help users process these kinds of information efficiently and quickly.

2.2 MULTI-ROBOT SYSTEM

2.2.1 Multiple Feeds Problem in Multi-robot Systems

The multiple feeds problem occurs at different scales in multi-robot system from multi-robot control (shown in the lower right corner in Figure 1) to multiple feedback surveillance (shown in the left part in Figure 1), in which the relationships across different perspectives are missing.

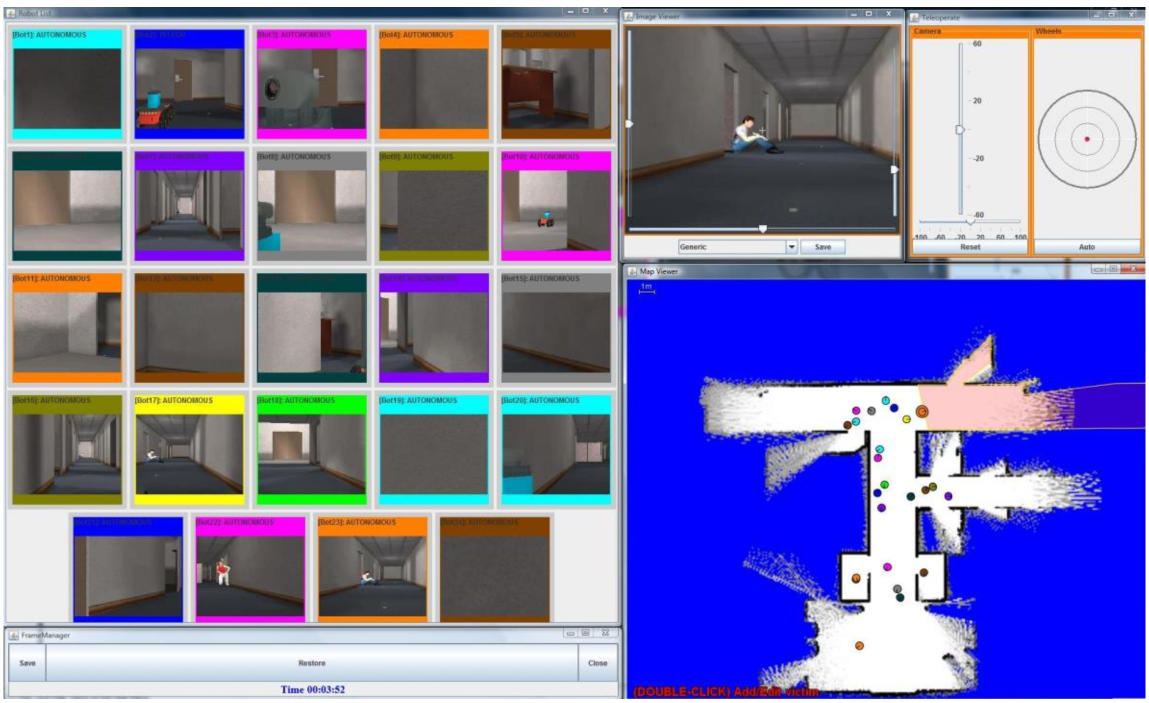


Figure 1. An example of the interface for multi-robot system

For example, in Figure 1, the 24 video feeds in the left part of the window shows a forward-looking view from the camera of the robot. This “wall of thumbnails” is a set of unrelated spatial views in a wide open environment. No spatial relationships are provided across sensors (video feeds).

In contrast, the lower right corner of the figure shows a top-down view map of laser sensor data that indicates the robot's nearby surroundings, while the colored circle indicates a different robot. These are uncoordinated spatial perspectives of the robot's surroundings, and as a result, the laser sensor data is only examined when the user encounters a problem with moving, based on the forward-looking video feeds.

To be more detailed, after a victim is displayed in the camera and once a victim is identified in a thumbnail, a complex and perhaps interrupted sequence of actions is initiated. The operator first needs to identify the robot and select it to view the camera in a larger window (shown in the upper right corner in Figure 1) and then must try to stop or teleoperate the robot. If this action is delayed, the victim may pass from the camera's view, and regaining the image of the victim may require further effort and teleoperation. After the user has successfully selected a robot, a series of actions must be performed to develop sufficient situational awareness to perform the victim-marking task. First, the robot must be located on the map by matching its window border color or numerical label. Next, the operator must determine the orientation of the robot and its camera by using cues, such as the prior direction of motion or matching landmarks between camera and map views. To gain this information, the operator may choose to teleoperate the selected robot in order to locate it on the map, to determine its orientation through observing the direction of movement, or to simply get a better viewing angle. The operator must then estimate the location on the map that corresponds to the victim in the camera view. If "another" victim is marked nearby, the operator must determine whether the victim he or she is preparing to mark has already been recorded on the map.

Without the means to combine the data gathered by all of the robots, the human operator is required to synchronously monitor their output. This requirement and load on the human operator may directly conflict 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 (Pepper et al., 2007; Lewis et al., 2005) confirmed that the search performance of these tasks is directly related to the frequency with which the operator shifts attention between robots.

This problem of high cognitive load required for the operator when trying to integrate information from multiple feed sensors is well documented, and several interfaces have proposed various solutions, such as occupancy maps (Yanco & Drury, 2007), snapshots for semantic maps (Nielsen et. al, 2007), 3D perspectives (Gómez et. al, 2010), and integrated/overlaid videos in a virtual scene (Nielsen & Goodrich, 2006; Nielsen et. al, 2007). However, too many non-integrated views can make the operator's task even more complex, while integrating all information will lead to some views getting ignored (Aubert et. al, 2010). Yanco & Drury (2007) also report that operators collide more often when relying only on the video, especially on the back of the robot. Thus, we will review the literature of visualization schema and interaction procedures for multi-robot systems, and attempt to propose a possible solution to this question of high cognitive loads.

2.2.2 Interaction with multi-robot systems

As the example has demonstrated, controlling multiple robots substantially increases the complexity of the operator's task, because operator attention must be shared among robots in order to maintain situation awareness (SA) and exert control (Drury et. al, 2003). The task of interacting with multi-robot systems (MrS), especially with large robot teams, presents unique challenges for human-robot interaction (HRI) researchers. Many different applications, such as interplanetary construction, search and rescue in dangerous environments, or cooperating unmanned aerial vehicles, have been proposed for multi-robot systems. HRI researchers have primarily focused on controlling these robot teams. These efforts have included the use of both the theoretical and applied development of the Neglect Tolerance (Crandall et al., 2006) and Fan-Out models (Olsen & Wood, 2006) to characterize the control of independently operating robots; predefined rules to coordinate cooperating robots, as in Playbook™ (Miller and Parasuraman, 2007) and Machinetta (Scerri et al., 2005); and techniques for influencing teams that obey biologically inspired control laws (Kira & Power, 2010; McLurkin et al., 2006; Ding et al., 2009). While our efforts to increase the span of control over unmanned vehicle (UV) teams appear to be making progress, the asymmetry is growing between what we can command and what we can comprehend.

Yanco & Drury (2007) performed a review of the human-robot interaction (HRI) interface design in the context of the AAAI Robot Rescue Competition in 2002, 2003, and 2004. The authors highlighted the importance of large video feeds and the integration of all necessary information and controls in a single window. The authors also related the percentage of the interface devoted to the robots' video feeds to the operators' performance, such as collisions and victim discoveries. These findings pioneered the field of rescue robotics and served as a basis for many recent interface designs.

Nielsen, Goodrich & Ricks (2007) presented an ecological interface paradigm that combines video, map, and robot-pose information into a three-dimensional (3D) mixed-reality display. The interface showed a 3D robot model, walls for the occupancy grid, and 3D located screenshots and videos. Their results showed that the interface increases the operator's ability to perceive, comprehend, and project the state of the robot, leading to a better SA.

Micire et al (2009) and Keyes et al (2010) explore the control of a single agent with a multitouch table that uses an adapted joystick-based interface. They performed a detailed analysis of users' interaction styles with two complex functions of the multitouch interface and isolated mismatches between user expectations and interaction functionality. The study leads to a better understanding of how to fuse information from multiple feeds sensors to lower the operator's cognitive load, as well as how to increase the relevance of the important elements.

More recently, Valero et al (2010) tried to design a HRI interface for multiple partially autonomous UGVs to enhance the operator's global (team) SA. The study uses different points of view (exocentric vs. egocentric) and coordinates reference frames (global map view, robot-centered, camera-centered) to enhance situation awareness.

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 to monitor and interpret the video that is being gathered by UVs. This can be a difficult task for even a single camera (Cook et al., 2006) and begins to exceed operator capability before reaching ten cameras (Humphrey et al., 2007; Lewis et al., 2010). With the increasing autonomy of robot teams, as well as plans for biologically-inspired robot "swarms" of much greater number, the problem of absorbing and benefiting from robots' output seems even more important than learning how to command them.

2.2.3 Information Fusion for Multi-robot Systems

One approach to such a problem is a picture-in-picture display (PiP), a specialized solution for integrating unmanned aerial vehicle (UAV) camera video (Draper et al., 2010; Hunn, 2005). In a PiP display, current video is scaled and transformed to reflect the position and orientation of the UAV and camera, and is projected onto a map of the area being surveyed. As the cameras move around the environment, the areas of the map falling under their view are refreshed. By presenting camera views in context, the PiP display eliminates many of the mental transformations and confusions that may occur when interpreting video from a remote camera (Gugerty & Brooks, 2001). Because the operator can attend to the information being gathered by a team of UAVs rather than to the video being received from any particular UAV, the PiP display provides a network-centric view (Alberts, Garstka & Stein, 1999) of the collected information and has been shown to lead to enhanced situation awareness (Draper et al., 2010). Another favorable property of the PiP display is that its ease of use should be independent of team size, because the size and resolution of the map remains constant. As UAVs are added to the system, regions of the map will be updated more frequently, which can reduce uncertainty without imposing additional cognitive load on the operator. In fact, as the update rate is increased, the operator will no longer have to predict or extrapolate and can simply observe events, which may make the task even easier.

Integrating unmanned ground vehicles (UGVs) into a network-centric view and reducing the difficulty of monitoring their video feeds is a more challenging problem, because there is no conveniently correlated aerially viewed map to provide a common representation. UGV camera views typically provide a projection of a 3D area onto a horizontal plane; namely, the camera plane of the UGV. The resulting image captures the space between a UGV and obstructions in the environment, such as walls or vegetation. Even though the view may be geo-referenced, its utility will depend on the degree to which it is relatively unobstructed. Differences in geometry also make UGV views much less informative. Because a UAV views the scene below it from a relatively long constant distance, objects and surroundings (obeying the inverse square law) have spatial relations that appear relatively constant. In contrast, for a UGV, objects will remain within camera view for long periods and will change drastically in size and appearance as the UGV approaches and moves away from them. In this case, views need to be chosen carefully for utility rather than simply transformed to fit a map.

While PiP views are inherently spatially organized and present collected data in their entirety, this setup may not necessarily work well for UGVs. Because UGV cameras collect many drastically different views of the same scene, there is no simple algorithmic way to simultaneously fuse all these views into a single, fully informative display. Simple projections on a map, as for PiP, are impossible, because such projections would involve a complete 3D reconstruction of the environment. Even if this were achieved through advances in 3D mapping and 3D reconstruction from images, viewing such a 3D reconstruction will not provide a simultaneous perspective of the entire environment. Once such simultaneity is lost, the user is no longer guaranteed to be able to see new events on screen as they occur. However, an asynchronous display of data may suffice for static environments, such as those that may occur in USAR situations that have immobilized victims, and may also prove to be valuable without a 3D reconstruction. This asynchronous display method will be discussed in Chapter 2.4.

2.3 SPATIOTEMPORAL INFORMATION FUSION

Many visualization techniques for analyzing complex event interactions only display information along a single dimension, which is typically one of time, geography, or network connectivity (Kapler & Wright, 2005). However, our problem lies in how to integrate and visualize spatiotemporal data from multiple feed sensors in an efficient way that is easy for a user to understand. Data may include video/images, time stamps, and correlated geographical information (laser scan/visual coverage of images), in addition to other types of information. As a result, we will describe this question in two levels:

How should we display and manipulate spatiotemporal data?

How should we organize multiple feed information using temporal and spatial cues?

2.3.1 Spatial and Temporal Data Visualization and Integration

One common approach to integrate space and time is to use a three-dimensional (3-D) display. For example, Andrienko and Andrienko (2003) developed a 3-D display interface called CommonGIS, in which data about events may be explored using the ‘‘space–time cube’’ representation. In the space-time cube, events are represented as circles placed vertically according to the time of their occurrence (Figure 3), with the earliest events shown at the bottom of the cube and the latest shown at the top. Variation of circle sizes or colors can additionally represent thematic characteristics of the events.

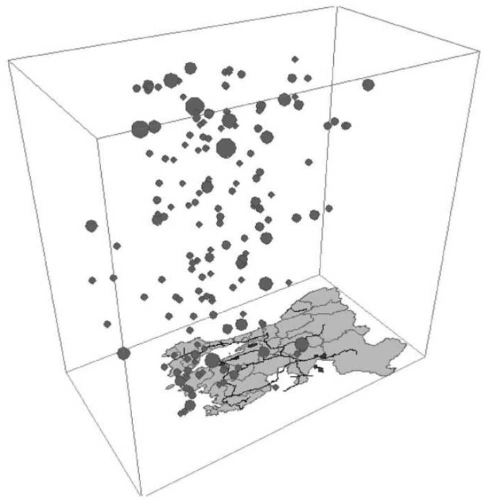


Figure 2. Space–time cube

(Andrienko and Andrienko, 2003)

Similarly, the GeoTime visualization (Kapler et al, 2004) represents events within an X,Y,Z-T coordinate space, in which the X,Y plane shows geographic space and the Z-T axis represents time in the future and past. Events are arrayed in time along time tracks, which are located wherever events occur within the spatial plane (Figure 3). For analysts, GeoTime's single view representation of a combined temporal-spatial three-dimensional space amplifies concurrent cognition of entity relationships and behaviors in both space and time. As a result, GeoTime improves perception of entity movements, events, relationships, and interactions over time within a geospatial context (Kapler et. al, 2007). However, these 3D approaches are not generally acceptable, due to the higher dimensionality of the spatial data. For example, the 3D space display could become cluttered with a large number of images that have a temporal order, and it may not be the easiest and efficient interface to use when browsing and searching for images.

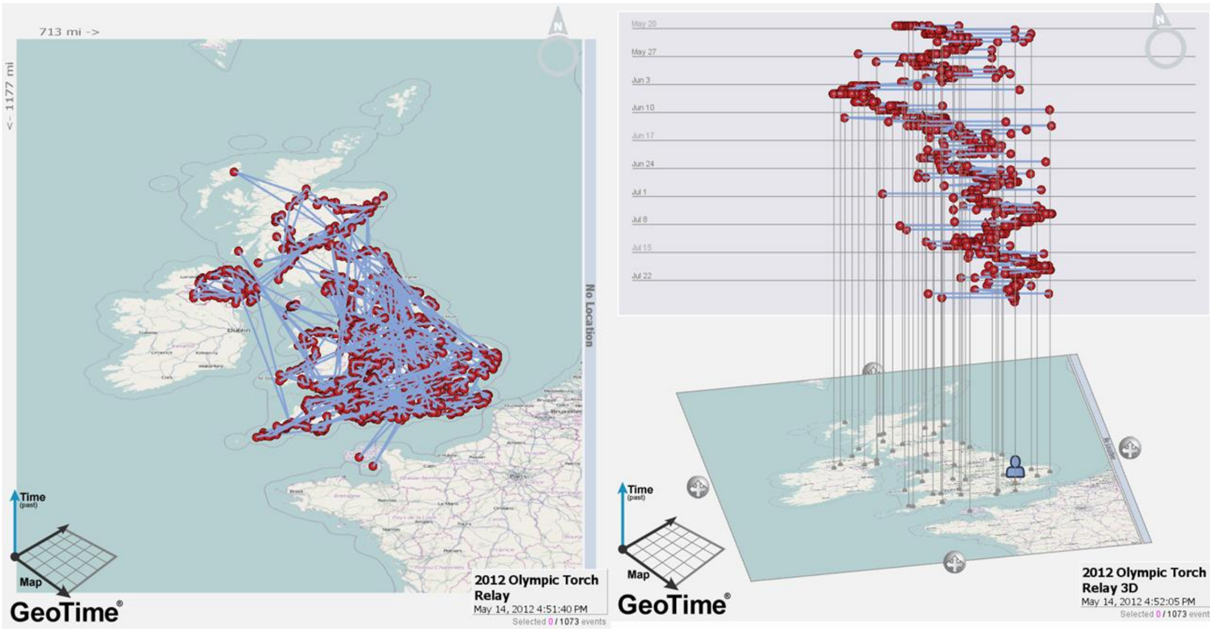


Figure 3. 2012 Olympic Torch Relay Route analyzed by GeoTime

(GeoTime, 2012)

Another attempt is in building a 3D virtual space from related images, such as Anabuki and Ishii (2006) and Tanaka et al. (2002). Spatial relations were found according to the relative orientation, translation between multiple images was based on the correlation of common visual features, and all images were connected to form a virtual space. The most well-known system of this type is Microsoft's Photosynth (Snavely et al., 2006), which is a purely software-based approach to the problem of position and orientation estimation. Photosynth takes unstructured collections of photographs, such as those from online image searches, and reconstructs 3D points and viewpoints to enable novel ways of browsing the photos (Figure 4). A user could carry out full 3D navigation and exploration of the set of images and world geometry, presented along with auxiliary information, such as overhead maps.

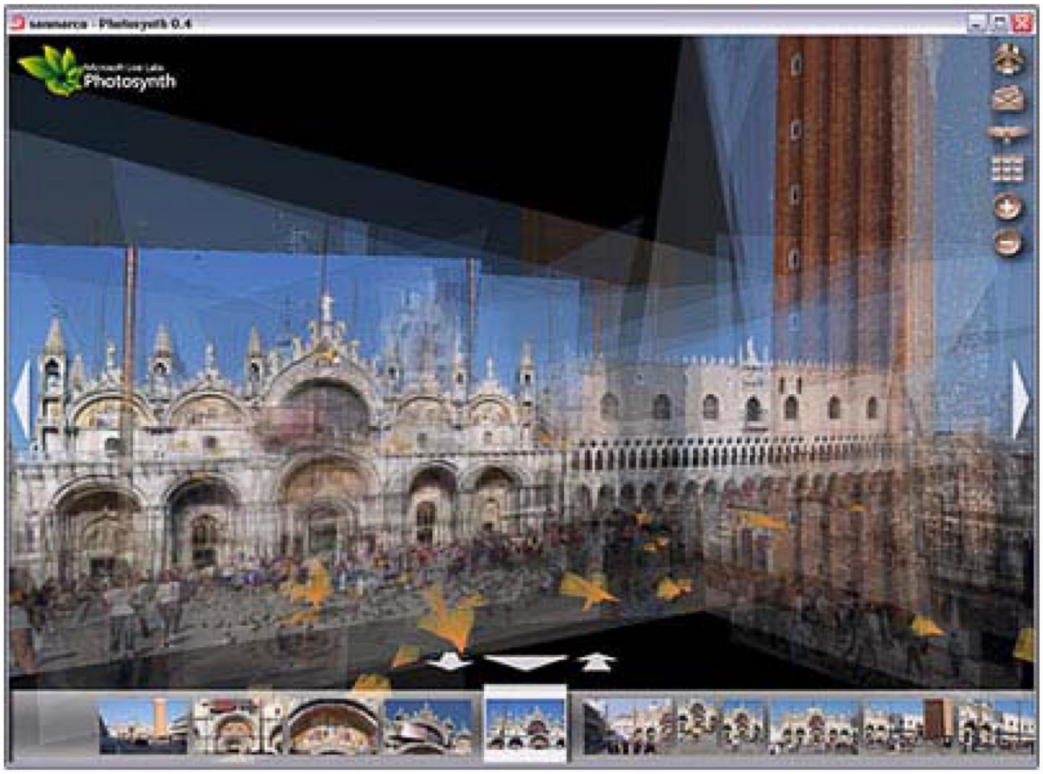


Figure 4. Demo of Microsoft's Photosynth Software

This photogrammetry technique is suitable for large amounts of high-quality images in related positions, such as well-photographed attractions (such as Notre Dame de Paris, shown above). Because the image-based rendering technique to smoothly transition between photographs depends on multiple correlated common visual features, there is unlikely to be enough information from the images shot by a regular USAR robot camera (60-degree field of view camera with a 320*240 resolution) to build a 3D virtual space. Less valuable information will be gathered when robots are traveling through a long hallway or turning quickly. Further, 3D spaces are difficult to navigate if they are based on a virtual world, as there are no such memorable landmarks (Lloyd et al., 2008). It would be easier for operators to map images to 2D floor plans than to a 3D space.

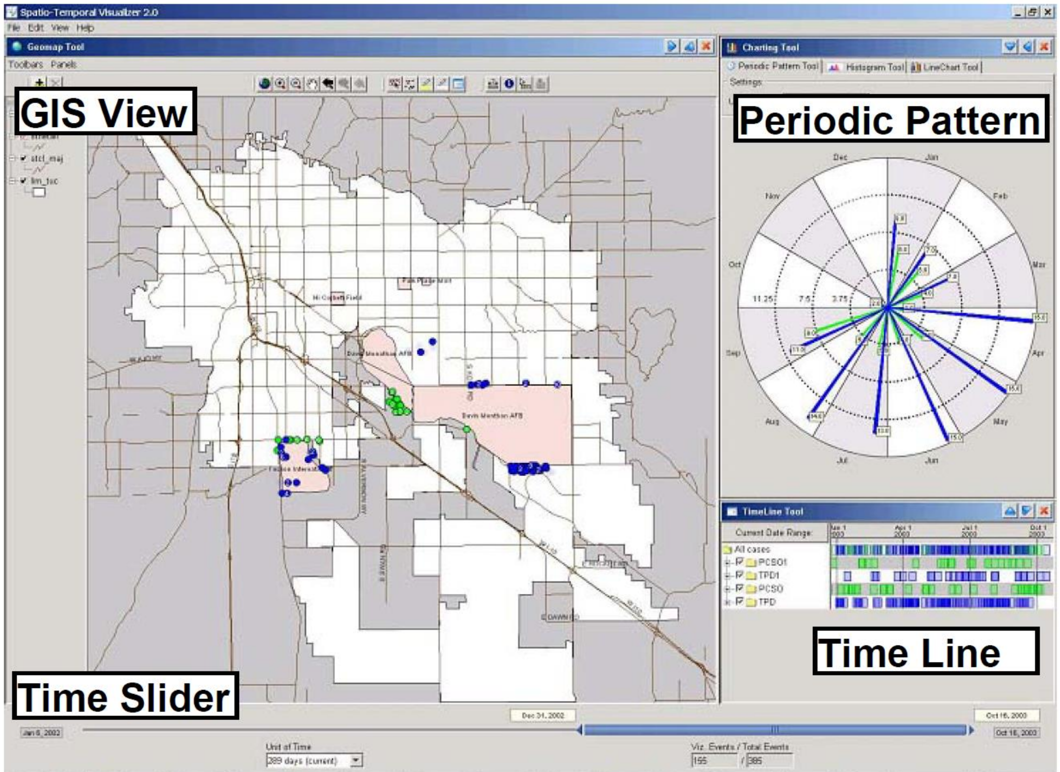


Figure 5. Major components of STV

(Chen et al., 2005)

Some other ideas involve trying to represent time and space in separate widgets that are linked together through brushing. Chen et al. (2005) have designed a system called spatial temporal visualization (STV) to integrate a synchronized view of three types of visualization techniques: a GIS view, a timeline view, and a periodic pattern view (Figure 5). STV was used to assist crime pattern recognition by including a GIS view to display incidents and a timeline to indicate crime density in the time dimension. They developed a 2D timeline that uses a block of color to represent the location of an incident in the time dimension. This crime density indication, along with a color hit of crime density, helps analysts identify crime clusters, which is an efficient way of presenting large-scale information.

Wu (2006) extended this work by showing spatial and temporal coverage of photos in a construction photo collection. Wu designed a visualization system called PhotoScope, which extends the standard photo browsing paradigm in two main ways: first, by visualizing spatial coverage of photos on floor plans; and second, by indexing photos using a combination of spatial coverage, time, and content specifications (Figure 6). PhotoScope generalized and reused the segmented timeline to represent the "density" of construction photos, and also used the idea of indicating "incidents" in the GIS view to show photo distribution on the floor plan. However, instead of using camera positions like STV, PhotoScope shows spatial coverage of photos on the floor plan, which provides more useful information to users.

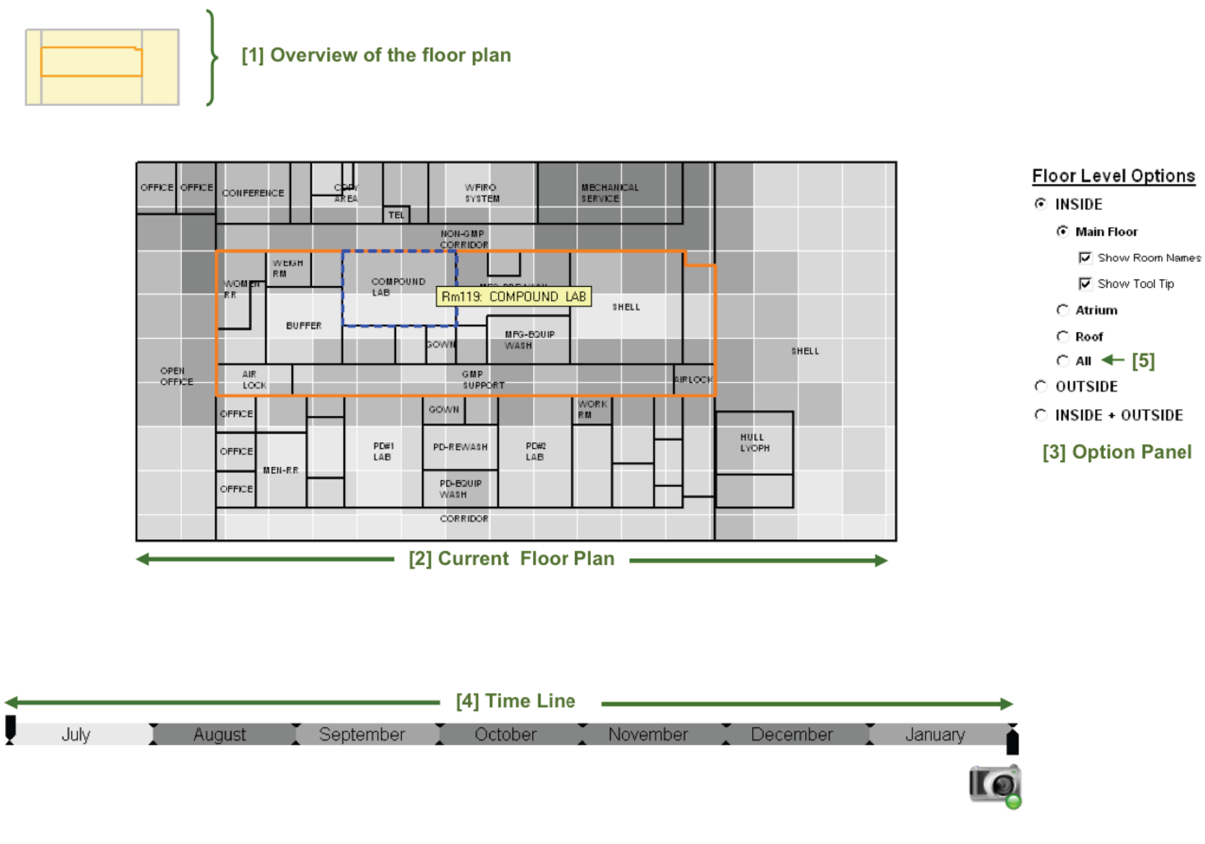


Figure 6. Overview of PhotoScope's main screen.

(Wu, 2006)

Yuri et al. (2007) tried to use a large number of simple motion sensors and a small set of video cameras to monitor a large office space. They developed a system to present this highly multidimensional data, which includes both spatial and temporal components. The interface visualized a sensor field overlaid on a floor plan and a timeline that showed the history of sensor activations (Figure 7). The system described in this paper is a manifestation of the idea that, in reality, most information that is of interest to users is contained in the context of multiple feed sensor networks, for which there is no need for all pieces information available.

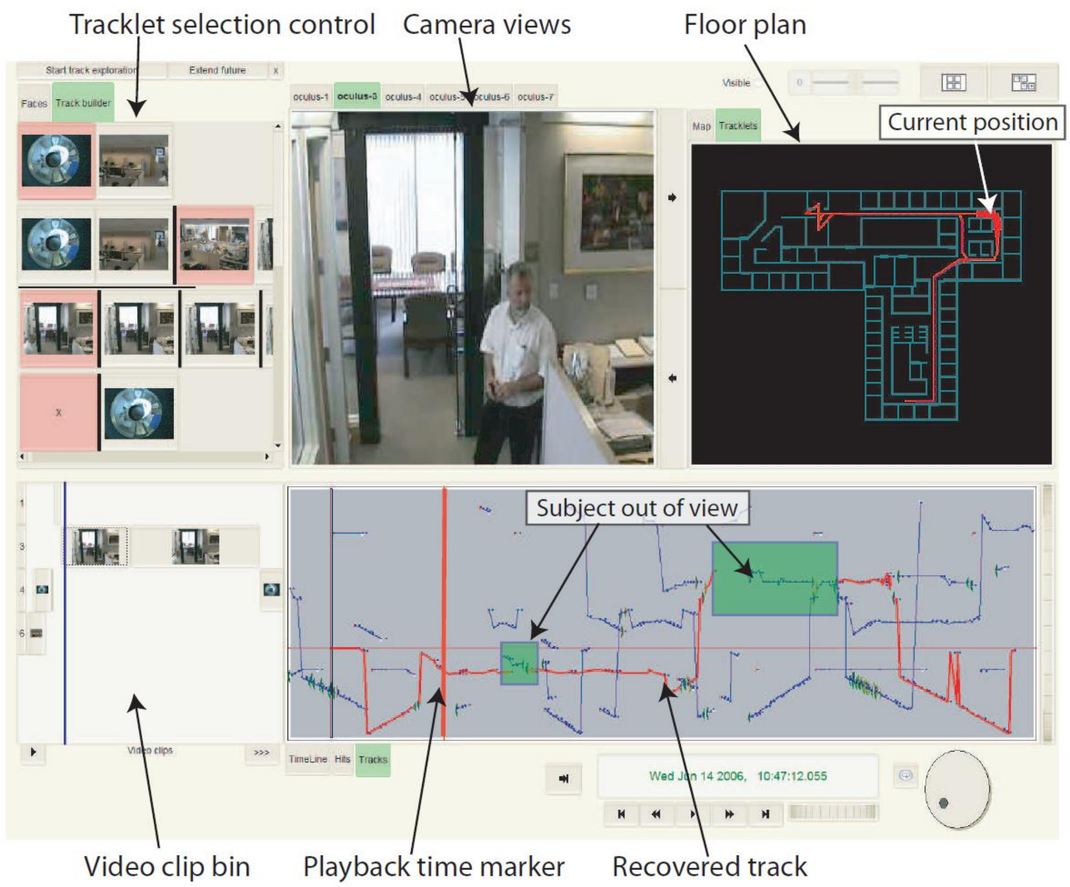


Figure 7. User interface of the MERL forensic surveillance system.

(Yuri et al., 2006)

Based on all the previous proposed solutions, and regarding the research questions we proposed, we need a visualization method that could do four major things: first, show the camera positions as well as spatial coverage of images on the 2-D maps; second, show spatial and temporal coverage of images; third, generalize and reuse the segmented timeline to represent the temporal distribution of images; and finally, realize the idea of overall spatiotemporal situation assessment.

2.3.2 Spatiotemporal Data Organization

A large quantity of image data collections or videos requires a correspondingly large organization effort. These data collections can be structured or viewed based on time and/or location. However, poor organization and layout may result in a sub-optimal use of screen space and may also require the user to engage in excessive panning. Researchers are struggling with a layout that conveys temporal and spatial order while making better use of space than simply a linear timeline.

Harada et al. (2004) creates a hierarchical structure of a user's photos by applying time-based clustering to identify subsets of photos that are likely to be related. They tried to associate a time line with album icons to indicate the time range that each album covered. The browser uses a novel interface that is based on a vertical, zoomable timeline (Figure 8).

Figure 8. Differently scaled time view of the Timeline browser

(Harada et al., 2004)

Time Quilt (Huynh et al., 2005) organizes photo albums into wrapped vertical columns in a temporal order. The system is designed to support visual searches and tries to maximize screen space usage, which can effectively convey temporal order (Figure 9). However, for a large image data collection and without zooming into the approximate date range of the temporally located image, it is difficult and inefficient to visually scan for a single image over the entire visualization. For this reason, researchers emphasize temporal ordering over space filling—but space filling is not guaranteed to minimize the amount of panning or facilitate faster visual searches.

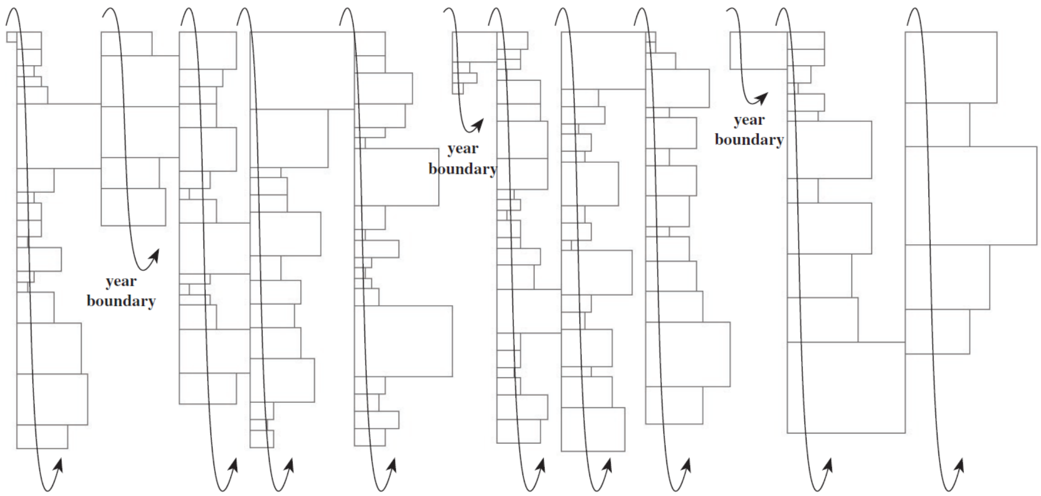


Figure 9. The Time Quilt layout "weaves" a linear timeline into columns of photo clusters (Huynh et al., 2005)

Other ways to organize images include PhotoSpread (Kandel et al., 2008) and Photo Tourism (Snavely et al., 2006). PhotoSpread is a system for organizing and analyzing biology field photo collections that are categorized by time, location, and other attributes. The changes of photos tags and location can be reorganized by drag-and-drop operations performed on an index spreadsheet. PhotoSpread was designed to meet the needs of field biologists who have large collections of annotated photos. This approach works well for categorical locations (e.g. field sites 1 and 2). However, visualizing the specific spatial region within a map is important to time-critical decision-making approaches.

Figure 10. The reconstruction of 3D points and viewpoints to browsing the photos

(Snavely et al., 2005)

Photo Tourism (Figure 10) is a 3D interface that can be used to interactively browse and explore large unstructured collections of photographs of a scene. The system consists of an image-based modeling front end that automatically computes the viewpoint of each photograph as well as a sparse 3D model of the scene and image to model correspondences. However, as we discussed in the previous chapter, the 3D interface may cause the cluster to fail when facing the multiple feed sensor problem. Photo Tourism does not address these issues.

Instead, Pongnumkul et al. (2008) have proposed a “map-based storyboard” (Figure 11) user interface for the user to reconstruct their tour, as documented by the video. In this interface, different pieces of video shot at different locations were directly correlated (tied) to their relevant locations on the digital map.

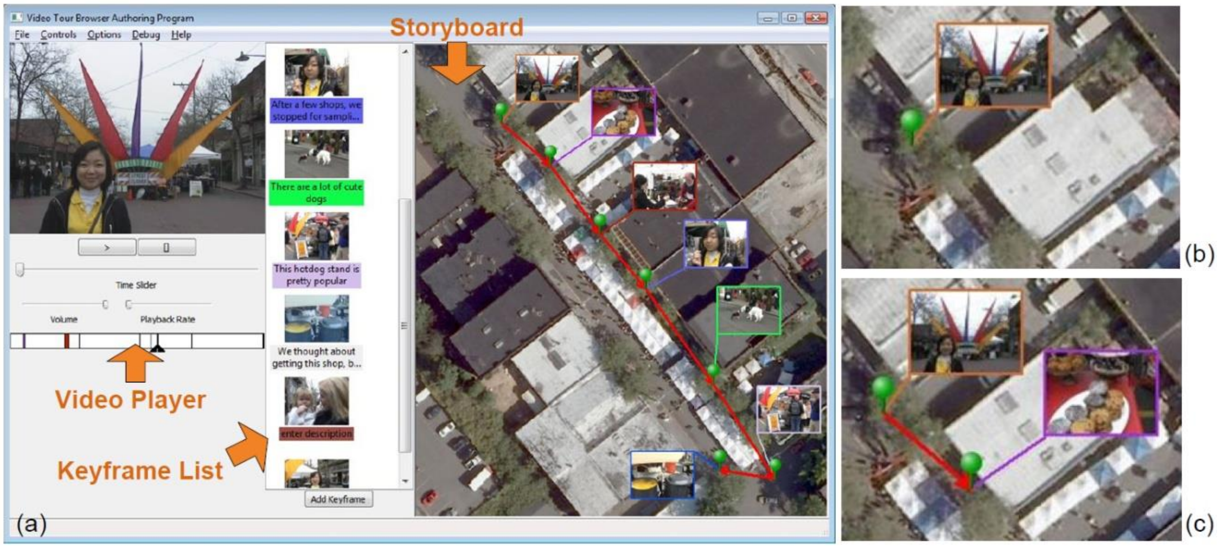


Figure 11. The UI layout of map-based storyboard with the location controller

(Pongnumkul et al., 2008)

In addition, this map-based storyboard approach provides a variety of interactions for the user to explore the tour, such as directly jumping to the video shot that correlates to a specific landmark, by only watching author-highlighted videos, or even by generating a unique virtual tour path that is different from the original one (Figure 12).

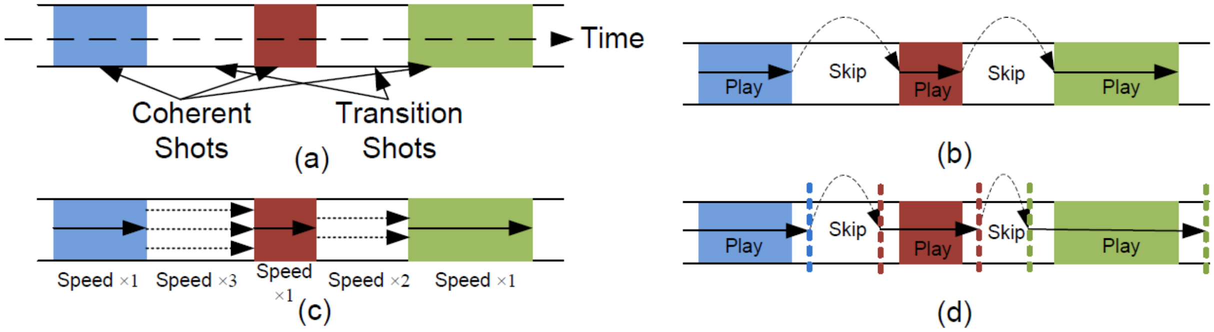


Figure 12. Advanced View Modes: pre-process the input video

by dividing it into highlighted shots and transition shots (a), which are viewing only high quality shots (b), Intelligent fast forward (c), and limited-time abstraction (d). (Pongnumkul et al., 2008)

However, storyboards do not facilitate indexing and searching for a specific shot, and geo-tags could be cluttered with multiple images collected from various moving sensors, since the same locations might be photographed repeatedly to document progress over time. A systematic method must be applied to reduce redundant spatial and temporal information.

The designed system offers two novel ideas: first, the concept of indexing images by space, time, utilities, and standardized content specifications; and second, the concept of showing spatial and temporal coverage of images in statistic charts with the timeline segmentation, which emphasizes the content of images rather than the mechanics of how the image was taken. We want our system to make navigating and exploring spatiotemporal data easier, enable users to extract more useful and reliable spatiotemporal information, and provide an overview of the progress of spatiotemporal objects over time.

2.4 ASYNCHRONOUS DISPLAY

The examples of spatiotemporal data visualization in listed Chapter 2.3 follow a similar scenario, which is an extra dimension added to a two-dimensional map display that shows the temporal order of events. Even if this were achieved, viewing such a 3D reconstruction does not also provide a simultaneous perspective on the entire environment. In other words, this method of visualization cannot be directly used in dynamic environments that have multiple sensors, as no retrieval interaction was offered in spatiotemporal components and the number of simultaneous events causes visual clutter. Once simultaneity is lost, the user is no longer guaranteed to be able to see new events on screen as they occur.

An asynchronous display method can alleviate the concurrent load put on the human operator and can disentangle the dependency of tasks that require direct attention to multiple video feeds. In other words, the asynchronous display method will turn force-paced tasks into self-paced tasks, and will also lower the user's mental workload. Furthermore, it is possible to avoid attentive sampling among cameras by integrating multiple data streams into a comprehensive display. This asynchronous display method allows the addition of new data streams without increasing the complexity of the display itself.

2.4.1 Panorama Pictures

An earlier approach to asynchronous display for USAR, called Panorama Picture Project, has been explored in Ulab at the University of Pittsburgh since 2007 (Velagapudi et al. 2008). This method, which is motivated by asynchronous control techniques previously used in extraterrestrial NASA applications, relies on substituting a series of static panoramas taken at designated locations for continuous video (Figure 13). The operator then searches through the panoramic pictures to determine the location of targets that are viewable from each of the selected locations. In a four-robot experiment that focused on comparing panoramas with streaming video, there was no difference in the number of victims found or area explored.

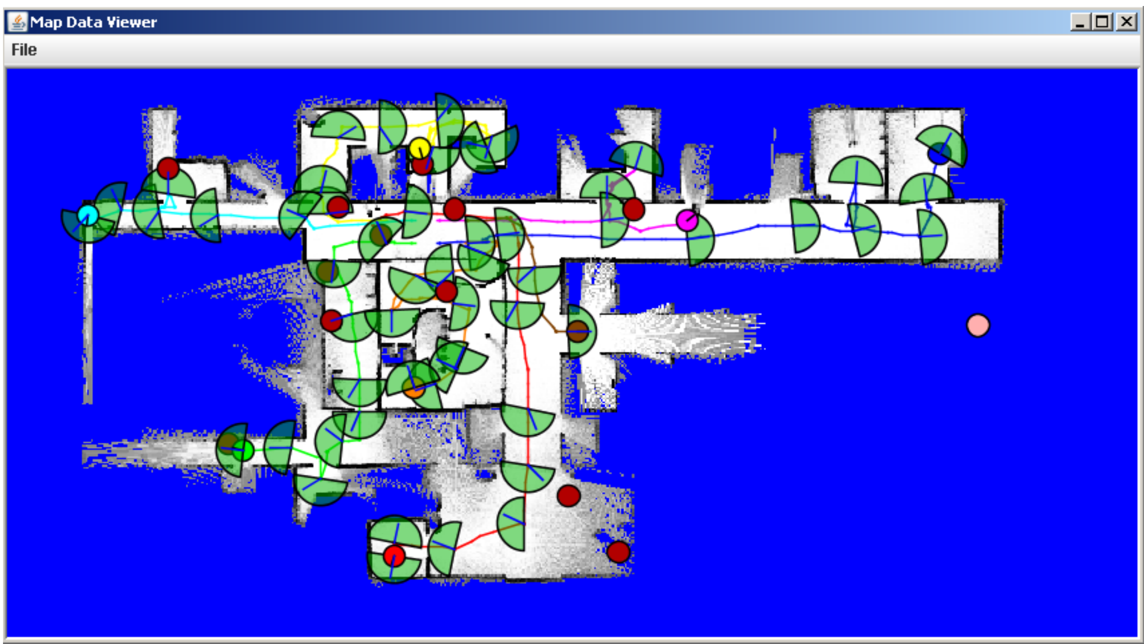


Figure 13. An example of the explored map for panorama picture experiment

A further experiment (Velagapudi et al., 2009) scaled the team size to eight and twelve robots, based on the premise that advantages for self-paced imagery search might emerge with increasing numbers of video feeds to monitor in the synchronous control condition, but again, no differences were found in these conditions. However, this approach did not utilize all the available data from the video feeds that robots gather, so a huge amount of potentially useful information from the panorama condition was discarded. Furthermore, the operator must give the robots additional instructions on where to sample future panoramas.

2.4.2 Image Queue

In contrast to previous work, the Image queue approach allows the use of autonomous exploration. Wang et al. (2011a) present an asynchronous display that mines all of the robot video feeds for relevant imagery, which is then given to the operator for analysis. This type of asynchronous display is called "image queue" as compared to the traditional synchronous method of streaming live video from each robot, which refers to "streaming video." The image queue interface (Figure 14) focuses on two tasks: (1) viewing imagery and (2) localizing victims. It consists of a filmstrip viewer that is 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.

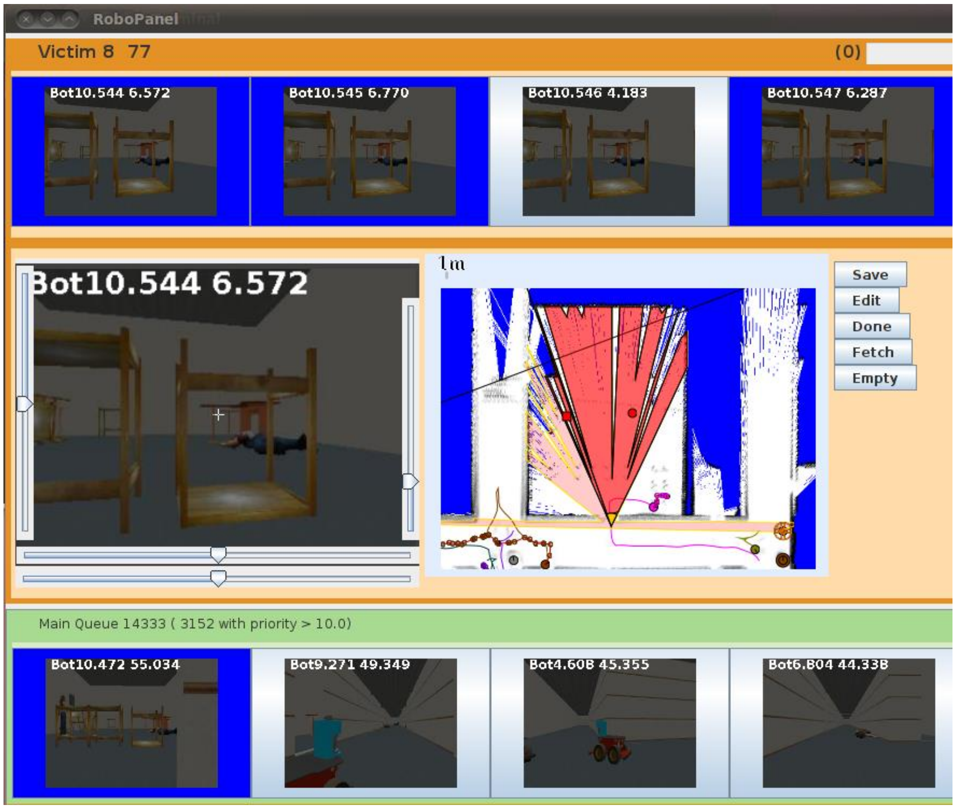


Figure 14. GUI for Image Queue interface

Experiment results show that the frequency of shifting focus between robots was correlated with performance for streaming video, but not for asynchronous images. As expected, the asynchronous display of information alleviates the need for excessive switching, which means that image queue participants have no need to teleoperate a robot, in contrast to streaming video participants, which may need to teleoperate a robot when they encounter a victim in the video feed. Most importantly, they do not need to stop the robot in order to precisely locate the victim. In essence, image queue interaction has decoupled the navigation and error-recovery tasks from the victim-detection tasks, allowing the latter tasks to be completed entirely asynchronously without any penalties for performance in terms of the number of victims. It was also found that the image queue reduces errors and operator's workload, as compared with the traditional synchronous display.

A further development for the Image Queue method involves automatic target recognition (ATR) (Wang et al., 2011b), which was used to augment utilities based on visual coverage in selecting imagery for presentation to the operator and that helps the user to identify targets of interest, such as injured victims (Figure 15). Originally, it was expected that the ATR display might reduce user workload and improve overall performance. However, the analysis of victim marking errors shows that the ATR group marked 52.9% more victims at the wrong location, but did not miss more victims. A similar disadvantage in reported workload suggests that substantial cognitive resources were required for the ATR group to separate false alarms from accurately placed boxes.

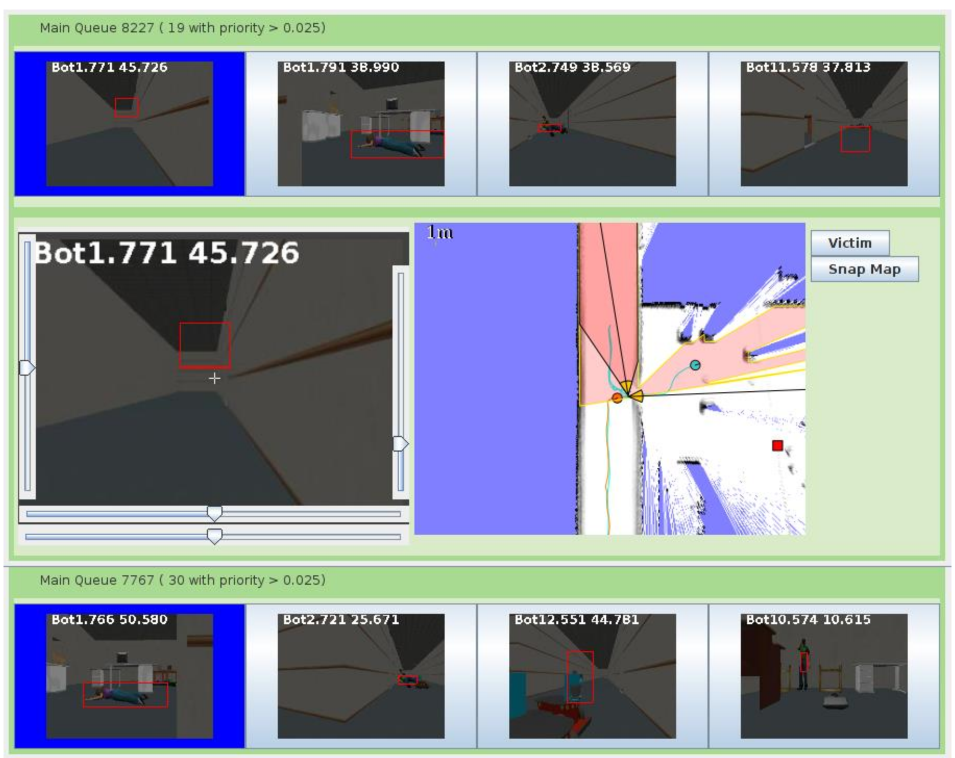


Figure 15. Image Queue GUI with automatic target recognition

An example of a final map for the ATR group illustrates this problem (Figure 16). When the operator viewed all images from a newly covered area or with newly detected victims, the system may continue to pull images from this general area, because priority is determined by victim probability as well as coverage. As a consequence, new images containing already marked victims may enter the queue, even though they represent only minor increases in coverage. Under these conditions, the ATR display frequently confused operators, which led them to mark the same victim twice or even three times at the same location. This problem could be alleviated by augmenting the priority computation by considering whether a target may have been already marked by the operator. This result indicates two important lessons: first, we may shift more cognitive tasks to users (such as short-term memory tasks or visual search) when we want to develop some automation assistance; and second, the significance of information may shift over time, especially when facing a dynamic environment.

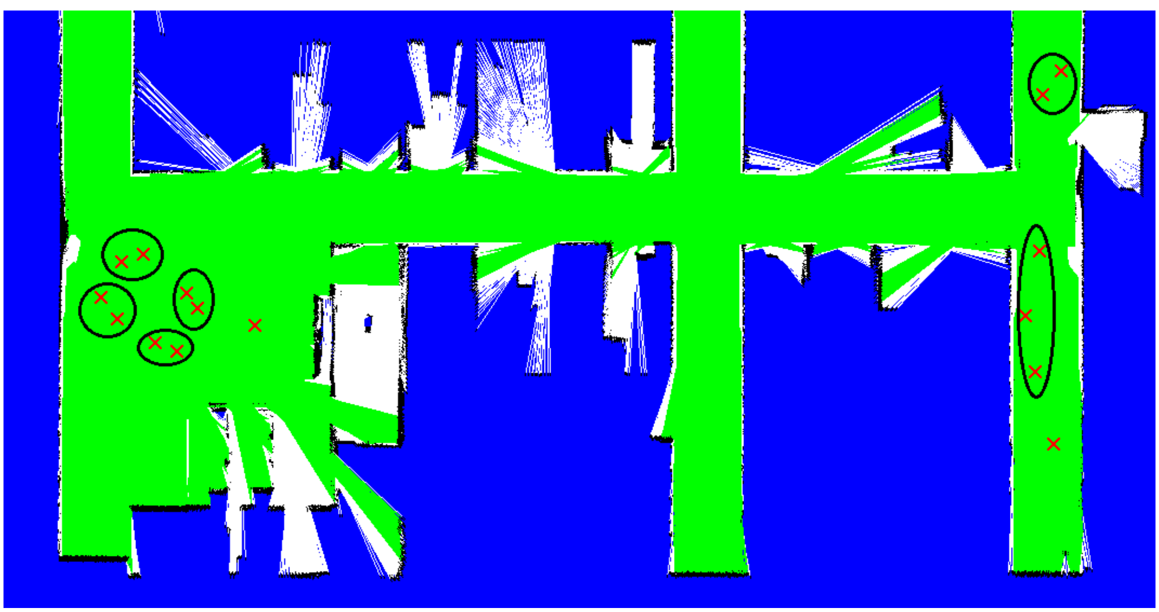


Figure 16. Example for ATR group map

Image queue display methods are great attempts at using synchronous interaction. However, errors that originate from a complex interaction between tasks will be difficult to detect by human operators in these synchronous interaction methods. Effectively, operators rely more on the autonomy and system design as they sacrifice the ability to have a complete overview of the system and its functions. Such a sacrifice is often made when scaling large and complex systems, but its impact on foraging tasks is not yet well understood. In other words, this display method may turn the operator into an image processor who has totally lost the global situation awareness and part of local situation awareness required for effective navigation. As a result, the user loses their control of the high-level searching or navigating strategies and totally relies on automation.

2.5 ANALYZING MULTIPLE MOVING TARGETS

Movement data link together space, time, and objects positioned in space and time. They hold valuable and multifaceted information about moving objects and properties of space and time, as well as events and processes occurring in space and time (Andrienko et. al, 2011). These multidimensional data raise significant challenges for analysis; however, it is also a great opportunity for movement data, such as that required to understand the temporal dynamics of events and processes, or to investigate behaviors of moving objects.

Analysis of movement is a popular visual analytics research topic that has been used in recent years. A wide variety of methods and tools for analysis of movement data have been developed, which allow analysts to look at the data from different perspectives and fulfill diverse analytical tasks (Andrienko & Andrienko, 2012). At the same time, visualization and interaction techniques are often combined with computational processing, which, in particular, enables analysis of larger amounts of data than would be possible with purely visual methods. The following fields are related to our research problem: visualization, visual analytics, geographic information science, database, date mining, and computational geometry. Thus, we need to find a suitable method to analyze our data and tasks from a complex combination of fields of study.

2.5.1 Spatiotemporal Objects

Peuquet (1994, 2002) distinguishes three components of spatiotemporal data: space (where), time (when), and objects (what). Accordingly, Peuquet defines three basic kinds of questions:

- when + where → what: Describe the objects or set of objects that are present at a given location or set of locations at a given time or set of times.
- when + what → where: Describe the location or set of locations occupied by a given object or set of objects at a given time or set of times.
- where + what → when: Describe the times or set of times that a given object or set of objects occupied a given location or set of locations.

Peuquet (1994, 2002) also defined three fundamental sets pertinent to movement: space, S (set of locations), time, T (set of instants or intervals, jointly called ‘time units’ or simply ‘times’), and objects, O. Based on these concepts, Andrienko et al. (2011) classify the following definitions according to the objects:

- Spatial object: an object that has a particular position in space at any time moment of its existence.

- Temporal object: an object with limited time of existence with respect to the time period under observation; or, in other words, an object that has a particular position in time (also called events).
- Moving object: a kind of spatial object that can change its spatial position over time (also called a mover).

As a result, we can jointly call these objects spatiotemporal objects. Based on the triad model of “what, when, where” given by D. Peuquet (1994), Andrienko et al. (2011) summarize the characteristics in a graphical form (Figure 17). From the figure, we can see how the spatial and temporal positions of objects are related to locations and times. Furthermore, trajectories of movers are linked to locations by spatial relations and to times by temporal relations.

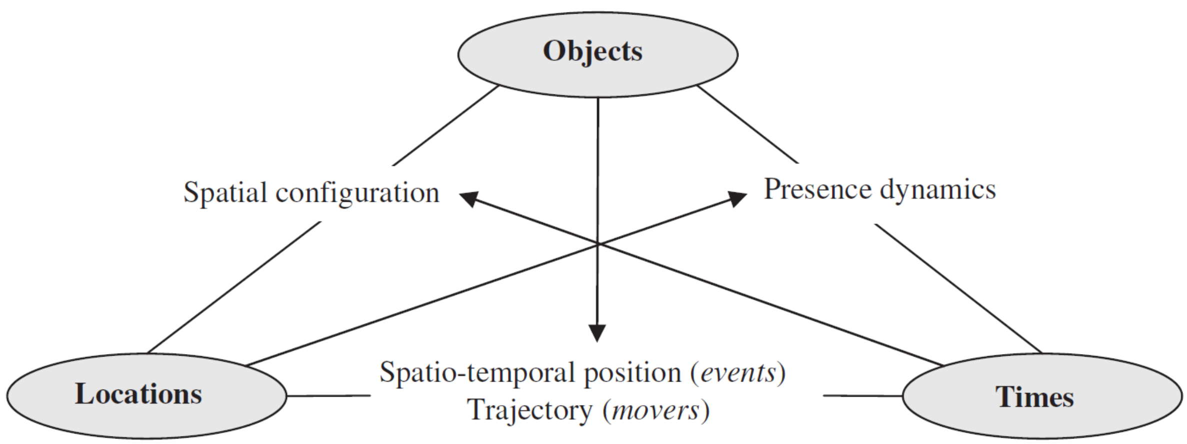


Figure 17. Characteristics of objects, locations, and times

(Andrienko et. al, 2011)

2.5.2 Movement Analysis Task

Blok (2000) and Andrienko et al. (2003) define analysis tasks for spatiotemporal data based on the types of changes that occur over time. These changes can include:

- Existential changes: appearance and disappearance of objects.
- Changes of spatial properties: location, shape or/and size, orientation, altitude, height, gradient, and volume.
- Changes of thematic properties expressed through values of attributes: qualitative changes and changes of ordinal or numeric characteristics (increase and decrease).

According to the three fundamental constituents of movement, analyzing movement may have three different foci: objects, space, and time. However, to answer our research question, we consider that the following analysis needs to be done for the moving targets:

- Trajectory extraction: This includes trajectory interpolation, summarization, and re-sampling (obtaining position records for regularly spaced time moments with a desired constant temporal distance between them (Yuan et al., 2010)). It is the top task for moving targets, as the shape of a trail allows for the rapid visual identification of patterns or anomalies in behavior (Kapler et. al, 2007).
- Events extraction: With the combination of space, time, objects, and contextual data, many kinds of relevant events can be extracted from movement data, such as attaining speed or visiting a place or target meeting.
- Significant place extraction: It is unreasonable and unfeasible to investigate the characteristics of all points in space or relations among the points. Therefore, it may be necessary to find significant locations that are relevant to the goals of the analysis (Andrienko et al., 2003). Possible criteria include the frequency that certain movement-

related events occur in certain locations or the amount of time that visitors spend in a certain location.

2.5.3 Visual Analytics of Multiple Moving Targets

Visualization is one of many methodologies that can be used to analyze moving objects. Although we spent a lot of effort on discussing spatiotemporal data visualization and organization in the last chapter, we will discuss a selection of visualizations particular related to moving objects in this section.

Visualization technology is introduced to support the analysis tasks that target the spatial and temporal characteristics of the objects (movers and events), as well as demonstrating their relations to space and to time. An animated map and a space-time cube (Andrienko et al. 2003, Kraak 2003, Kapler and Wright 2005) are used for movement data. Multiple views are used to take account of additional attributes in the analysis of moving objects, performed by Dykes et al. (2003). Riveiro et al. (2008) also present a novel system that supports the insertion of the user's knowledge and experience in the creation, validation, and continuous update of the environment. As demonstrated in Figure 18, users can drag and modify the probability values if they consider them unrealistic. If one probability value is modified, the others are recalculated.

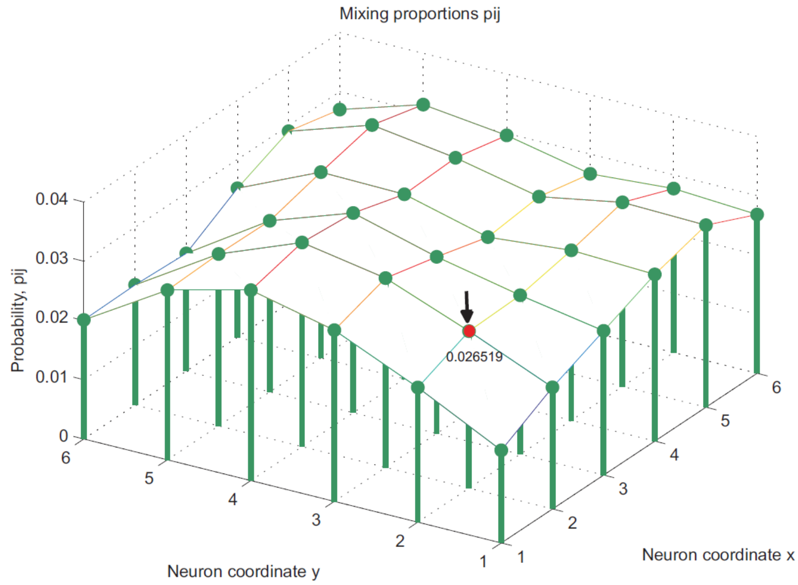


Figure 18. Visual analytics for the detection of anomalous maritime behavior

(Riveiro et al., 2008)

Another challenge lies in how to show thematic characteristics of the moving targets. One way is to show the thematic characteristics in same display by visual properties (color, size, shape, etc.) of the graphical elements that represent the objects (Bak et al., 2009). Alternately, additional visualizations, such as scatterplot, parallel coordinates, and time graphs may be used to indicate the thematic characteristics (Andrienko et al. 2011). Temporal bar charts, also known as Gantt charts, can represent the temporal positions of objects. Typically, the horizontal dimension of this display represents time, and bars positioned according to the lifetimes of the objects represent thematic characteristics (Figure 19). Time-variant thematic characteristics can be represented by colors or shades of bar segments (Kincaid and Lam 2006). In our visualization design, we will choose the temporal bar chart to assist a user's temporal navigation.

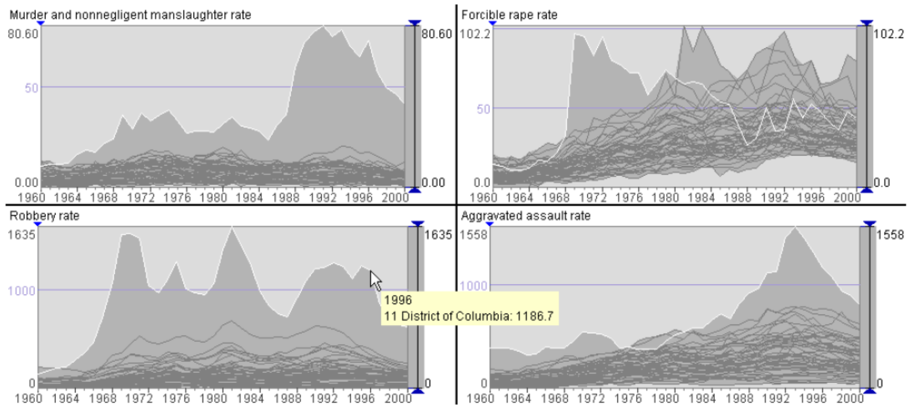


Figure 19. Time graph display of the temporal variations of the crime rates

(Andrienko et al., 2008)

Hurter et al. (2008) designed FromDaDy, a trajectory visualization tool that tackles the difficulties of exploring the visualization of multiple trails. This multidimensional data exploration is based on scatterplots, brushing, pick and drop, juxtaposed views, and rapid visual design. Their innovative use of brushing trajectories information across views gave us great inspiration.

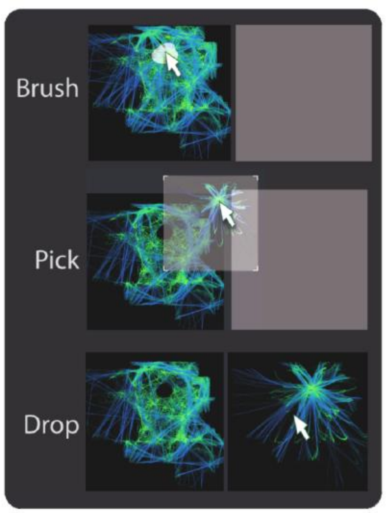


Figure 20. Pick and Drop interaction in FromDaDy (Hurter et al., 2008)

3.0 DESIGN OF THE SYSTEM

3.1 SCENARIOS AND MOTIVATION

When supervising a team of semi-autonomous robots, the operator's task is not only to manipulate each individual robot, but also to achieve the top goal assigned to the entire team of humans and robots (Furukawa, 2010). It is important to support the operator's skill-based behaviors, but it is more significant to support human operators in their understanding of the overall state of a study-in-progress and the situation around it by using a system-centered view. Although the cognitive resources of humans are limited, operators are required to understand highly complex states and make appropriate decisions in dynamic environments (Calhoun et al., 2005; Goodrich et al., 2005). Additionally, based on the information collected by the sensors, the system should be able to help user trackback the existing data.

Fong and Thorpe (2001) claim that human-robot interfaces "must provide a mechanism for the operator and the robot to exchange information at different levels of detail or abstraction" because the requirements of a given operator's task are different at different levels. The information that an operator needs for manipulating an individual robot is clearly different from the information needed to analyze the overall situation and make decisions to achieve overall goals.

As a result, the operator must use a multi-scale visualization, which is a consequence of modeling global-scale events with small-scale details. Multi-scale visualization in this context means that information exists at multiple levels of detail and that these levels of detail are not presented all at once (Jul & Furnas, 1998). The multi-scale feature introduces two related concepts: zoom and information scale (Humphrey, 2009). Zoom means that the "level of visualization detail can change in a navigable manner." This type of multi-scale system is also called a zoomable user interface, to highlight the zooming capability over the information scale (Pook, Lecolinet, Vaysseix, & Barillot, 2000). Information scale is the concept that "a particular piece of information is not necessarily present at all levels of detail, because the information may be too small, too large, or too dense to be presented at a particular level of detail" (Humphrey, 2009).

These questions lead to a human-aided multi-robot (sensor) information fusion problem. First, the relevant information cannot be completely represented by computer models (interfaces). As a result, it is necessary to implement mechanisms that permit human operators to apply the full range of knowledge that only they can master, and it is important to identify the fundamental requirements for such a system. Second, while the necessary information must be present, it must also be presented in such a way as to maximize its utility. Information can be present in separated areas of the interface, which would require users to manipulate windows to gain an overall picture of the system's state. Such manipulation takes time and can result in an event going unnoticed for some time. Information fusion is an underestimation of presentation in HRI, as time delays and errors occur when users need to fuse different pieces of information both spatially and temporally.

When many images are collected over a region of interest during certain period of time, how are these images presented according to users' spatial and temporal concerns? This overarching system goal could be considered in the following levels of requirements.

System Requirement 1

The system should allow the assessment of the situation in the entire map at any time period that has already happened, not only from time periods in front of a single sensor. In other words, the system must enable users to manipulate the time coverage and retrieve corresponding images collected within the time range from all sensors.

System Requirement 2

The system should allow the assessment of the situation at any time period or time point within any concerned region. The system must have a retrieval function in both spatial and temporal scale.

System Requirement 3

The images that users select should be presented in an appropriate form (interface), so that the user can correlate the image with spatial and temporal parameters. Specifically, the system needs to show spatial coverage of images with a time stamp, as well as the shooting location and direction, to provide an overview of which regions have been documented by images.

System Requirement 4

Redundant data or information should be contained out of the way of the main interaction. We want to make the information spatially and temporally unique, while ensuring that there is no distortion or loss of data fidelity. We attempted to achieve this by adding filter and statistic function in both spatial and temporal scale, and added assistance to the user with a clear and apparent display.

System Requirement 5

All data processing required for indexing and storage of the data should be fast and efficient. If the searching and sorting process takes longer than the real-time synchronous procedure, then the system will not scale well.

The main goal of the proposed system, called Event-Lens, is to generate asynchronous prioritized images that will provide the operator with a manageable but comprehensive view of the information collected by the robot team (from multiple sensors). An asynchronous, scalable, and comprehensive human-robot interaction method will be proposed, which allows operators to manipulate images of spatial and temporal resolution, as well as to maintain necessary global and local situation awareness.

3.2 SYSTEM ARCHITECTURE AND OVERVIEW

The goal of the Event-Lens interface is to best utilize the advantages of an asynchronous display and to maximize the amount of visualization of information in spatial and temporal dimensions that are available for human operators. As the number of robots in a system increases with improved autonomy, the demands on operators for these tasks increase as well. As a result, another requirement for the Event-Lens interface is to provide the potential for scaling to larger numbers of robots and operators. The proposed Event-Lens interface implements the idea of asynchronous display of a collection of images that have top utility values (spatiotemporally covered areas) that allows operators to navigate through the data. At the same time, images will be displayed with spatial information, which is intuitive for users.

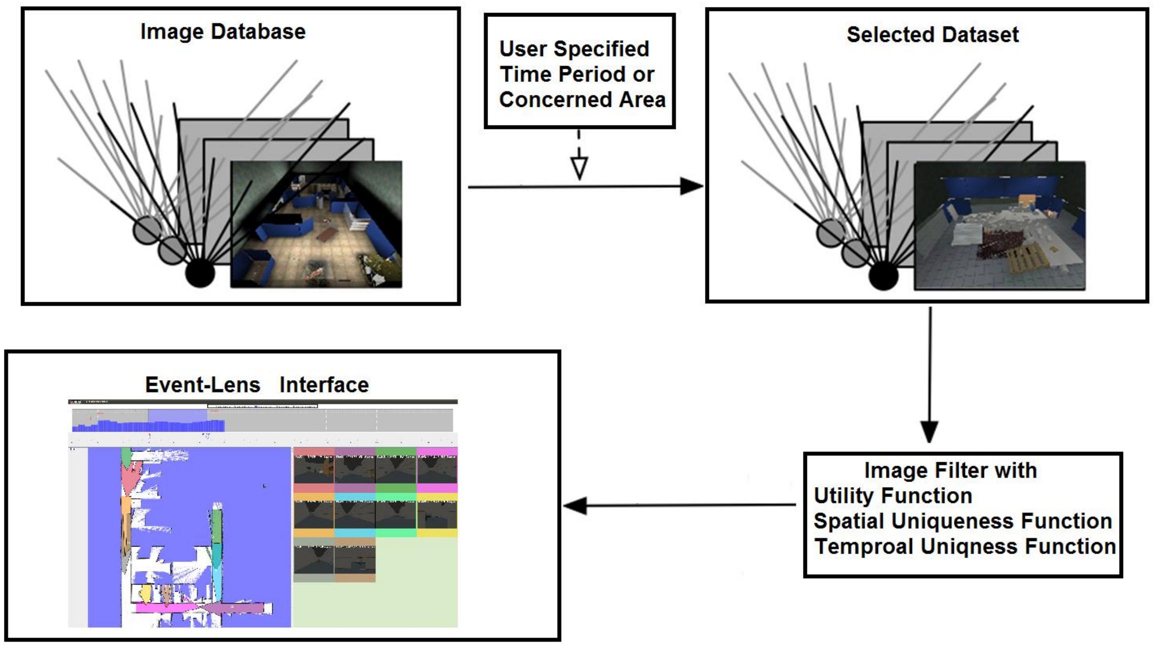


Figure 21. An illustration of the system architecture for Event-lens

Figure 21 presents an overview of the steps involved in the Event-Lens system. A database stores observations that are made by multiple robots, such as laser scans, video frames, and time stamps. From these, an image filter extracts high-utility images when a user specifies the period of time or the area to which images belong, which are then filtered and presented for analysis in a parallel image-in-map display.

The main screen of the Event-Lens system (Figure 22) consists of a number of components. An overview map of the floor based on the laser scan sensor data lies at the lower left corner, and the selected image in the priority queue with current visual coverage is shown in the lower right. The visual areas could generally be subdivided into the top selected images, with sequenced smaller regions, according to the utility scores of images.

Once the user selects an image of the subarea, a zoomed-in picture will display automatically and a tool tip appears that displays detailed information, such as the covered area's numerical value, the time stamp of the image, and the robot's number. The image can be turned off when a user clicks on another area of the map.

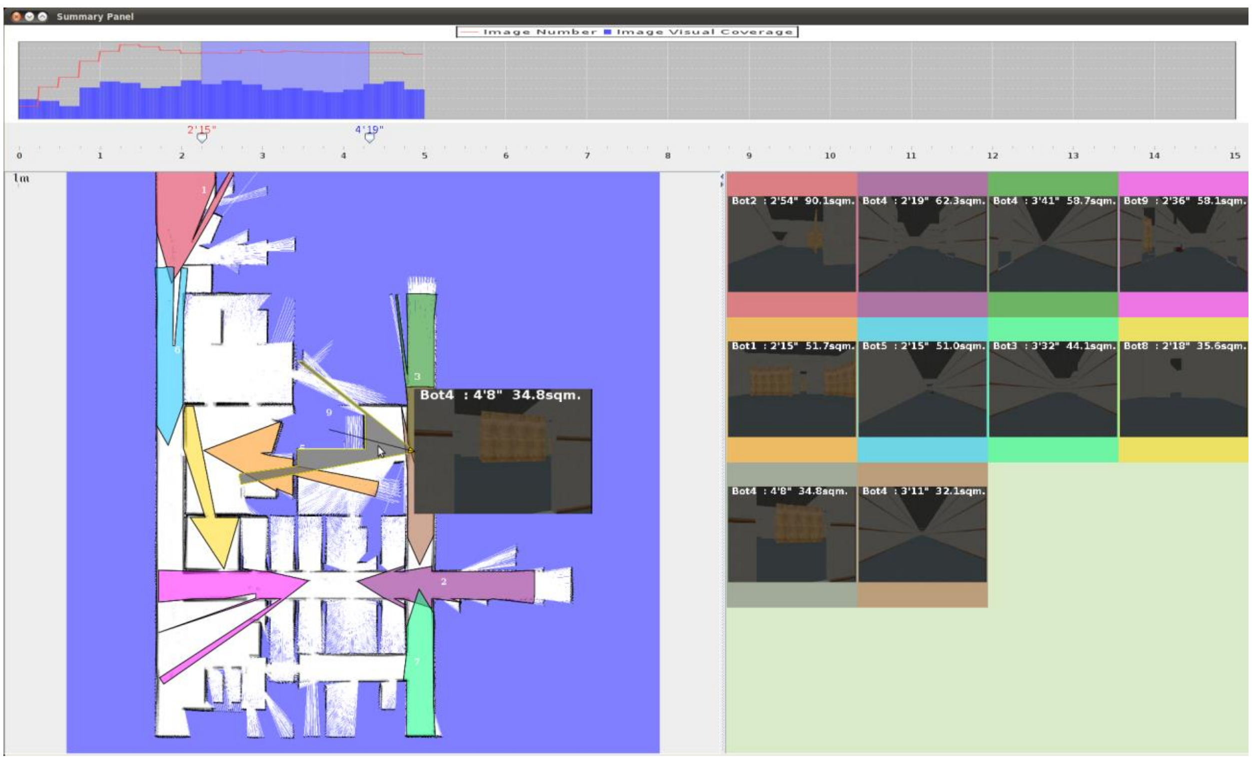


Figure 22. Prototype of GUI for the Event-Lens interface

The Event-Lens interface (Figure 22) focuses on two tasks: (1) viewing imagery while considering the spatial and temporal coverage of certain images, and (2) helping users to specify the period of time or concerned area where images were selected and calculating their utilities. It consists of a parallel image-in-map viewer that is designed to present the operator with a global view of where the image was taken, with view coverage showed alongside this location data. 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 within a short period of time. The image filter attempts to reduce this 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 goal, 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 individual image and compute its utility. The computation of utility can be adapted to a particular application. For our experiments, we computed utility through the area covered, as seen in a particular image.

3.3 DATABASE DESIGN

The first element of the Event-Lens system is the information database, which receives data streams from the robot team and stores them as information frames. The data collected by each robot consists of time-stamped laser scan geometry, streaming camera images, and inertial navigation system (INS) pose estimates. For our experiments and for image queue implementation, ground truth sensor information was provided.

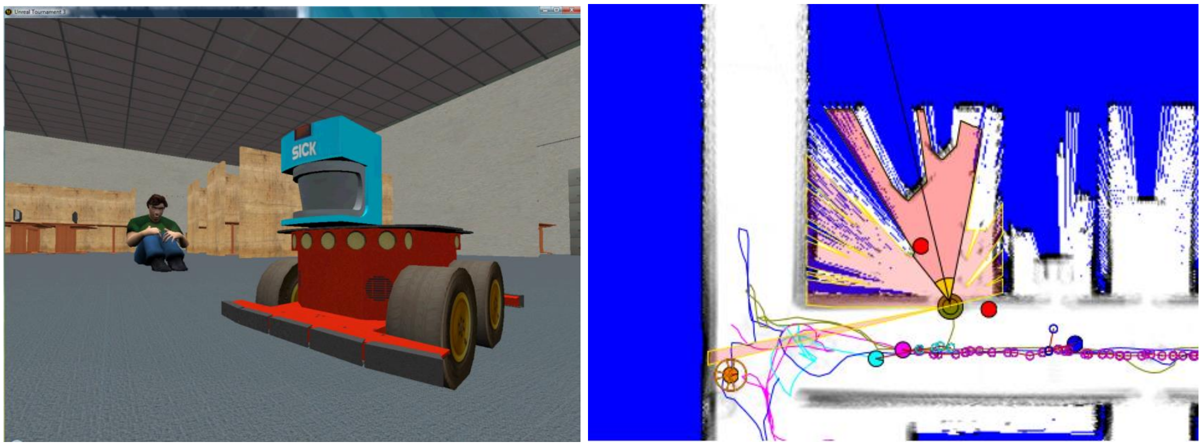


Figure 23. P3AT robot (left) with laser scan area (light pink) and camera viewed area (dark pink) showed on the map right

The P3AT robots (Figure 3) simulated in urban search and rescue simulation (USARSim) were equipped with a pan/tilt camera with a 45-degree field of view (FOV), a front laser scanner with 180-degree FOV, and a resolution of one degree. Using the camera FOV and the minimum and maximum camera distance from which points of interest (POI) can be detected, the laser scan was reduced to the camera's parameters and a 2D shape of the geometric area the camera image captures was calculated. Using the pose estimate, the reduced laser scan shape is translated from local to global coordinates, which creates the view shape. Similar to PiP interfaces, this process allows the view shape to be overlaid on the operator's map and eliminates the context switching and localization operator tasks. The camera image and view shape are then packaged as an information frame and added to the information database.

3.4 PRIORITY QUEUE

The next element of the Event-Lens system is the priority queue, which maintains a prioritized copy of the information database for the operator. When frames are added to the information database, they are also selected and added to the priority queue, which is sorted by the image filter by using a heuristic design to reduce redundancy in image processing. The heuristic image filter calculates utility scores based on the viewed area (visual coverage) of each frame. The priority queue maintains a total viewed area that is equivalent to the combined total of each viewed area from every information frame (image) processed by the image filter. The utility score of a given information frame is the area of its viewed area, after the total viewed area is subtracted from it.

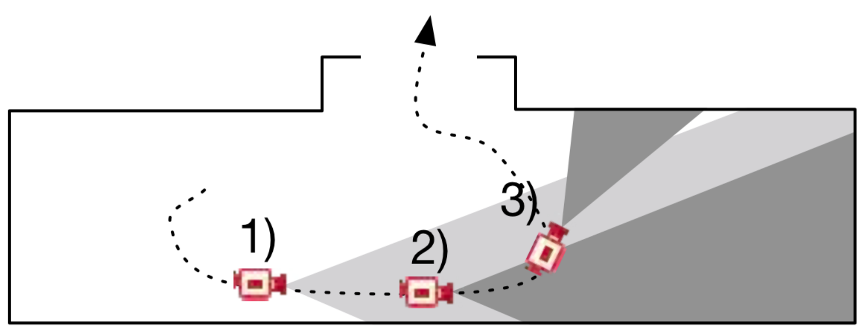


Figure 24. Example database with an empty total shape containing images 1, 2 and 3

For instance, Figure 24 shows a representation of the information database with three information frames and an empty total viewed area (no frames have been processed by the image filter). Images with a larger amount of new visual areas receive higher utility scores. Areas that have already been covered by selected images in the priority queue do not count towards utility. In colloquial terms, this kind of utility picks images that cover large areas with minimal overlap.

Figure 24 illustrates this concept of utility with a simple example. The frame taken at position 1 has the largest visual coverage and highest utility; while considering that the frame at position 1 has been selected by the image filter, note that the frame at position 2 has no utility, since it is entirely overlapped by the frame at position 1. The frame at position 3 has some utility, since it provides coverage in an area not covered by the frame at position 1.

3.5 SPATIOTEMPORAL INFORMATION INTERACTION

3.5.1 Overall Assessment

Since the Event-Lens system is designed to facilitate visual searching for top utility images during a user-specified time period, it is necessary to establish a connection between the images and their essential visual coverage and time stamp information. We interpret this connection by providing a big picture of spatial coverage of the top selected photos in the map, as well as their shooting position and direction during a given time range.

The individual image's visual coverage is filled with a specified color that is correlated to the thumbnail of the image list (shown in the right part of Figure 25). We used color as well as numbers (the white figure) to represent the sequence of images that cover larger unselected regions of the map, which we call the "utilities." The smaller number of the area that is marked, the larger the utility in this image. We originally considered displaying this information with all of the camera positions and directions together, but realized that this method would not provide a visual cluster and might result in chaos when regions were documented by several images. Instead, the image as well as camera positions and directions only showed up once the user selects a visual cover area. Displaying the image with its coverage on the map could be prompted by clicking the image in the top image list (shown in the right part of Figure 25) or the colored area on the map. We consider this to be a better way of helping user finds images in a specified area.

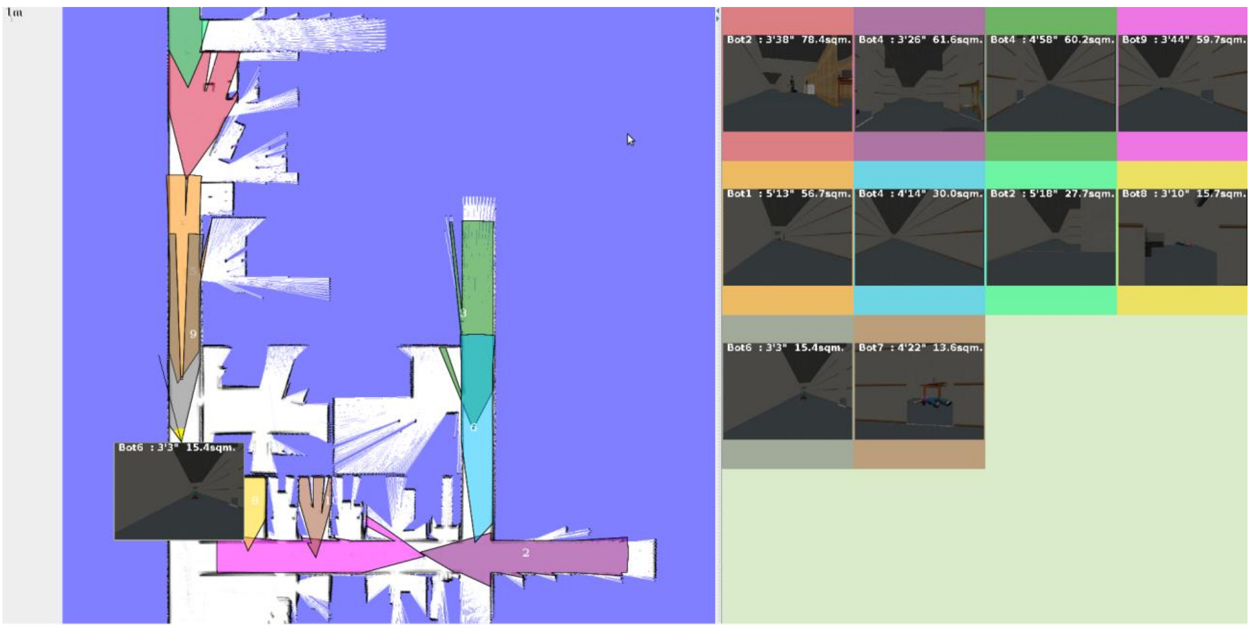


Figure 25. Top utility image list with visual coverage

3.5.2 Temporal Navigation

The toolbar of timeline (Figure 26) is placed over the overview map (the top part in Figure 22). Each 15-second interval is represented by a segment in the timeline. The graduation of each minute is showed for the user's overall assessment of time. The red line represents the number of photos that the system has gathered within that segment. In addition, the statistics of the total visual coverage of selected images (top ten images of utilities) is showed by the blue histogram for each segment period. Neither histograms of covered areas nor the number of images are cumulative values, which are exclusive to the 15-second interval. For instance, more photos are chosen in the third segment than in the second segment; however, the total utility of the images in the second segment is larger than the coverage of images in the third segment.

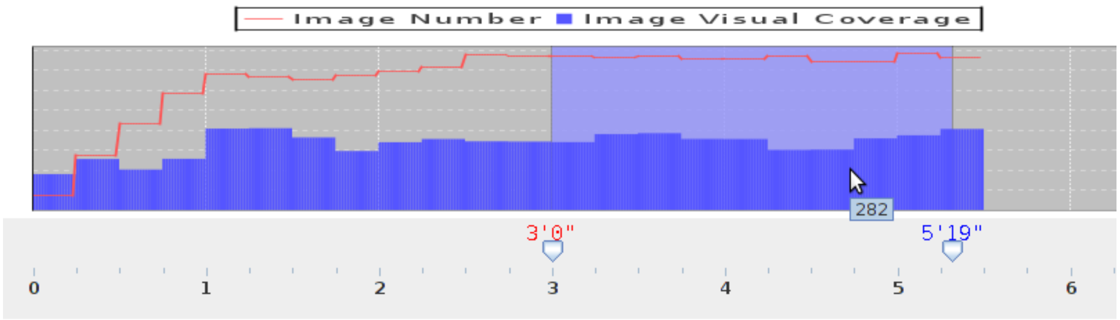


Figure 26. The toolbar of timeline

Two sliders sitting on the timeline can be dragged and released to narrow down the time range of interest. Sliders only stay at the edge of each second for system performance considerations. If the user releases the slider at an interval between two seconds, the slider will automatically jump to the closer edge. Users could also select each 15-second segment by simply clicking the blue histogram in each segment. Images with visual coverage in the map implement the concept of dynamic queries and update correspondingly when the user changes the time range in the timeline. For example, the coverage of images may shrink if the time range is narrowed down from three minutes to one minute.

3.5.3 Spatial Navigation

A region may also be selected by dragging and releasing a rectangle (shown in the lower left of Figure 27). The selected region is highlighted in red. If a subarea, room, or region is selected, a sub-queue (shown in the right side of Figure 19) that contains images retrieved from the database taken within this region will be opened, and a timeline toolbar that only shows the statistics of images within the selected region will be displayed on the top of the sub-queue. The total visual coverage of the selected rectangle area is showed by the blue histogram that denotes each second. For example, it is easy for a user to realize that the selected area under surveillance ranges from 7 seconds to 2 minutes, 30 seconds in Figure 27.

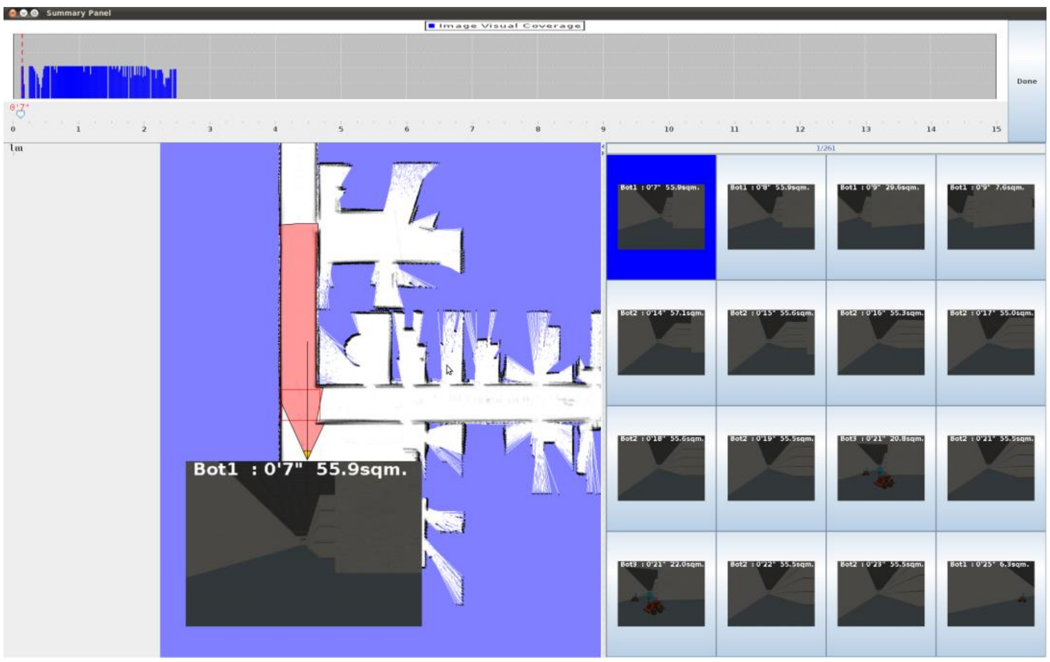


Figure 27. An illustration showing the sub-queue of a user's selected area

The images shown in the sub-queue are also processed by an image filter, which attempts to reduce redundancy by only showing either spatially unique or temporally unique images from the database. This sub-queue image filter inherits the operating principle behind the priority queue, but tries to emphasize the temporally unique images, which is critical for showing status changes in the specific region. The detail will be discussed further in the methodology chapter.

The image in the sub-queue may be viewed through multiple procedures. First, a user could view the image by scrolling the wheel of the mouse to “walk” through the relevant images in a given time sequence. Second, the user could carry out a visual search on in the sub-queue image list part (shown on the right side of Figure 27) and click directly on any particular image. Third, a slider sitting on the timeline can be dragged and released to specify a particular time point. Images near the selected time point will be shown on the top of the sub-queue image list. Once an image in the sub-queue is selected, the zoom-in image with its spatial coverage will be shown on the map as well.

The sub-queue of a region can help an operator to properly localize the moving target, and to find better spatial and temporal perspectives with which to view the tracked target. More importantly, the sub-queue also can provide a user with the temporal navigation ability within the information regarding the region in question. Once a given subarea is selected, a timeline bar will also be provided, with updated statistics that indicate the time distribution of photos that were taken in this subarea. This feature provides the necessary contextual information, but in a much more efficient way. In other words, a user is given the ability of expanding or shrinking the timeline that is offered by the asynchronous display and database design.

3.6 MARKING OF MULTIPLE MOVING TARGETS

In the simplest sense, trails allow the user to quickly characterize movement throughout a time period without the need for animation. The shape of a trail allows rapid visual identification of patterns or anomalies in behavior, and the slope of a trail shows the maximum possible average velocity of an entity during a segment of travel (Kapler et. al, 2007). In other words, the best way to show a history of a moving target is to show its trail or trajectory. As a result, we developed a target trajectory analysis function to assist users with marking a moving target.

When participants detected a target in an image, they could click the image to see the visual coverage of the selected image. With the assistance of the shape and direction of the points-of-view (shown in Figure 28), the user could double-click on the map to directly mark the target (in this case, a robot with a yellow head). As we were dealing with a problem of multiple moving targets (MMT), target identification was needed to finish the target marking task. Since we use a different color for multiple moving targets to simplify the problem, the user only needs to select a color from the waiting list (shown in the right part of the panel in Figure 28).

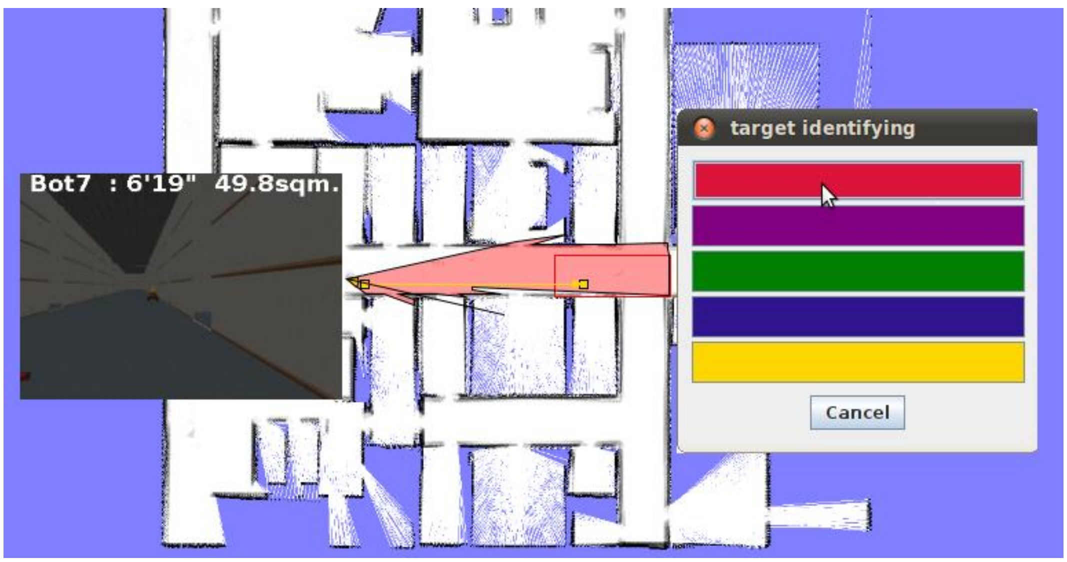


Figure 28. Target identification and marking interaction

Once the target is marked with a certain color (identified), the image as well as the laser scan and time stamp information will be permanently correlated with this mark. All the targets marked with the same color will be connected by an arrow of the same color to show the direction in which the target has moved and the marking time sequence (Figure 29).This target trajectory analysis function will assist human users with marking the MMT. It will lower the user's mental workload by reducing memory requirements, as well as the spatial and temporal reasoning task for the MMT.

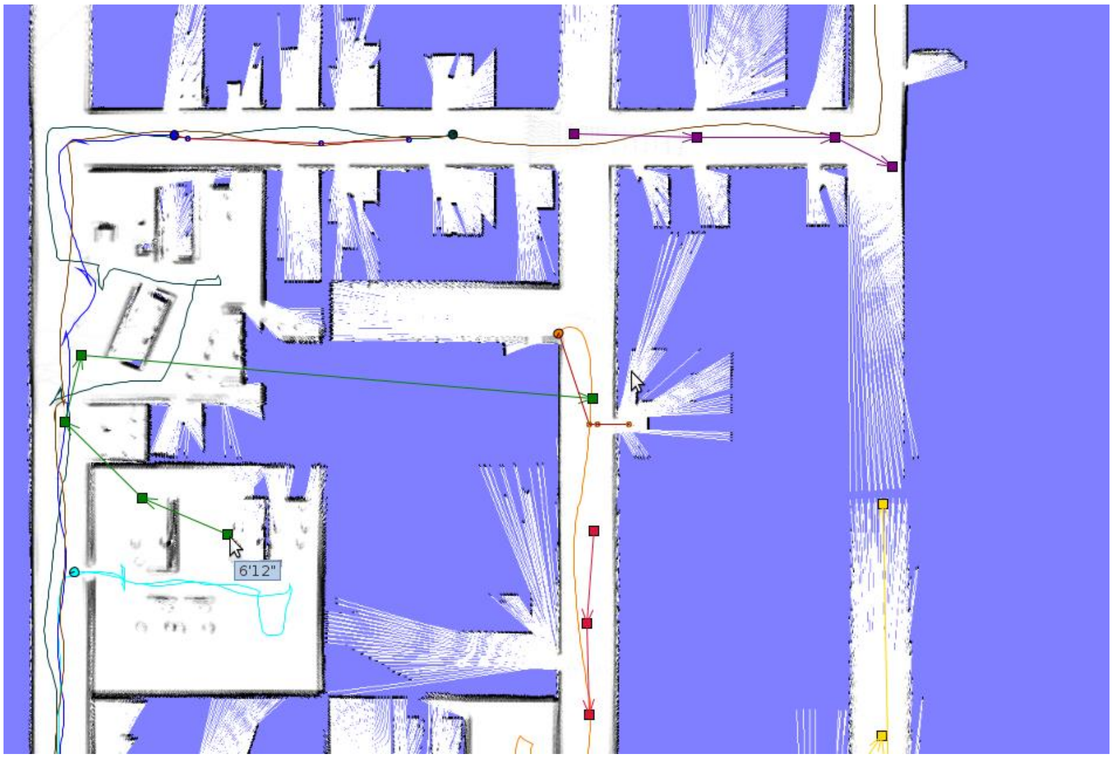


Figure 29. An illustration of target trajectory analysis function

4.0 METHODS

4.1 USARSIM

The Unified System for Automation and Robot Simulation (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 (Lewis et al., 2011).

USARSim uses Epic Games' UnrealEngine3 (UE3, 2010) to provide a high fidelity simulator at a low cost. The Unreal Engine provides fast, high-quality 3D scene rendering that supports mesh, texture, lighting, and material (such as reflective, transparent, and semi-transparent surfaces) simulation, which allows us to simulate realistic camera video (Wang, 2007). This is one of the most critical features of current approaches to human control of mobile robots. Validation studies showing agreement for a variety of feature extraction techniques between USARSim images and camera video are reported in (Carpin, Stoyanov, Nevatia, Lewis, & Wang, 2006b). Other sensors, including sonar and audio, are also accurately modeled. Validation data showing close agreement in detection of walls and associated Hough transformations for a simulated Hokuyo laser range finder are described in (Carpin, Wang, &

Lewis, 2005). The current UnrealEngine3 integrates MathEngine's Karma physics engine (UE3, 2010) to support high fidelity rigid body simulation in instances of collision, friction, joint simulation and force, and torque modeling. This feature allows the simulation to replicate both the physical structure of the robot and its interaction with the environment. Validation studies showing close agreement in behavior between USARSim models and real robots being modeled are reported in many studies (Carpin, Lewis, Wang, Balakirsky, & Scrapper, 2006; Lewis, Hughes, Wang, Koes, & Carpin, 2005; Pepper, Balakirsky, & Scrapper, 2007; Taylor, Balakirsky, Messina, & Quinn, 2007; Zaratti, Fratarcangeli, & Iocchi, 2006).

The Unreal Engine uses efficient client-server architecture to support multiple players. This embedded networking capability also allows USARSim to support human control of multiple robots without modification (Wang, 2007). Figure 30 shows the Unreal Engine's components and the expandable library of robot-themed models, environments, and control interfaces to acquire sensor data and issue the commands that we have added to create the USARSim simulation.

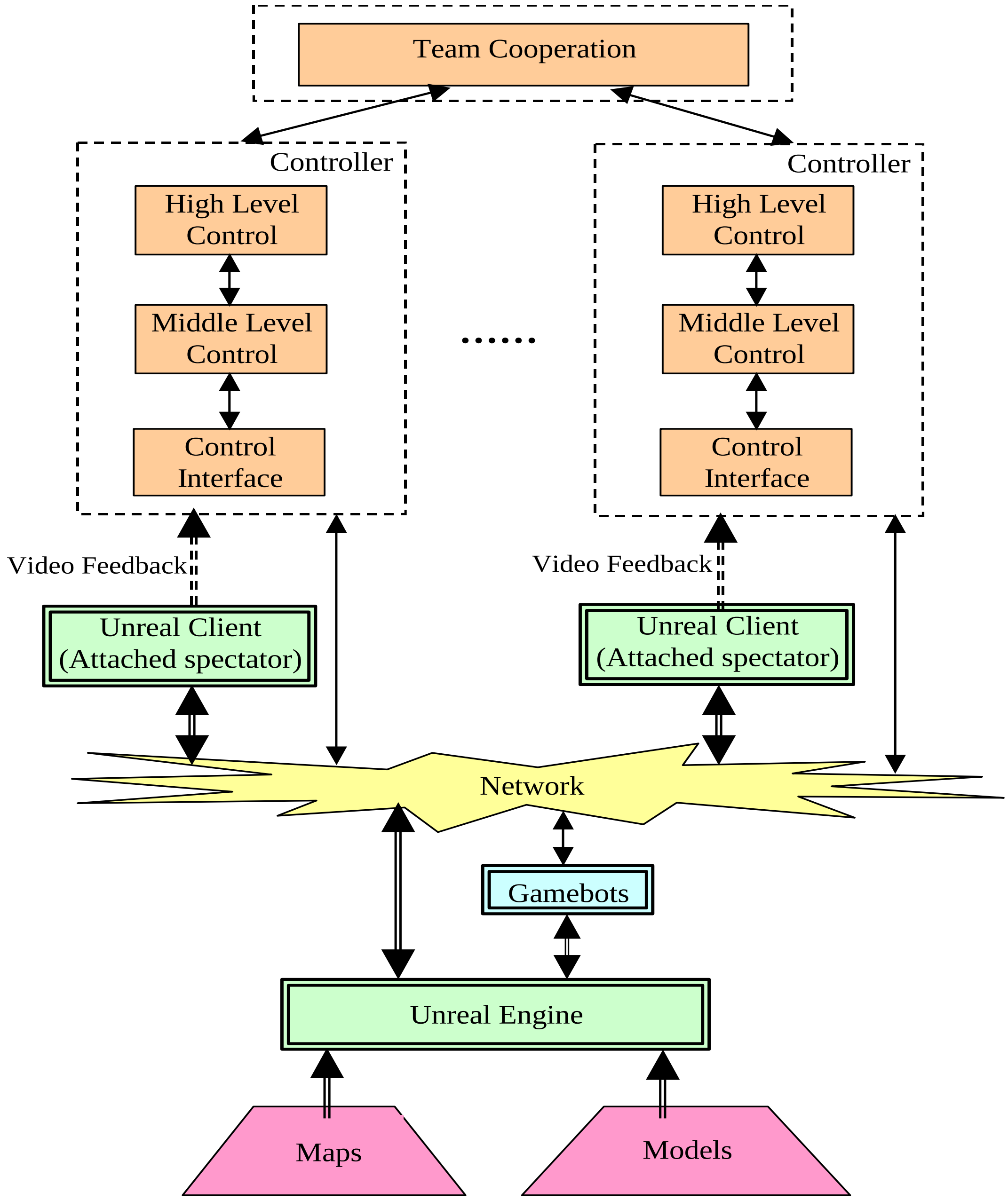


Figure 30. USARSim architecture

(Wang, 2007)

4.2 MRCS – THE MULTI-ROBOT CONTROL SYSTEM

To suit the research questions of this thesis, the system had to be scalable, to allow us to control different numbers of robots; reconfigurable, to enable us to study different human-robot interfaces; and reusable, to facilitate testing different control algorithms (Wang, 2007). With these requirements in mind, we selected the distributed proxy-based multi-agent framework Machinetta as our system's baseline, which is named here as a multi-robot control system (MrCS). MrCS is a multi-robot communications and control infrastructure with an accompanying user interface that was developed for experiments in multi-robot control and RoboCup competitions (Balakirsky et al., 2007). The Event-Lens system was designed and developed based on the MrCS infrastructure. MrCS provides facilities for starting and controlling robots in the simulation, displaying camera and laser output, and supporting inter-robot communication through a distributed multi-agent system, like Machinetta.

Each robot connects to Machinetta through a robot driver that controls the robot on both low and middle levels of control. For low-level control, the driver serves as a broker that translates robot sensory data into local beliefs and also translates the exploration plan into robot control commands (e.g., wheel speed control). For middle-level control, the driver analyzes robot sensory data to perceive its states and local environment. Then, based on this perception, the driver overrides the control commands to ensure safe movement when it is necessary. Possible adjustments may include changing the direction of motion to avoid obstacles and recovering from becoming immobilized or from a dangerous pose. When the robot is in an idle state, laser data analysis allows the driver to generate potential exploration plans (such as the eventual destination and the path to it).

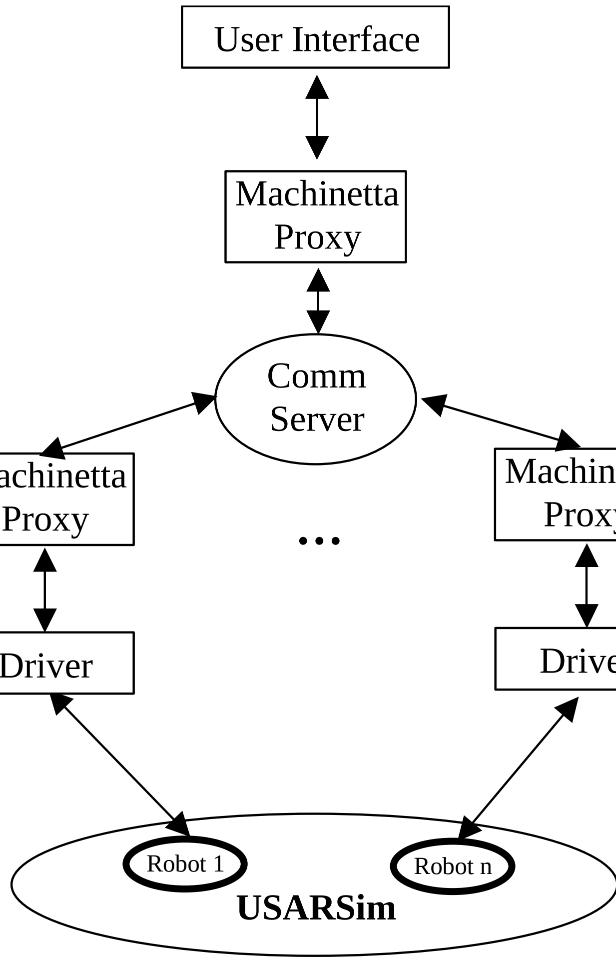


Figure 31. MrCS architecture

(Wang, 2007)

Figure 32 shows the elements of the conventional MrCS GUI for the streaming video condition. The operator selects the robot to be controlled from the colored thumbnails at the top right of the screen. These thumbnails also show the live video streams from all robots. To view more of the selected scene shown in the large video window, the operator uses pan and tilt sliders to control the camera of the selected robot through a teleoperation widget (upper right). The current locations and paths of the robots are shown on the Map Viewer (bottom left). When under manual control, robots are assigned tasks by designating waypoints on a heading-up map on the Map Viewer or through the teleoperation widget. Apart from the interface for displaying relevant information and allowing controlling robots manually, MrCS also contains a number of autonomous functions, which are described in more detail in Section 4.3.

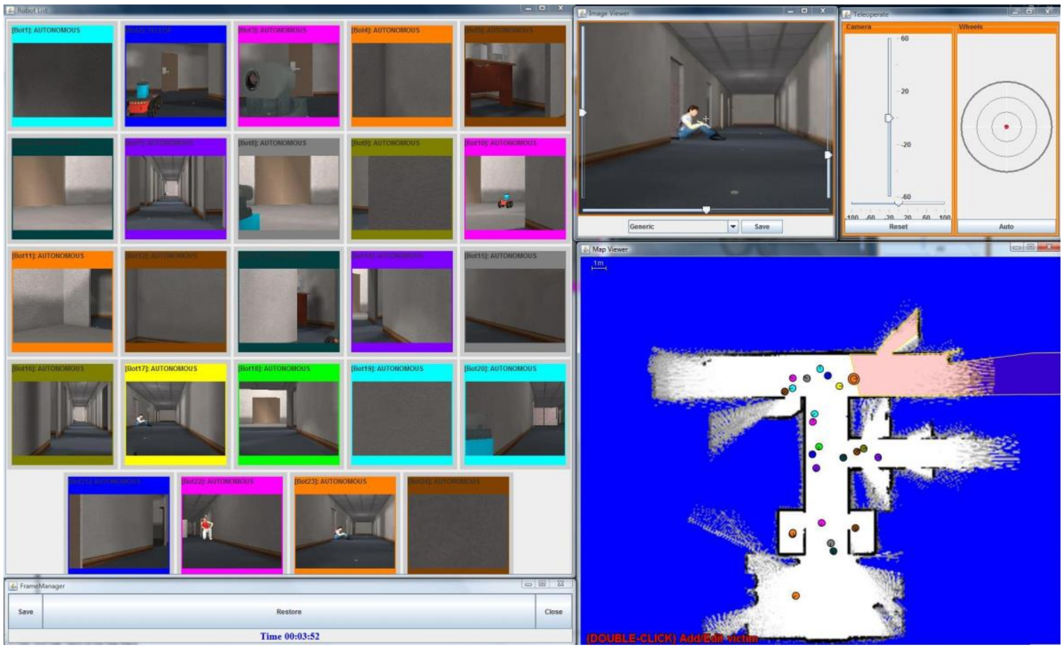


Figure 32. The conventional MrCS user interface for the streaming video condition

4.3 SYSTEM AUTONOMY

Our previous study (Lewis et al., 2010) demonstrated that a high degree of autonomy can improve performance in foraging tasks, especially in search and rescue scenarios. Consequently, the current version of MrCS includes further autonomous functions, while also improving existing capabilities. A new Segment Voronoi Diagram (SVD) path planner replaced the random tree planner that was used in earlier studies (Takahashi & Schilling, 1989). The new planner generates paths that maintain a safe distance to nearby obstacles. Such paths are generally longer, smoother, and more human-like. Previous experiments (Chien, Wang & Lewis, 2010) showed that operators are easily able to follow paths that are less smooth, but the added safety of the paths also benefits autonomous navigation. The use of a path planner and autonomous navigation to drive the robots is one major distinction to the panorama study (Velagapudi et al., 2008) where participants manually generated paths to reach specified panorama locations.

A new simultaneous localization and mapping (SLAM) 2D map algorithm has replaced the previous Carmen occupancy grid that was used in MrCS. The SLAM algorithm uses the pose and laser scan data from the robots to build an evidence grid describing the probability of cells on the map that contain an obstacle. The new algorithm has greatly improved the quality of the maps when noisy data was provided. To autonomously guide the exploration process and further relieve the operator, we used a frontier-based exploration approach. The SLAM algorithm identified frontiers cells on the map, which are areas on the boundary between known and unknown occupancy space. Each robot identifies these frontier cells locally within their own map (which are created from their laser scans), ranks them according to the expected benefit from visiting them, and communicates these cells to other robots. Machinetta is then used to coordinate the assignment of exploration locations to robots. Paths are generated for every robot to visit its assigned locations to gather further laser and video data, improve the map, and the cycle is repeated by computing new exploration locations.

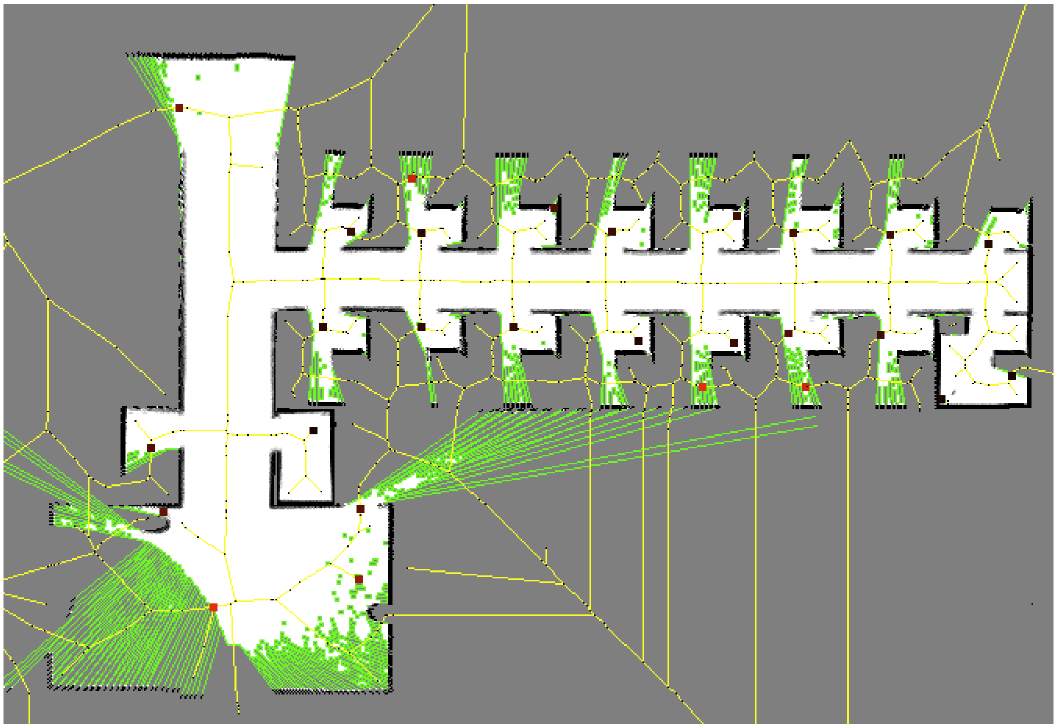


Figure 33. A sample map with the structures used and computed by the autonomy

Figure 33 visualizes a few of the structures that were generated by the SVD planner and SLAM algorithm. In the sample map, there are spaces with obstacles, such as walls (black areas), obstacle-free spaces (white areas), undetermined spaces (gray areas), frontier cells (green boxes), path planner roadmap (yellow lines), and frontier locations for exploration (squares). High-value exploration locations are red squares. Robot autonomy is handled using the Machinetta communication framework to distribute tasks using the LA-DCOP (Scerri et al. 2008) distributed task allocation algorithm. There are two types of tasks in the system: go-to tasks that specify a location in world space for the robot to visit, and look-at tasks that specify a pose in world space for the robot to achieve. Frontier points received from the SLAM algorithm are packaged as go-to tasks and are entered into the LA-DCOP algorithm.

Unfortunately, frontier-based navigation does not operate in a way that ensures complete visual coverage of the environment, as the FOV parameters of the laser scanner and RGB camera are different (Figure 34). While the relatively narrow FOV of the camera operates well in urban environments, as they consist predominantly of lots of narrow hallways and small offices, there are still small patches of incomplete visual coverage that need to be filled. Using the SLAM occupancy grid, random valid poses are generated along with the camera view shapes. The utility of the view shape is determined using the information database, and if this utility is above a system threshold, the pose is packaged as a look-at task and sent to LA-DCOP.

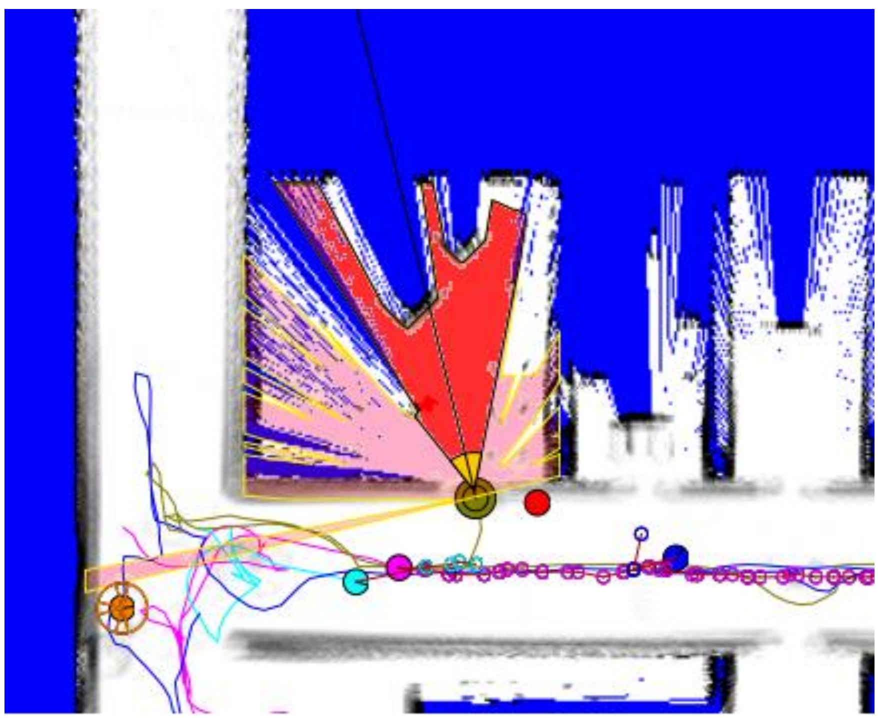


Figure 34. FOV of the laser scanner (180 degree) and RGB camera (45 degree)

4.4 IMAGE PROCESSING IN IMAGE FILTER

4.4.1 Image Utility Calculation

The operating principle behind the priority queue is to reduce redundant images. Take the previous example: when image 1 is processed, its view shape is added to the total shape and the utility of each frame in the database is updated, which results in the database representation shown in the lower part of Figure 35. After this update to the total shape, image 2 has no utility and image 3 has a very small utility.

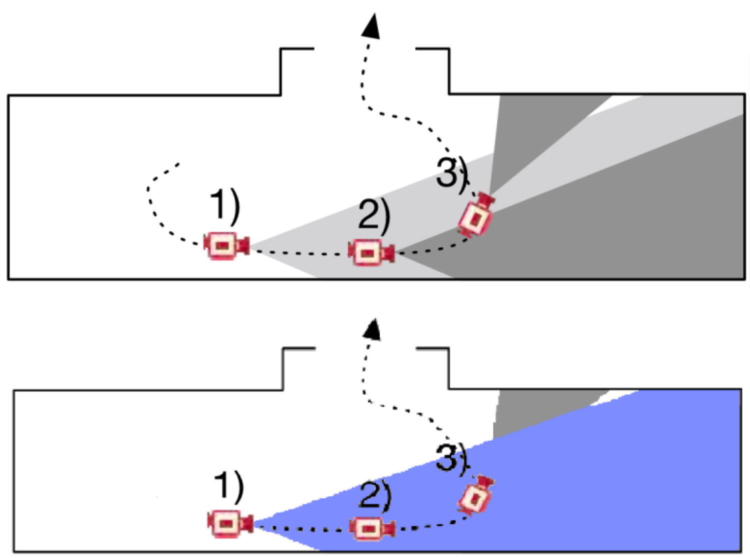


Figure 35. Example database with a blue total shape after selecting image 1.

More precisely, we compute a utility $u(I)$, for every image I, as follows. Let $F \subset \Re^2$ be the workspace of the robots, i.e., the free space in which they can navigate. F is uncovered by the robots as they explore and build a map. Let S be all images already been selected and processed by the image filter, and let D be all images stored in the database. Initially, S is empty and D grows as robots explore the environment. Every image I in D has an associated area $a(I) \subset F$, which is the area visible in the image. This area can be computed by referencing the image in the map of F. The relevant area for an image I is all of the as-yet unseen area.

$$\partial(I) = a(I) \setminus \bigcup_{I' \in S} a(I')$$

Now we define $u(I) = area(\partial(I))$, according to how we sort our images in D, which becomes the total viewed area. Among these, we present all those images that have the highest priority. Obviously, once the image filter selects and processes an image, the set S changes the total number of viewed areas that have to be updated. To avoid frequent updates, we can compute the utility for I supposing that all higher-priority images are already in S. As a result, the utility of an image only updates when new imagery is added to the database D at a location nearby.

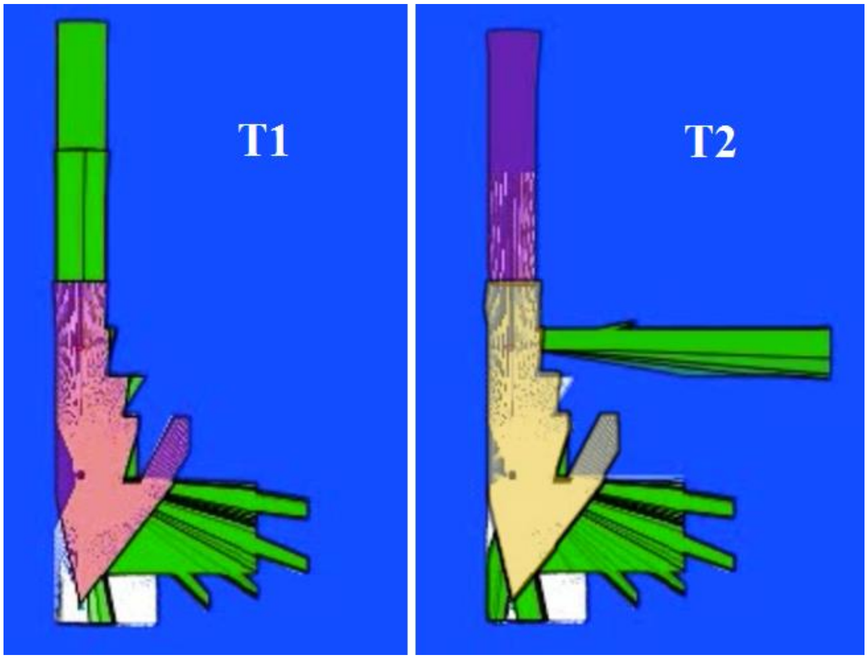


Figure 36. An illustration of how to calculate redundant imagery

Figure 36 shows the calculation procedure of prioritized imageries in the real database. In T1, the pink area indicates the highest-priority image coverage in the database. Once this pink area is added to the priority queue, the pink area becomes yellow to indicate the total viewed area that has been collected in this area (shown in T2). The next highest priority image is recalculated (pink area). Green areas are images with lower priority.

4.4.2 Spatial and Temporal Uniqueness in Sub-queues

In the Event-Lens system, a user could also specify a region by dragging and releasing a box, generating a sub-queue, and navigating through the time within this region. It is possible that an operator will have multiple images collected, which will allow him or her to view the region from a different angle. Using the SLAM occupancy grid, the system creates a number of valid poses approximated to the selected region. Figure 37 shows such a condition. The operator could make use of the sub-queue function on the map to monitor changes in circumstances within the selected area for a long duration of time.

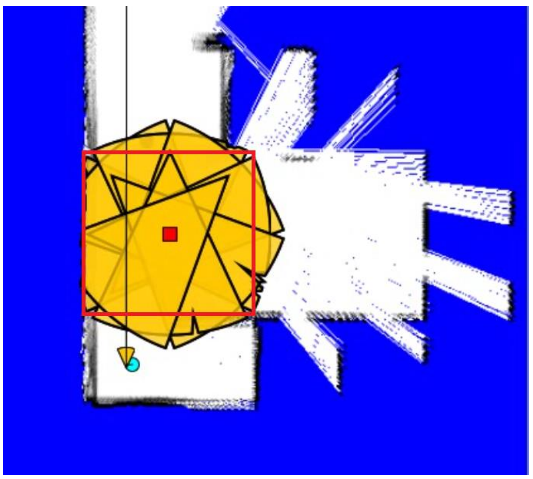


Figure 37. Visualization of images collected in a region

The image filter is also responsible for the image process of the sub-queue with three steps, as shown in Figure 38.

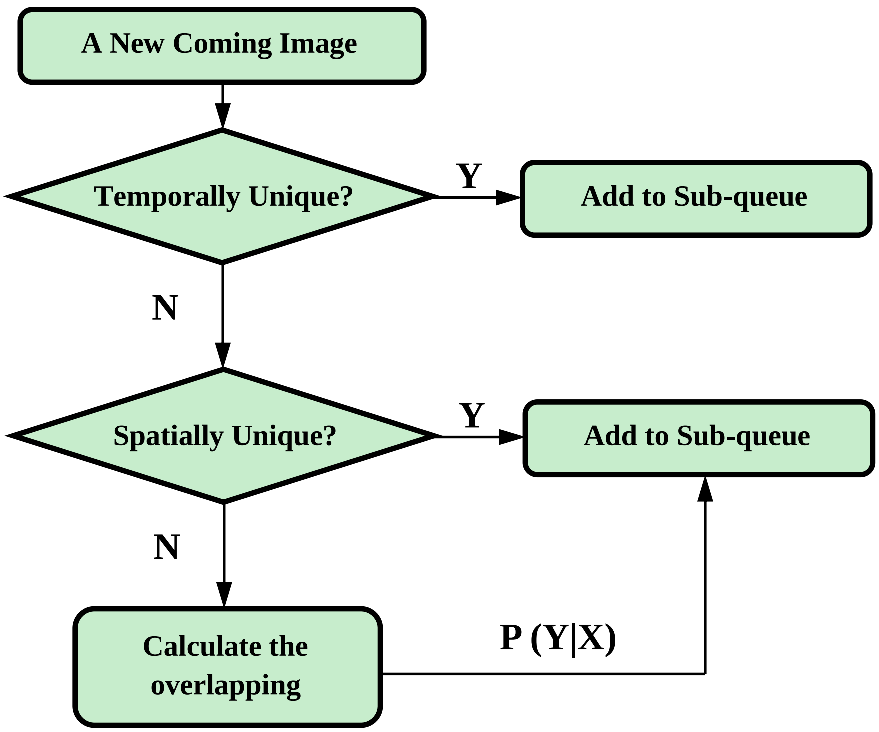


Figure 38. Image filtering process in sub-queue

First, the frames in the information database will be filtered if they do not fall into the selected region. Once the image within the region is selected, we will filter out the temporally unique pictures, as we want the operator have the ability to monitor a region for the whole process without losing any valuable pieces of information, as well as avoiding any image redundancy. When the time interval of the two pictures is more than one second, these two pictures will be considered as temporally unique and will be retained. When the two pictures' time interval is less than one second, a further process will be carried out to determine the uniqueness of these two pictures in space. We define spatial uniqueness in the following way: the two images have absolute spatial uniqueness if there is no intersection of the two FOV of the image. Otherwise, the two images are only partially spatial unique. We introduce a function to determine the probability of retaining image Y, given image X and selected rectangle A. The probability of retaining image Y is proportional to the ratio of the occupied area of intersection of O_X and O_y within A to the union of occupancy grid O_X and O_y within A, using the following equation:

$$P(Y) = 1 - \frac{(O_X \cap O_Y) \cap O_A}{(O_X \cup O_Y) \cap O_A}$$

As a result, if the images are completely spatial unique, P(Y) = 1, whereas if the FOV of the two images are identical, then P(Y) = 0.

4.4.3 Pre-Index of the Images

One of the challenges is to design Event-Lens as a powerful system that is both interactive and dynamic. The most important factors in such a system are speed and scalability, which are crucial to a productive and pleasant user experience. However, we are facing a large data set, which increases dramatically when we are dealing with image data. For instance, 12 robots with camera and laser scan sensors will generate about 1GB of data per minute, including about 21600 images (480*360), as well as correlated coordinates, FOV information, and time stamps.

Furthermore, problems will become more complex when we are trying to filter the image, as calculating the utility will dynamically affect the database. For example, if image X was considered to be a top-utility image, then the FOV of image X will be reduced from the unviewed shape. This simple step will actually affect all the other images in the database. Thus, top ten means ten times of refresh the database, let alone the user may change the concerned time interval. On the other hand, it will take a huge amount of CPU offers and a great deal of valuable time to simply read the endless image list.

Comparing the method by which all pictures were retrieved (called the "fetch all" algorithm), we want to pre-index each image as it comes to the database. We sequence the image by its utility with a 15-second interval, and save it to the cache. Once the user specifies the time interval, the image filter will first search the covered 15 seconds, as well as the uncovered time pieces (parts of less than 15 seconds). We use the advantages of database and caching in this algorithm, as its name would indicate.

We did a pre-experiment test to compare results between these two algorithms. For the dataset size, we started from a 60-second image set (about 20000 elements), which is fairly large for existing visualization systems. As a result, it takes about 17 seconds to process the images from a 600 second (15 minute) time interval with the "fetch all" algorithm; in contrast, it takes less than 1 second to process the images with the pre-indexed cache images (Figure 39).

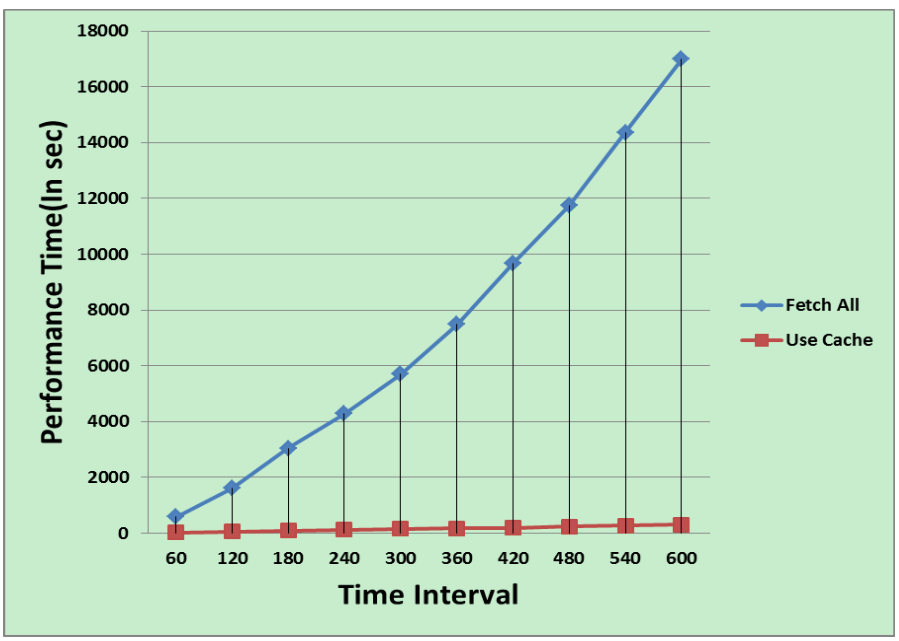


Figure 39. Visualization of images collected in a region

4.5 INDIVIDUAL DIFFERENCES

Coordinating information from multiple sensors to search an environment for moving targets requires operators to continually shift their attention from sensor to sensor or image to image, change their perspectives to maintain situation awareness (SA), and locate targets as they appear within a camera's field of view (spatial relations). As a result, the individual difference such as attentional control ability, spatial ability, and video gaming experience (which is connected to SA and operation efficiency) will be important factors for this multi-task procedure.

The way in which to test and measure these individual differences will be discussed. Furthermore, we sought to evaluate whether individual differences in spatial ability, attentional control, and video gaming experience might affect an operator's performance.

4.5.1 Attentional Control

When interacting with information from multiple sensors, especially moving sensors, the demands on the operator are likely to be extremely high, as there is a need to switch attention between sensors and develop an understanding of the environment from many different perspectives. Operators may encounter problems in integrating information because of the reference across different sources is not well presented, which has been shown in several studies (Olmos, Wickens, & Chudy, 2000; Thomas & Wickens, 2001, Chen & Clark, 2008).

In cognitive neuroscience, attentional control is defined as an individual's capacity to choose what they pay attention to and what they ignore (Astle & Scerif, 2009). Attentional control also refers to one's ability to focus and shift attention in a flexible manner (Derryberry & Reed, 2002). Bleckley et al. (2003) have shown that individual differences of attentional control could affect multitasking performance. Feldman Barrett, Tugade, & Engle (2004) examined individual differences in the capacity to control attention as a major contributor to differences in working memory capacity. Kahneman, Ben-Ishai, & Lotan, (1973) also showed that the measure of proneness to attention disruption was significantly related to accident rate.

In a recent U.S. Air Force survey of subject matter experts on the operators' performance of multiple unmanned aerial systems (Chapelle, McMillan et al., 2010), attentional control is one of the most important abilities that would affect an operator's performance, since the robotics control task is inherently multitasking (e.g., sensor manipulation, tracking, communication) (Chen & Barnes, 2012).

Feldman Barrett et al. (2004) tried to explain this phenomenon. They used the term "cognitive miser metaphor" (conserving the amount of cognitive resources), which refers to severely limited attentional resources that result in adopting strategies that simplify the need for controlled attention. In other words, when dealing with complex information processing, people with lower attentional control tend to reduce their attentional control requirements by taking the "cognitive miser" approach.

Therefore, we hypothesize that participants with better attention control can allocate their attention more flexibly and effectively. This flexibility in switching attention could be considered as a prediction factor of multi-tasks performance, such as the marking of multiple moving targets and monitoring robots.

4.5.2 Spatial Ability

In the current practice of the USAR task, robots equipped with laser range finders use simultaneous localization and mapping (SLAM) to build a map based on laser scans and position the robot on the map relative to those scans. At the beginning of the task, the map is entirely unknown, but as exploration continues, features such as walls, objects, and open spaces appear on the map. The laser map, however, cannot provide sufficient resolution to perform complex perceptual tasks, such as victim identification. The images or video can only provide the operator with partial information about the environment, because they are limited by the camera's field of view, the robot's orientation, and its trajectory through the environment. In order to locate the targets and clarify the relationship between a camera view and the robot's location on the map, operators must expand their perceptive ability to maintain global and local mental models of the environment.

Consequently, the differences in orientation between the map and camera views might require mental rotation. Several researches have shown that the track-up map, which is ego-referenced with rotating viewpoints, is better for local navigation, and the north-up map, which is world referenced with fixed viewpoint, is better for global awareness (Aretz, 1991; Casner, 2005; Darken & Cevik, 1999; Lohrenz, Gendron, Edwards, Myrick, & Trenchard, 2004; Wang, 2004; Werner, 2002; Chen & Clark, 2008). Individuals with high spatial ability are able to adapt to using either the track-up or the north-up map with less effort than individuals with low spatial ability (Darken & Cevik, 1999).

Werner et. al (1997) conclude that individual diversity in spatial cognition, which includes acquisition, organization, use, and revision of knowledge about the spatial environment, could directly affect the results. Individuals with higher spatial ability have been shown to perform significantly better at navigation tasks than those with lower spatial ability (Cassenti et. al, 2002).

Regarding cognitive models of geographical space, Hirtle (2012) reviewed a number of modeling approaches that are specifically focused on issues of spatial cognition. Specifically, the field of robotics has provided a rich background for the development of spatial cognition, way-finding strategies, and human-robot communications models, which provided a "better understanding of human spatial cognition by examining what components are needed to build a useful and complete system" (Hirtle, 2012).

In another recent study, Chen (2010) found that individuals with better spatial ability performed significantly better in a target search task under night conditions than those with a low sense of direction. Baldwin and Reagan (2009) note that "individuals with poor sense of direction relied more heavily on verbal rather than visuospatial working memory resources, and, conversely, individuals with good sense of direction exhibited more route-learning disruption from a tapping task, suggesting a greater reliance on visuospatial working memory resources." Our previous research also showed that individuals with greater spatial ability exhibited more effective visual scanning and target detection performance during multi-robot control tasks (Chien et al. 2011).

Hegarty & Waller (2004) found that most of the current spatial tests cannot purely examine either object rotation ability or perspective-taking ability, and that the variance in strategies could affect the performance directly. According to their findings, the revised version of the object perspective test (Kozhevnikov & Hegarty, 2001) is both reliable and a largely strategy-free measure of perspective-taking ability. After judging the requirements of the USAR task and comparing various spatial ability tests, we adopted the newer version of the object perspective test (Figure 40) for measuring participants' spatial abilities.

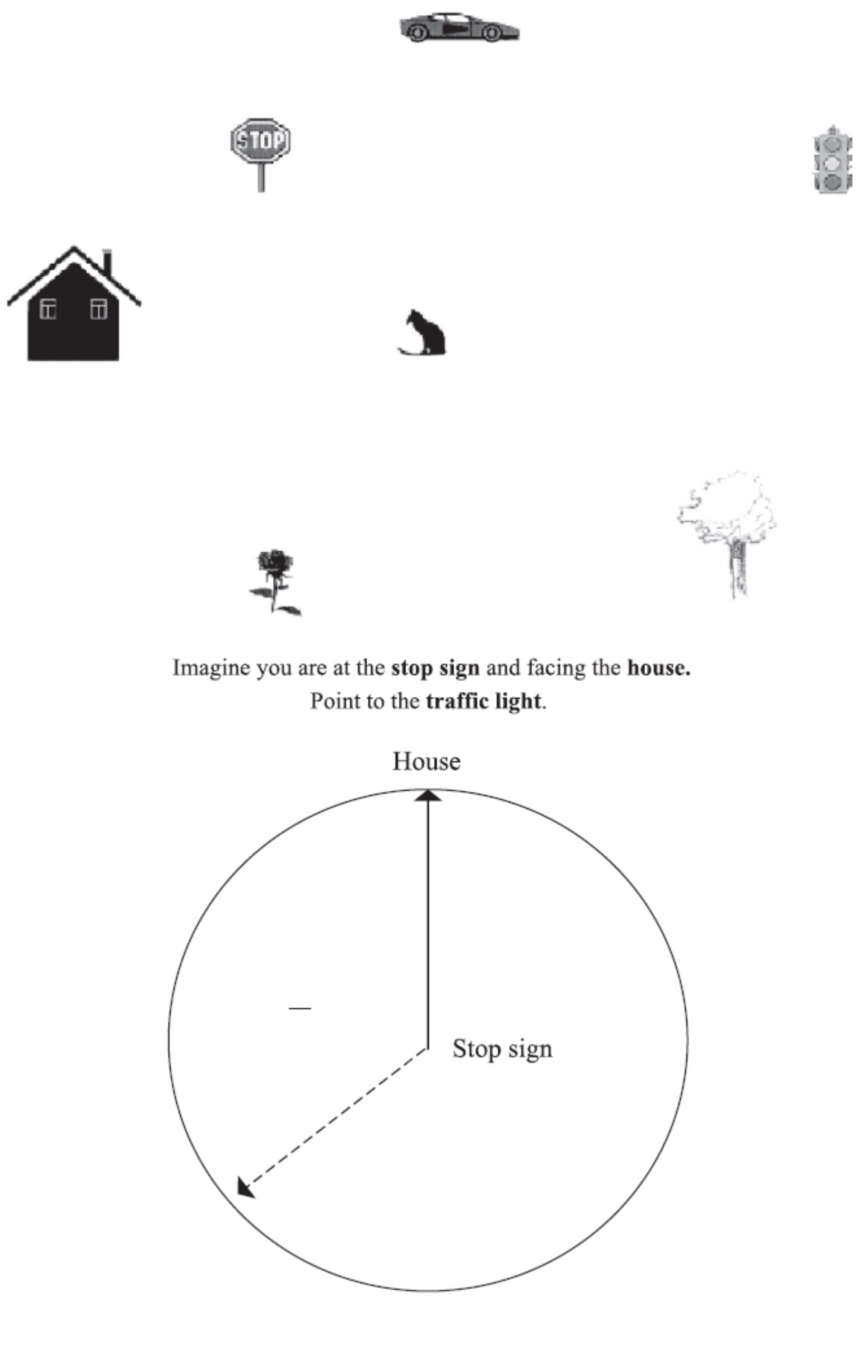


Figure 40. Example of the spatial orientation test (Hegarty & Waller, 2004)

In the Object Perspective Test, seven objects are drawn on the top of a sheet and the bottom half of the page shows a circle marked with a standing point and a facing direction. Participants are asked to imagine being at the position of one object and facing another object, and then are asked to indicate (draw) the direction to the target object. Participants are prevented from physically rotating their body or the booklet and have to identify the target object and then mark it on the circle (as the dotted line shown on the Figure 40), completing twelve questions within five minutes. The initial score on the spatial orientation test is an error score; a higher value represents less spatial ability. Scores are then linearly transformed (by subtracting the average error score from 180 degrees) so that higher scores correspond to better performance. This transformed score, in which higher scores correspond to higher spatial ability, is used in the analyses we report.

4.5.3 Action Video Gaming Experience

Many of the participants in our former researches have commented that interacting with the simulated robots with our MrCS is more or less like playing a video game. The similarity can be derived from the two aspects required: cognitive skills and correlated task requirements.

Regarding the cognitive level, the effects of video game playing on tasks that require a wider range of cognitive abilities, including attention, memory, and executive control, has been well explored. Research has suggested a causal relationship between playing action video games and improvements in a variety of visual and attentional skills. Green & Bavelierv (2003) carried out the pioneering research that action video games could modify visual selective attention, which was published in Nature. Boot et al. (2008) conclude that expert gamers and non-gamers differed on a number of basic cognitive skills: “experts could track objects moving at greater speeds, better detected changes to objects stored in visual short-term memory, switched more quickly from one task to another, and mentally rotated objects more efficiently."

From the task perspective, the similarity of multi-robot control and playing video games could be summarized as following categories: visuospatial selective attention, multiple object tracking, rapid process of visual information and imagery, flexibility in attention allocation, and multitasking.

Concerning visual selective attention, the tasks of controlling robot teams and playing video games are similar. During the task, certain chunks of information deemed relevant to the observer are selected for further processing, while others are ignored. The purpose of this process is to prevent sensory overload and promote effective functioning in the face of the nearly overwhelming and infinite visual information received.

In other words, visual attention is the set of mechanisms by which relevant visual information is selected while irrelevant information is suppressed (Hubert-Wallander, Green & Bavelier, 2010). This is consistent with our system design goal; however, this attention cannot be solely fulfilled by the system. As a result, individual diversity in visual attention, which includes acquisition, organization, use, and revision of knowledge about the spatial environment, could directly affect the task performance.

Consequently, those who habitually play video games have been documented to outperform novices in a variety of visual attentional capabilities, including attention in space, in time, and to objects (Green and Bavelier, 2006a). Correspondingly, experienced action video game players, when compared with infrequent or non-gamers, were found to perform significantly better on tasks that required visuospatial selective attention, multiple object tracking, rapid process of visual information and imagery, and flexibility in attention allocation

(Hubert-Wallander, Green & Bavelier, 2010; Chen & Barnes, 2012). In addition, there is also a strong relationship between video game experience and multitasking performance (Hambrick et al., 2010).

Therefore, we hypothesize that experienced action video game players would outperform infrequent/non-gamers in our visually demanding Event-Lens interface and interaction procedure in terms of asynchronous control, target and event detection, and monitoring of the robot team and the dynamic environment.

To summarize, the evaluation studies sought to examine the relationship between participants' action video gaming experience and their task performance, as well as the global and local SA of the spatial temporal environment.

5.0 EVALUATION

To ensure validity of the visualization and interaction design, we decided to conduct a series of laboratory studies. Deploying the Event-Lens system and asynchronous visualization methods to the real world would be a useful next step to improve the system's usability and to gather user requirements.

The research questions we hoped to answer from the studies were suited to an exploratory approach that allowed us to learn users' spatial and temporal search strategies and explore possible directions and design guidelines for future visualization technologies that could support large-scale multiple sensors for information fusion and management. Additionally, to the best of our knowledge, little previous information is available on similar problems or research issues regarding asynchronous display and interaction method. We therefore conducted two exploratory studies: first, evaluate the Event-Lens system and help define user requirements for asynchronous spatiotemporal navigation, which could improve future spatiotemporal visualization designs, and at the same time, try to identify the strategies that users used in searching and locating moving targets; and second, compare the user behavior and performance between synchronous and asynchronous control methods for large-scale multiple sensor systems with tasks that involve multiple moving targets.

5.1 APPARATUS

The reported study and experiments were conducted using the USARSim robotic simulation with 12 simulated UGVs performing Urban Search and Rescue (USAR) foraging tasks. A large USAR environment previously used in the 2010 RoboCup Rescue Virtual Robots competition was selected for use in the experiment (Robocup Rescue VR, 2010). The environment was an office-like hall with many rooms that were full of obstacles like chairs, desks, and bricks. Targets were evenly distributed within the environment. Robots could enter the environment from one of two possible entrances. The environment was 5026 m^2, a size sufficient to guarantee that no single participant could complete exploration of the environment. There were 100 victims distributed in the environment. A simpler environment was used for operator training purposes.

The computer used for MrCS was a Dell Precision T5500n with a dual six core Intel® Xeon® 3.33GHz processor and with 12GB DDR3 RDIMM memory; the other computer used for UT3 and the image server was a Dell XPS 730x with an Intel® Core (TM) i7 2.67 GHz processor and with 6GB memory. The displays were both 24-inch LCD monitors with a screen resolution of 1920 by 1200. Participants interacted with the software using a standard mouse and keyboard.

5.2 BEHAVIOR ANALYSIS USER STUDIES

5.2.1 Situation Awareness Survey Question Design

As described in Chapter 2, the multiple robot control task places substantial demand on the operator to maintain situation awareness. Endsley (1995) presents three levels of SA: perception of the environment, comprehension of the current situation, and projection into the future, to characterize the progressive deepening of awareness. For our multiple moving target searching and tracking tasks, these levels map roughly in the following fashion:

- Level 1: the display scanning perceptual search task, in which the operator seeks to identify the presence/absence of targets
- Level 2: the target marking task, in which the operator must find matches between image and map views to determine the orientation of the camera and victim location
- Level 3: in our experiment, foraging occurs in a static environment, but with moving targets. Projection to future states is primarily relevant to the target tracking, in which operators plan search strategy and specify interested areas to explore.

Convergent evidence from out-of-the-loop effects in interaction with automation (Endsley & Kiris, 1995, Endsley, 1996) and learning of survey knowledge in virtual environments (Peruch, Vercher, & Gauthier, 1995, Carassa, Geminiani, Morganti, & Varotto, 2002, Wallet, Sauzéon, Rodrigues, & N'Kaoua, 2009) suggest that lack of active involvement in navigation may lead to a loss of SA. In the multi-robot task, lack of SA is most likely manifested in difficulties with the target marking task. Performance of this task will be affected if the operator has difficulty in recognizing the target associated with a camera view, cannot determine its heading or camera orientation of image, or has difficulty associating camera landmarks with features of the laser generated map (Lewis et al., 2010).

Based on the system assistant multiple-moving-target-trail analysis tasks, SA questions were created based on possible questions that users with different roles in the team (such as team leader, operator, or observer) might ask, and were essentially derived from combining major factors of information needs extracted from different scenarios that relate to navigation, exploration, observation, and tracking. All the types of SA questions are listed in Table 1, where the SA questions were divided into six categories:

Table 1. Task type and example SA questions

Task Type	**Existing Knowledge**	**Example SA Questions**
Time Point	Location and Activity	Approximately when did target A enter/exit room X ? Approximately when was room Y fully explored?
Time Period		Approximately how long did target A stay in room Y?
Location	Target/Activity and maybe time	Where is the yellow target located at 5'30? Where are the locations that are on fire?
Regions		Please generally describe the route of the red target on the map. Are there any intersections that have been blocked? If yes, please mark it on the map.
Status	Time, Location, and maybe Target/Activity	What was the status of the green target after 10 feet?
Progress		What progress was made in exploring the upper left part of the map in the first 5 min?

5.2.2 Participants

Thirty paid participants were recruited from the University of Pittsburgh community, balanced between genders. No prior experience with robot control was required and a frequent computer user was preferred. The participants took a pre-experiment survey regarding their familiarity with computers and video games. Each participant was offered $15 per hour to compensate for their time.

5.2.3 Procedure

An Ishihara color vision test (with ten test plates) was administered via paper-based presentation. The user interface of the Event-Lens system employs several colors to represent the robot and moving target identification, as well as to track marking of a moving target. To effectively interact with the Event-Lens system, normal color vision was required. After providing demographic data, participants read standard instructions on general information about the Event-Lens system. The paper-based Spatial Orientation Test (Hegarty & Waller, 2004) was used to assess participants' spatial ability (SpA), especially regarding perspective. An on-line questionnaire of the Attentional Control scale (Derryberry & Reed, 2002) was used to evaluate participants' perceived attentional control. There are 20 items in the attentional control survey, in which attention focus and shifting are measured. The measure of attentional control is shown to have a good internal reliability ($\alpha = .88$). The study had two sections: a training section with three SA questions, and a real section with six SA questions. In the training section, participants followed instructions to use specified functions of the Event-Lens system for each question. For example, to complete the second question, we asked participants to select a specific region that contained a certain target and to view the images in the sub-queue. By doing this, participants could familiarize themselves with all of the functions of the Event-Lens system. In the real task section, participants could use whichever parts of the Event-Lens system seemed most helpful to complete their tasks. We did not require participants to give us a single correct answer or to use certain functions of the Event-Lens system, because we hoped they would explore the GUI and tell us their opinions of the Event-Lens system. By doing this, we hoped to learn more about users' spatial and temporal search strategies and to explore possible requirements and design guidelines for the future. In other words, all the tasks and SA questions were simply possible directions that could guide exploration through the Event-Lens interface and the environment. After the task, the participants were asked to complete workload ratings on the NASA task load index questionnaire (NASA–TLX; Hart & Staveland, 1988) to evaluate their perceived workload.

Additionally, a pure autonomic control test was run to assess the impact of human assistance on system performance. The autonomous path planner was run ten times to estimate the extent of the region that was likely to be explored without human assistance.

5.2.4 Results

Moving Target Marking Performance

Measures that relate to operator performance, such as target marking numbers and error marking numbers, have been logged and calculated. In addition, the individual differences, such as spatial ability scores, perceived attentional control scores, and perceived workload have been surveyed.

In terms of target marking performance, the analysis revealed that participants with higher SpA (those with higher composite scores of spatial tests) detected significantly more targets than did those with lower SpA, $F(1, 28) = 5.454$, $p = .027$. At the same time, participants with higher SpA made significantly fewer errors when marking moving targets than participants with lower SpA, $F(1, 28) = 12.607$, $p = .002$.

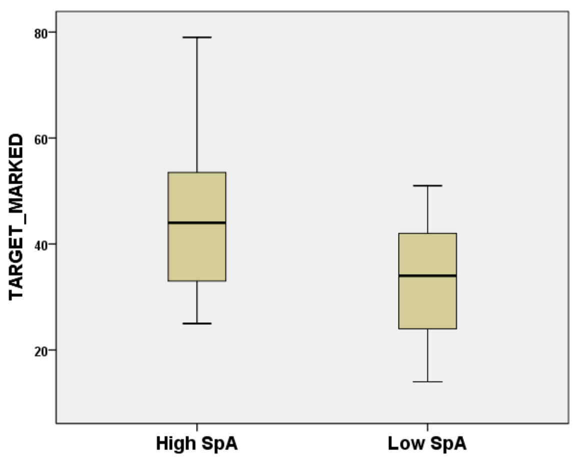


Figure 41. Target marking performance and effects of operator's spatial ability (SpA)

Figure 41 shows the box-and-whisker Plots for each condition group. The median (Q2) of the data set is the middle (horizontal) line of the box. The lower quartile value (Q1) is at the lower horizontal line; the upper quartile value (Q3) is at the upper horizontal line. The lowest datum is still within 1.5 interquartile range (IQR) of the lower quartile, and the highest datum is still within 1.5 IQR of the upper quartile. Outliers may be plotted as individual points (Massart et al., 2005).

In addition, participants' perceived attentional control (PAC) score was significantly correlated with the number of correct target marking: $r = .369$, $p = .045$.

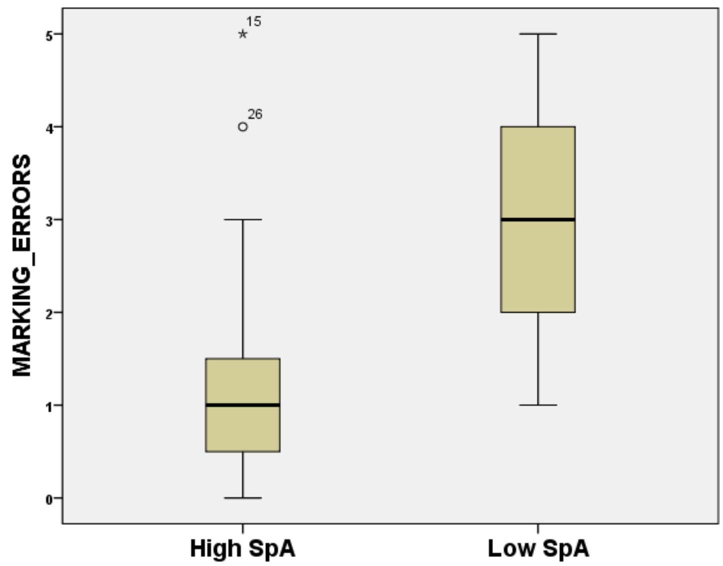


Figure 42. Target marking errors and effects of operator's spatial ability (SpA)

Situation Awareness

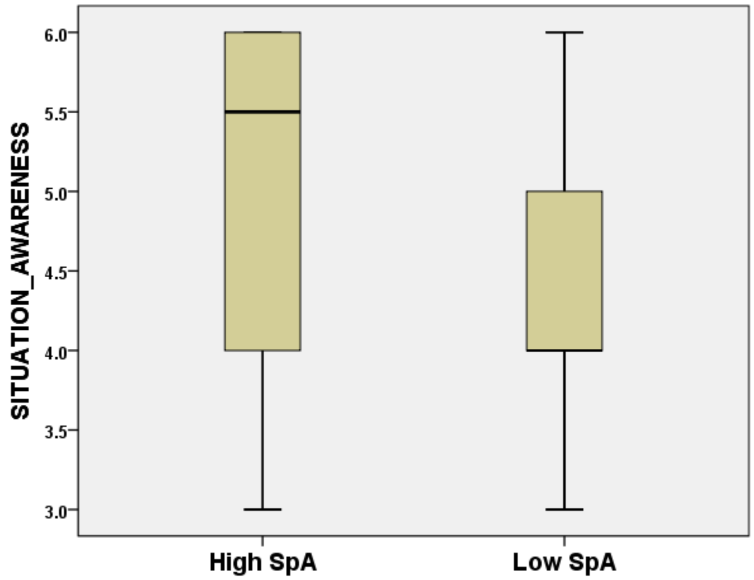


Figure 43. Situation awareness queries and effects of operator's spatial ability (SpA)

The analysis revealed that the SA score was significantly better for the participants with higher SpA compared to participants with lower SpA, $F(1, 28) = 6.642$, $p = .016$. Frequent video game players had significantly better SA than infrequent gamers: $F(1, 28) = 5.621$, $p = .025$.

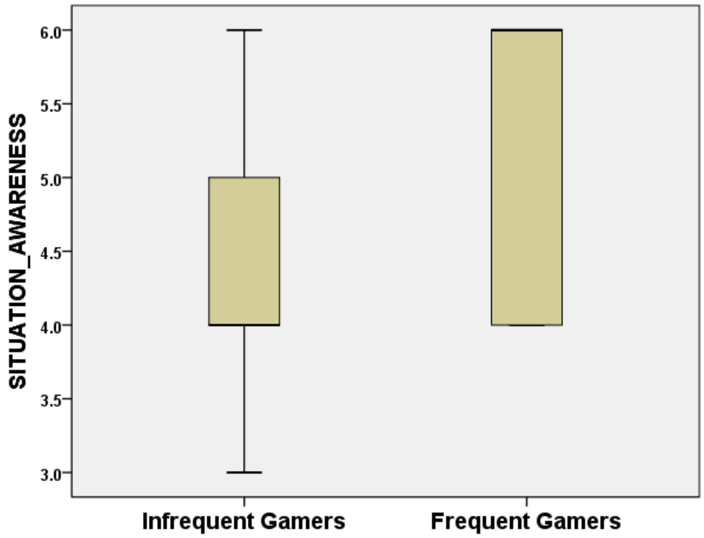


Figure 44. Situation awareness queries and effects of operator's video game experience

The analysis also revealed that the participants' SA score was significantly correlated with the perceived attentional control (PAC) score: $r = .502$, $p = .005$.

Perceived workload

Participants with higher PAC rated their workload as significantly lower than did those with lower PAC, $F(1, 28) = 4.774$, $p = .037$; however, the difference of workload assessments failed to reach statistical significance when correlated with SpA and video game experience.

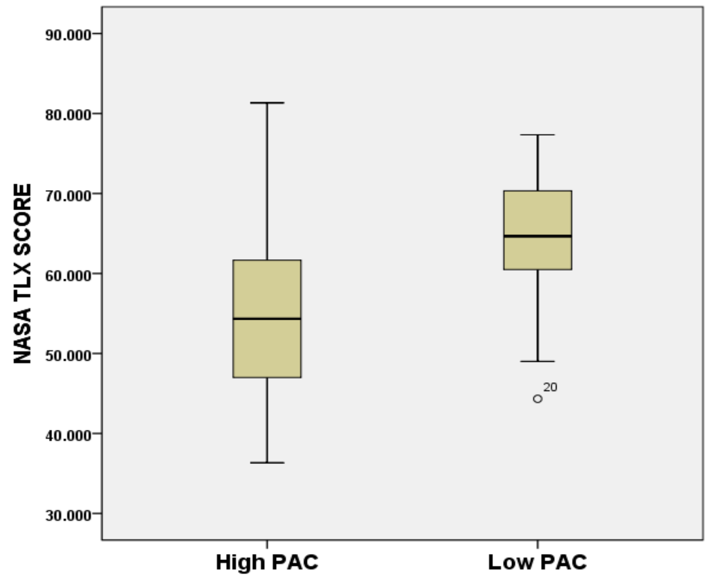


Figure 45. Perceived workload and effects of perceived attentional control (PAC)

Operator's interaction with the Event-Lens system

Participants' interaction with the Event-Lens interface have been logged and calculated, and information such as the number of images viewed and average time spent viewing each image was analyzed.

Participants' SpA was significantly correlated with the average time spent on each image (the time spent to detect a target and mark it on the map): r = -.496, p = .005. Additionally, participants' PAC was significantly correlated with the target marking number per image (to visually search and select image with a target rather than select no target image): r =.393, p = .032.

5.2.5 Discussion

Participants completed all tasks and were able to answer most of the SA questions. Our analysis focuses mainly on usage strategies for searching and tracking moving targets, operators' interaction with the Event-Lens system, and any effects of individual difference. Therefore, we begin with describing categories of user strategies that emerged from our analysis, and then focus on evaluation performance outcomes and individual difference issues from our study.

5.2.5.1 Strategy of Searching and Tracking Moving Target

For a spatial-temporal object, such as a moving target, time, location, and key activity are the three most crucial factors. Time links to targets' temporal positions, location relates to coordinates in maps, and key activities are represented as important specifications, such as a target's status change, which could link to both time and location. Our Event-Lens interface focus on these factors when searching for relevant moving target images. In most cases, participants already know one or two factors when they are following our issued tasks. As a result, they search for images that could indicate or verify the rest. For example, when a participant is trying to find images which indicate a robot has finished exploring a room, she or he already knows the location information, and thus focuses on images that could verify the activities and time, which is the image when the robot left the room. This logic leads to the development of three categories of strategies, which are illustrated in Table 2.

Table 2. Categories of Moving Target Search and Tracking Strategies

Task Type	Existing Knowledge	Example Tasks and Questions
Time Point/ Time Period	Location and Activity	Step 1: Specify a region, generating a sub-queue Step 2: Navigate through the image within this region Step 3: Apply own knowledge and scan through the images Special Case: Examine the timeline only
Location/ Regions	Target Activity and maybe Time	Case A: General location is known (e.g. The whole interested area) Step 1: Specify a region, generating a sub-queue Step 2: Navigate through the time by examining image content.
		Case B: Location information is not known Step 1: Navigate through time within the entire image set Step 2: Examine photo content by scanning content
Status/ Progress	Time, Location, and maybe Target/Activity	Step 1: Select desired time and/or location Step 2: Examine photo content by scanning

Task Type 1: Given location and activity, what is required for a specific time point/period?

Related questions usually ask "when" or "how long" did an activity occur in an interested area (Table 1), and a concrete answer was usually expected, such as a time stamp or a time range. Participants used relatively consistent strategies that included two steps (see Table 2): (1) Specify a region by dragging and releasing a box, generating a sub-queue for the area being examined; (2) Directly navigate through the images within this region by examining the contents of photos. Most participants scanned through thumbnails of the sub-queue window and occasionally viewed the image on the map in the covered region that was interesting to them. A few participants scanned through all images shown on the map using the mouse's scroll wheels. Organizing sub-queue images by time and providing only spatial and temporal unique images were critical to enabling this workflow. Most participants also used the timeline toolbar to perform temporal navigation throughout the sub-queue. The larger the size of the sub-queue, the more likely it was that the participants would use the timeline toolbar.

A couple of users also applied SA to help search the image. For example, some participants "guessed" the possible time range and then adjusted the timeline toolbar to reduce the number of images. Later explanation indicates that these participants noticed that a certain activity was not supposed to occur before some pre-requisite activities. This case did not happen very often in our study, probably because most of our users were not directly involved in robot team exploration. However, we speculate that users who are involved in this type of exploration would be more likely to use their SA to filter images.

Task Type 2: Given time and activity, what is required for the location?

Strategies used to identify moving target locations were highly dependent on the activity or item that was involved (see Table 2). If participants knew an approximate location, they scanned through the sub-queue for that location, and then may have specified a smaller area to create a more detailed sub-queue to scan through. When the general location was unknown or if there were multiple locations, participants usually brought up the entire timeline toolbar and narrowed it down using 15-second predefined sections, then scanned through the top 10 images. In this type of task, the time was not always available. Most of our participants were not immersed in the robot searching and explore process, so they rarely had an idea of precisely when the activities took place. On the other hand, if the time point or time period was known, our participants successfully adjusted the time range by setting up a beginning and ending point in the timeline toolbar before examining images.

Task Type 3: Given time and location, what is required for activity and status?

The inquiry of moving target status change and progress is an extremely important task. Due to the continuity of the mobile progress, it is impossible and unnecessary to mark the entire track of moving target. As a result, the status change would be the progress check points for the moving target, and could be a good display moving process.

Two steps were included in the strategies (see Table 2). Some participants first narrowed down the set of images by selecting the designated time range with the timeline toolbar, then focused on map to specify the locations, while other participants first specified the locations by creating a sub-queue for the interested area and then tried to narrow down the time range. When time and location were narrowed down, both groups of participants then examined contents in the images with visual scanning. The majority of participants scanned through most images in the thumbnails list, while other users directly viewed the image on the map with timeline toolbar or mouse scroll wheels. Since the images were ordered by time and were preselected by the utility functions by only showing the spatial and temporal unique images, it was possible to determine status/progress by viewing a small number of photos at critical positions, which is a powerful shortcut. This also suggests the critical importance of a time-based order of images.

After a successful marking, the image as well as the laser scan and time stamp information will be permanently correlated with this mark. In addition, all marking for certain moving targets with the same color will be connected by an arrow of the same color to show the direction in which the target has moved. By showing the target moving sequence rather than the order of marking, our Event-Lens system allows viewers to have an overview of the Multiple Moving Targets (MMT) status changes or progresses with a quick glance of the map.

The strategies used within each task category were relatively consistent, partly because the order of task steps was imposed by the asynchronous MrCS. The Event-Lens interface provided clear hints to lead operators through the interaction workflow, which seemed natural and intuitive to participants. Among all strategies that engineers used, the timeline toolbar was used most frequently. The majority of participants (93.3%) frequently started with the predefined segment of 15-second intervals to further narrow down the size of an image set. The use of sub-queues and images on the map were dependent on known information, such as locations and status changes. It could infer that the mental models of most people for moving targets is time-based, and thus would cause them to start their searching progress through the temporal dimension.

5.2.5.2 Impact of Individual Differences

According to the performance result, those participants with higher SpA marked more targets and made significantly fewer errors when marking moving targets than did those with lower SpA. These results are consistent with previous findings (Chen et al., 2012) that "individuals with higher SpA tend to exhibit more effective scanning performance and, therefore, are able to detect more targets than do those with lower SpA, especially when visual processing load is heavy".

Our Event-Lens asynchronous interaction procedure required the most visual monitoring by the participants. Detailed action analysis shows that high-SpA participants were able to view 17.75% more images than their lower-SpA counterparts, and spent significantly less time (2 fewer second) on each image viewed.

In terms of awareness of the overall task environments, participants' SA scores were significantly better with high-SpA participants than their lower-SpA counterparts. Baldwin & Reagan (2009) explained that individuals with a good sense of direction exhibited more route-learning disruption from the tapping task, suggesting a greater reliance on visuospatial working memory resources. Although eye movement and eye gaze data were not collected during this study, we can still infer that the high-SpA participants scanned the map more frequently during the task and were immersed more deeply with the robot exploration.

Similar effects on SA was also found for game experience difference, as the evidence shows that that frequent video gamers tended to have somewhat better SA than did infrequent gamers. As Green and Bavelier (2006) found that frequent gamers tend to have better visual short-term memory, it is not surprising to find them exhibiting better SA of MMT searching and marking tasks. Additionally, Cummings, Clare, and Hart (2010) claim that frequent gamers have higher degree of consent and collaborate more with systems than infrequent gamers. Their finding was confirmed by our performance and operation results, as frequent video gamers worked better with the Event-Lens system on every dimension than did the infrequent gamers. During the task, frequent video gamers issued 17.33% more temporal navigation actions (specified a time period) and 44.4% more spatial navigations (selected an area to create a sub-queue), viewed 31.9% more images than the infrequent gamers, which resulted in 25.4% more corrected markings and 21.9% more markings per image viewed than the infrequent gamers. This more effective and efficient collaboration with the Event-Lens system has contributed to the frequent gamers' better SA. Our performance and operation results, which were impacted by the participants' video gaming experience, also support the conclusion of an U.S. Air Force study (Triplett, 2008). According to the research interviews of UAS pilots, "gamers' superior visual information-processing skills may be able to translate into superior robotics management performance."

The result also revealed that the participants' SA score was significantly correlated with the perceived attentional control (PAC) score. From this detailed operation data, we can infer some of the attentional processes and strategies of the participants. In their effort to maintain optimal performance across the tasks, the low-PAC participants appeared to allocate more attentional resources than to the high-PAC participants, as the elevated actions (15% more images viewed) suggested. However, this biased attention allocation came with a price, as the degraded target detection performance in the lower-PAC group had 9.8% fewer targets marked per image viewed. Additionally, the result of the after-task NASA-TLX workload survey, which indicated that participants with low PAC experienced significantly higher workload compared with the participants with high PAC, confirmed our deduction.

Participants' PAC was found to have a significant effect on their MMT marking performance. This finding is consistent with those of Chen et al. (2012) that participants' PAC has a significant effect on their multitasking performance, as participants with higher PAC were more able to allocate their attentional resources effectively in the multitasking environment than those with lower PAC, "especially when tasking environments became more challenging."

5.3 INTERACTION MODE EXPERIMENT

5.3.1 Hypothesis

In recent experiments (Wang et al., 2011), we have found that the image queue method alleviates the need for excessive switching, and that it reduces errors and operators' workload, as compared with the traditional synchronous display. However, the synchronous method has an inherent advantage for interacting with the dynamic environment, and it has been the most accepted way of handling this type of service for years. For moving data, the synchronous interaction method is intuitive, clear, and easy to understand; as it is "what you see is what you get (WYSIWYG)." It is deterministic in its behavior, due to the fact that all signals are sampled at a well-defined time interval. Synchronous designs rely on few timing parameters to guarantee operation, while the asynchronous method may involve process time and delay. Meeting these advantages of the synchronous method will ensure that our asynchronous design is more compatible with dynamic environments.

As Section 2.1 describes, we want to solve the multiple feeds problem for multidimensional data by using the Event-Lens system's asynchronous method, which enables user to assess the situation at any spatial and temporal scale and to manipulate the multidimensional spatiotemporal data. Thus, our hypotheses are:

- H1: The Event-Lens system will lead to improved performance of the multiple moving targets marking task, as compared to conventional streaming video methods.

- H2: The Event-Lens system is easier for users to abstract and understand information availability on the time dimension, and will lead to improvement of analyzing and reasoning for the trajectories of multiple moving targets.
- H3: By providing spatial and temporal unique information with top utilities, the Event-Lens system will reduce errors and operator workload, as compared to the traditional synchronous display.

5.3.2 Experiment Design

In the interaction mode experiment, the effects of display methods and interaction procedure on operator performance were investigated. The participants' task, as in the previous behavior analysis study, was to perform a supervisory control task in which 12 robots navigated autonomously with the assistance of MrCS user interface while searching for moving targets. The display methods and interaction procedure was manipulated to be either traditional synchronous display with live streaming video feeds from the robots (streaming-video condition), or asynchronous display with the Event-Lens MrCS user interface (Event-Lens condition).

Furthermore, the experiment simulated an unidentifiable moving target task, compared with the identifiable moving target task as in previous behavior analysis study. For the unidentifiable moving target, all the targets had an identical appearance, even down to the same color. As a result, in addition to the target detection and marking tasks, the participants had to simultaneously perform a target identification reasoning task.

The experiment followed a three-condition, repeated measures design. In conditions 1 and 2, the users carried out an identifiable moving target (IDT) task that compared the conventional traditional synchronous display (streaming video) with MrCS augmented by the experimental Event-Lens asynchronous display. In condition 3, the users were asked to fulfill the unidentifiable moving target (UIDT) task with the Event-Lens asynchronous display. All three conditions are counterbalanced with the task sequence, entrance points for robots, and moving target route (Table 3). In all conditions, automated path planning to improve search performance and autonomous exploration was used. Because the laser map was built up slowly as the environment was explored and the office-like environment provided few distinctive landmarks, there was little opportunity for participants to benefit from prior exposure to the environment from a different entrance location.

Table 3. Experiment Design

Sample Size	Experiment Conditions		
10 participants	Streaming Video with IDT Task	Event-Lens with IDT Task	Event-Lens with UIDT Task
10 participants	Event-Lens with IDT Task	Event-Lens with UIDT Task	Streaming Video with IDT Task
10 participants	Event-Lens with UIDT Task	Streaming Video with IDT Task	Event-Lens with IDT Task

5.3.3 Participants

30 paid participants (15 male and 15 female) were recruited from the University of Pittsburgh community and assigned proportionately to experimental conditions. No prior experience with

robot control was required and a frequent computer user was preferred. The participants took a pre-experiment survey regarding their familiarity with computers and video games. Each participant was offered $15 per hour to compensate for their time.

5.3.4 Procedure

Participants were randomly assigned to the different condition sequences (with 10 participants per group, as shown in Table 3) before their sessions started. After providing demographic data and completing a perspective-taking test, participants read standard instructions on how to control robots with MrCS. Each condition session followed the same procedure and lasted for approximately 30 minutes.

The interaction methods with relevant MrCS and the type of moving targets were matched in the training and experimental scenarios. In each subsequent training session, participants practiced control operations for the same MrCS and tasks as the following real condition for 10 minutes each. Participants were encouraged to find and mark at least two targets in the training environment under the guidance of the experimenter. After the training session, the real-task experimental session began immediately and lasted about 15 minutes.

The environment was 5026 m^2, a size sufficient to guarantee that no participant could complete exploration. There were five moving targets distributed in the environment that were both with and without different colors. During the scenarios, participants used their 12 robots to track the 5 moving targets.

A 5-min break was given between the experimental scenarios. Participants assessed their workload using an electronic NASA-TLX workload survey (Hart & Staveland 1998) immediately after each condition. Following completion of all conditions, participants were asked to evaluate the usability of all interaction methods by filling out the post-experiment interview.

5.3.5 Performance Metrics

Performance is a complex concept to assess. Part of the performance variable could be physical, such as precision, errors, and reaction time; other aspects are more mental, such as situation awareness or workload. Given the demanding task of marking multiple moving targets in a dynamic environment, it is not surprising that requesting additional operator responses will have a negative impact. At the same time, completing the task with less time improved user efficiency, but it also increased workload and task error rate (Hendy et al., 1997; Mosier et al., 2007).

Through the experiment, task performance variables will be coded in the following categories: 1) the number of moving targets located and identified, the number of moving target routes edited; 2) task errors (incorrect markings); 3) precision (marking distance with the ground truth location); 4) operator's efficiency (time of task completion when user identified a target to successfully mark it); and 5) the perceived workload (NASA-TLX scores). In addition, participants' composite SpA test scores and their Attentional Control survey scores were used as covariates.

5.3.6 Results

Moving Target Marking Performance

Similar to the behavior analysis user study, measures relating to operator performance and operation have been logged and calculated. In addition, the individual differences, such as spatial ability scores (SpA), perceived attentional control scores (PAC), and perceived workload have been surveyed.

In terms of the operator's moving target marking task performance, the analysis showed that both interaction methods and target identifiability significantly affected the moving target marking numbers. The analysis revealed that participants with the Event-Lens condition marked significantly more identifiable targets (IDT) than did those with the streaming video condition, $F(1, 28) = 11.15$, $p = .003$. At the same time, with the Event-Lens interface, participants marked significantly more identifiable targets (IDT) than unidentifiable targets (UIDT), $F(1, 28) = 16.87$, $p < .001$. (Figure 46)

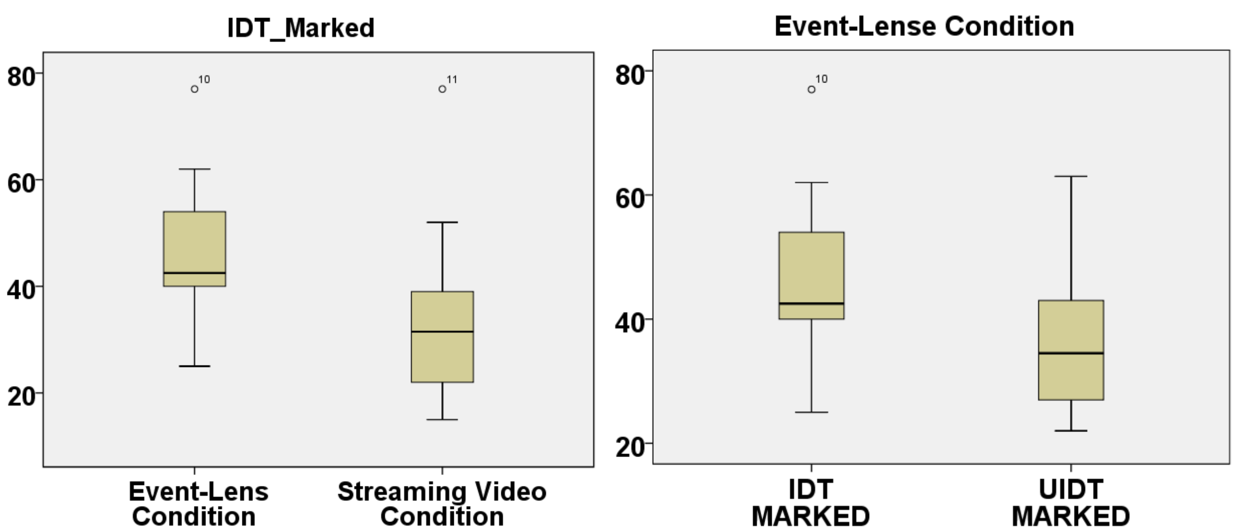


Figure 46. Effects of interaction methods (left) and target identifiability (right) on target marking performance

Further analysis of markings with ground truth (moving target track) revealed that participants with the IDT task made significantly fewer errors when marking moving targets than the participants with the UIDT task, $F(1, 28) = 16.92$, $p < .001$. However, when reducing the wrong identification errors (marking as a wrong color) from the total numbers, the error making rate of the two conditions remain at the same level (Table 4). In addition, the effect for interaction methods was not found to be significant, as regarding the marking errors.

Table 4. Target marking errors and effects of target identifiability

Variables	IDT Task $\bar{x}$	UIDT Task $\bar{x}$	$F_{1,28}$	P
Total Errors	2.05	4.64	16.92	<.001
Errors without wrong identification	2.05	2.18	.380	.544

Perceived workload

The analysis showed that both interaction methods and the target's identifiability contributed significantly to the participants' perceived workload, with $F(1,28) = 24.48$, $p < .001$ and $F(1, 28) = 12.43$, $p = .002$, respectively (Figure 47). Participants experienced higher workload in the streaming video condition, as well as when the moving target was unidentifiable.

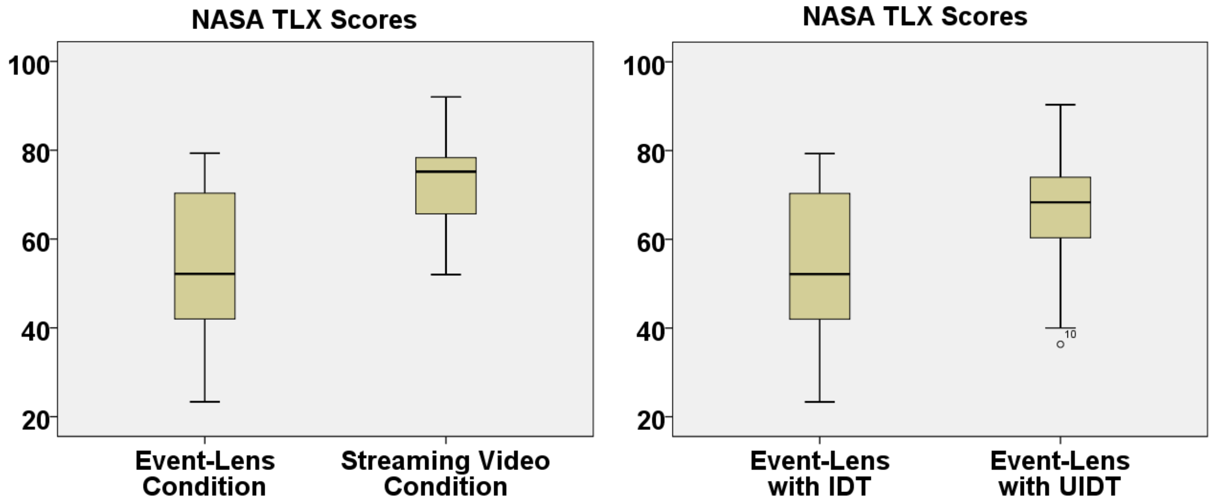


Figure 47. Effects of interaction methods (left) and target identifiability (right) on perceived workload

Operator's interaction with the MrCS

Participants' interaction with the MrCS during each condition have been logged and calculated, The number of images viewed, the average target marking time, and average target marking number per image viewed were analyzed.

Participants used significantly less time for marking each target in the Event-lens condition than in the streaming video condition, $F(1, 28) = 51.69$, $p < .001$. There was a significant difference between the IDT task and UIDT task for the average target marking time, $F(1, 28) = 4.699$, $p = .039$, which favors the IDT task, as it had less time spent on it.

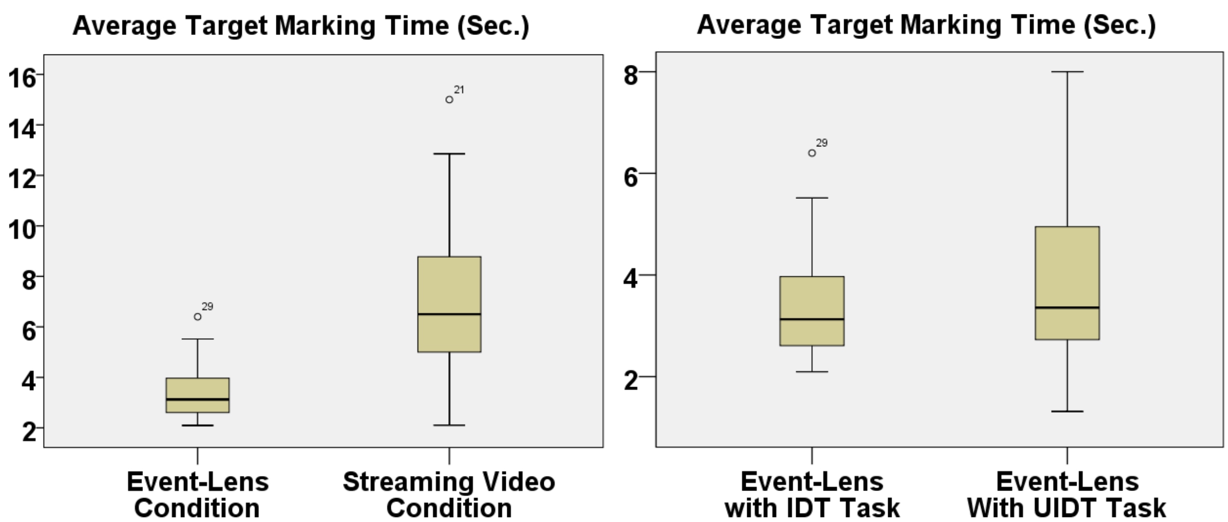


Figure 48. Effects of interaction methods (left) and target identifiability (right) on average marking time

When comparing the average number of images which were selected and viewed between the IDT task ($M = 153.3$, $SD = 44.21$) and UIDT task ($M = 149.9$, $SD = 67.96$), there is no significant difference. However, participants with the IDT task marked significantly more targets per image as compared to those with the UIDT task, $F(1, 28) = 7.52$, $p = .010$.

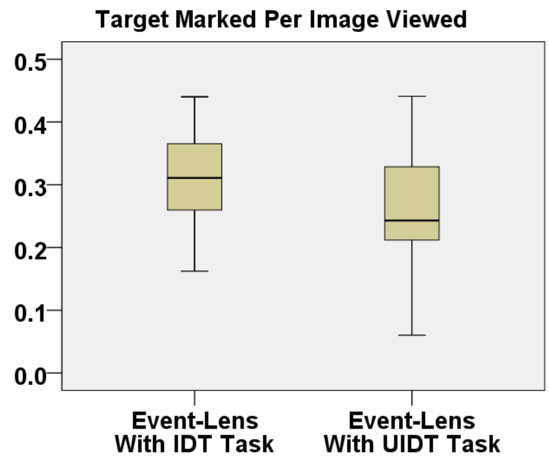


Figure 49. Effects of target identifiability on number of target marked per image viewed

5.3.7 Discussion

5.3.7.1 Synchronous Control vs. Asynchronous Control

One purpose of this interaction mode experiment was to examine the impact of our Event-Lens display on overall performance for searching and tracking multiple moving targets. The Event-Lens system presents information to subjects asynchronously, but ordered by a quality metric that relates to the spatial and temporal utility of the information. This stands in contrast to the conventional synchronous video stream display that presents information as it becomes available. The performance advantage of the multiple moving targets marking task for the Event-Lens condition has confirmed our hypothesis. Additionally, participants in the Event-Lens condition made 39.8% more marks for each MMT route, as compared with the conventional streaming video condition. The advantage of the Event-Lens system in analyzing and reasoning the multiple moving targets trajectories indicates that users of the Event-Lens system find it easier to abstract and understand information availability on the time dimension, which also proves our hypothesis.

One of the main design objectives for the Event-Lens interface is to create effective abstract and intuitive spatiotemporal representations that help users maintain awareness of key information and actions. To be more precise, measures of participants' behavior linked to target detection were annotated and analyzed in the data log. The significant difference of average marking time confirmed the advantage of the asynchronous interaction methods and interface design. The diagram in Figure 50 compares the interaction procedures of participants with the Event-Lens and streaming video conditions. Once a target entered into a camera's field of view (video or image) and was potentially detectable, a series of actions were performed to develop sufficient situation awareness (SA) to perform the target marking task.

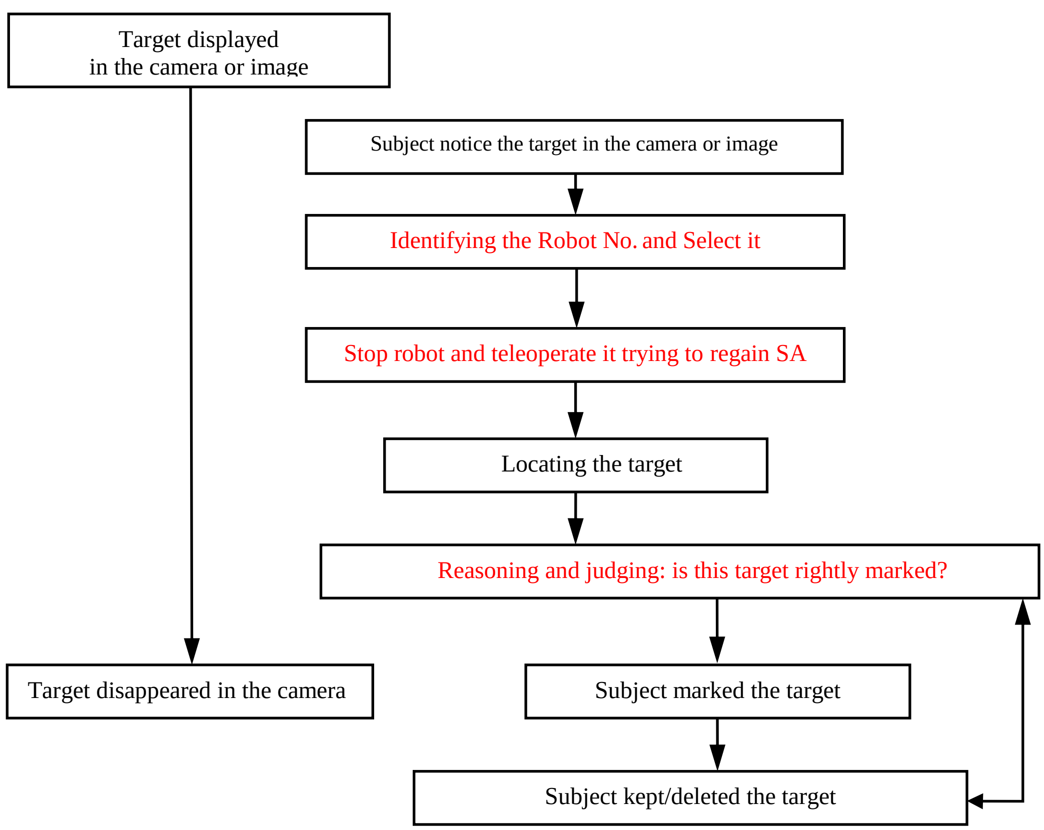


Figure 50. Target Marking Procedure Analysis for Operator with streaming video and Event-Lens conditions

(All actions in red color are only for streaming video conditions)

In the streaming video condition, the operator first needs to identify the robot and regain SA of the robot by matching the robot's color and numerical label on the map. Next, the operator has to determine the orientation of the robot and match landmarks between camera and map views. In order to clarify the relationship between the robot and target, the operator may choose to teleoperate the selected robot to help locate it on the map and determine its orientation through observing the direction of movement. The operator must then locate the target on the map corresponding to the camera view.

Clearly, for the Event-Lens system, the actions such as stopping the robot, viewing the map, and locating the robot have been eliminated, and thus the time required for an operator to regain SA has been shortened. Furthermore, to gain enough information, the operator in the streaming video condition may choose to teleoperate the selected robot to locate it on the map and determine its orientation either through observing the direction of movement or simply to get a better viewing angle. However, for the Event-Lens condition, as it provides information of visual coverage, camera direction of image captured, and marking history shown on the map, an operator could easily figure out where the image was taken and determine if the target has been marked before. Additionally, with the temporal navigation function, it is much easier for operators to find a better view image of a given target.

Although navigation tasks benefits from immersion and situation awareness (Wang et al., 2011), Event-Lens participants have no need to teleoperate a robot in contrast to streaming video participants when they encounter a moving target in the video feed, which may make the Event-Lens participant less immersive. However, Wang et al. (2011) conclude that the added situation awareness from teleoperating robots while searching for victims provides little to no benefit for such a system with many robots. This performance result confirmed this conclusion.

Most importantly, this asynchronous display method decoupled the navigation, exploration, and error recovery tasks from the target detection tasks, allowing the latter to be completed entirely asynchronously without any penalties for performance in terms of the number of marking numbers (Wang et al., 2011). Even more so, by decoupling these tasks, we reduced the number of errors that occur for marking moving targets, while at the same time reducing overall workload, which verified our hypothesis.

Above all, when people are monitoring multiple sensor systems, normally they are confronted with a bank of videos and other information displays, much like a security guard monitoring many surveillance cameras. Continued monitoring of video feeds synchronously during system operation puts a greater effort on operators (Wang, et al., 2011). Not only does it require continuous attention, but participants switch between tasks more often (whenever a target appears in a video feed). This leads to an impression of time pressure that may contribute to the increased number of errors. The significant difference of individual dimensions of workload, temporal demand, and effort all support this observation.

5.3.7.2 Identification of the Moving Target

Marking unidentified multiple moving targets is the exploration part of the experiment, which is totally unthinkable when working in the conventional streaming video condition. This is also the reason we didn’t apply the 2*2 repeated measure experiment design for our interaction mode experiment. By providing the target trajectory analysis functions to assist users with marking multiple moving targets, the Event-Lens system makes it possible to deal with unidentified MMT.

When participants detected a UID target in an image, they could click the image to see the visual coverage of the selected image. By comparing the timestamp of the image and other target marking histories (different color trajectory on the map), the user could select a color from the waiting list to give the UID target a customized identification.

However, the significant difference of performance results shows that marking multiple moving targets is still a great challenge for operators, as they need to build more immersive SA, allocate their attention more properly, and carry out many difficult reasoning procedures. Thus, the significant difference of the marking errors is mainly caused by the wrong identification, which consisted of nearly 50% of the total error numbers. A further analysis result indicates that, beside wrong identification, other error making rates of the two conditions remain at the same level.

These result shows that participants with higher PAC make significantly fewer errors during the UID condition than those with lower PAC, which suggest that participants with higher PAC were more able to allocate their attentional resources effectively when tasking became more challenging with the multiple UID moving targets. It was also found that participants with higher PAC consistently performed better in the UIDT condition across different parameters than those with lower PAC, such as 16.62% more corrected markings and 18.2% more images viewed.

However, when two or more UID targets were shown in the same image, or images of UID targets that were located close to each other, it is technically extremely hard for user to individually clarify the identities of MMT with asynchronous interaction methods. For instance, the case of two identical color and shape UID targets that have passed each other, and the case of two UID targets that met each other and turned back, will provide the same set of image results. The only way to clarify this difference is to make a record of the whole progress available to the operator in as much detail as possible, which depends on the information availability of the surveillance sensor network.

5.4 GENERAL DISCUSSION

Time, location, and activities are relatively closely connected in multiple moving target (MMT) tracking and visualization. Bringing these three crucial factors together into a single interactive tool enables users to more efficiently locate certain target images, which will better support finding the status change of targets. Participants were observed to have fairly clear logic using the timeline (spatial navigation), sub-queue (temporal navigation), and target trajectory analysis functions. Operators who are tracking MMTs often need to browse images in a specific spatial region and manipulate the time range at the same time. The Event-Lens system provides more flexibility to enable users to browse and search in various ways. Our marking history representation also provides an overview of spatiotemporal coverage of MMTs on the map, which shows the marking history through spatial and temporal dimension, enabling users to see trends in activity over time and space and to answer certain questions from the visualization alone.

In the current study, we investigated the effects of our prototype system, Event-Lens, on human operators' performance of supervising multiple robots to complete searching and tracking MMT tasks in a multitasking environment. Overall, it appears that the Event-Lens system was effective in reducing the operators' marking times in target search tasks, and significant benefits on the operators' concurrent task performance and workload were also observed. The results of the interaction mode experiment show that the Event-Lens system improved operators' performance on target detection, route editing, and errors, as compared with conventional synchronous interaction methods. More importantly, the Event-Lens system enabled users to perform more challenges, such as marking and tracking unidentified targets. Participants' self-assessed attentional control and video gaming experience was found to affect their overall performance. Across experiments, participants with higher SpA consistently outperformed those with lower SpA in tasks that required the most visual scanning joined with perspective shifting.

All the benefits of the Event-Lens system come from extending human perceptual ability by allowing operators to control their point of view spatially and temporally. Our Event-Lens system makes the information existing in the fusion of the points-of-view available for operators, and our interface helps users process these kinds of information more efficiently and effectively.

In addition, the following design guidelines and implications for the future system design were also discovered:

1. Organizing multiple sensor information based on time, space, and content provides an efficient way to search and track MMTs. Some types of tasks can be performed directly from time and spatial coverage visualizations.
2. Time-based organization of information supports visual scanning, which is the most frequently occurring action. This also reflects the major mental model of operators who participate in MMT-based tasks.
3. Providing simple navigation facilities for the users' selected information enables smooth workflow of information extraction, as even simple window-switching actions can disrupt the users' workflow.
4. Redundant data or information should be contained out of the way of the main interaction.

6.0 CONCLUSIONS

6.1 THESIS CONTRIBUTION

As stated in Section 1.2, the ultimate goal of the present research is to solve the multiple feeds problem of multiple sensor systems for dynamic environments, and to devise a way to overcome the difficulty in visualizing multidimensional data. More specifically, we aim to develop a new interaction approach that provides multiple interaction methods by which users can manipulate spatiotemporal data.

This work contains several significant contributions to human-sensor systems. First, the state-of-the-art spatiotemporal information visualization, organization, and integration have all been comprehensively reviewed. The problems of data overload and redundancies have been reduced by providing spatially and temporally unique information maximize usefulness to the end user.

Combining spatial and temporal information with multiple interaction methods for users to manipulate spatiotemporal data, such as timeline toolbars and area concerned sub-queues, provide additional facilities for operators to quickly locate MMTs with related status changes. This leads to a second contribution. By using the Event-Lens approach, we also learned about users' spatial and temporal search strategies. We also expect that this knowledge will contribute to usability requirements for future systems dealing with multiple moving targets.

A third contribution is in the practical application of our approach. We designed and developed the Event-Lens system, which generates asynchronous prioritized images and facilitates a manageable but comprehensive view of the information collected by sensor systems with multiple feeds. Additionally, significantly improved performances of participants empirically prove the viability of our approach.

The fourth contribution is to summarize guidelines for future visualization and interaction methods to support tasks with large amounts of spatiotemporal data.

6.2 FUTURE WORK

There are some limitations of our work, which also indicate the related issues that require future investigation.

First, we have tried to simplify our conditions by assuming that there is no automation failure. However, during actual human-robot interaction procedures, humans are likely to be called upon to assist with a variety of low-level robot problems, such as sensor failures or obstacles that robots cannot solve on their own. This actually adds several new types of tasks that human operators excel at, such as fixing stuck robots, solving robot traffic jams, or avoiding potentially dangerous situations. The system should gather and analyze this valuable information and provide both spatial and temporal trending information. A new robot assistance queue could be simulated by viewing information from the perspective of a troubled robot. The information should be available, such as images taken by the robot prior to becoming stuck, nearby frames taken by other robots, and frames from look-at tasks, and a carefully designed robot assistant queue function centered on the stuck robot could gather, integrate, and organize all of this relevant information, which would increase the operator's situational awareness beyond what a simple video feed could provide.

There are also areas for future work addressing the two other main assumptions made by the implemented Event-Lens approach: ground truth sensors and a static environment (geometry level). A probabilistic manner information database could handle the noisy sensor data. It is similar to the way in which mapping algorithms handle noisy pose and laser scan data using an occupancy grid. On the other hand, a similar approach could be used to naively address dynamic environments, as with a decay function used to reduce the certainty of areas of the information database occupancy grid as time elapses.

Future work may also focus on the task of estimating the numbers of moving targets as well as their distribution. To do so, other types of sensors may be introduced to the sensor system. This raises the question of what further spatial and temporal information should be presented to operators.

The overall significance of the Event-Lens system goes further than the mere reduction of errors, workload, and performance improvement—it allows for the design of a system that treats multiple moving target detection tasks as notifications for a call center, which points the way toward how to design the visualization and interaction system with large-scale, complex, multiple-dimension sensor systems. The experiments' conditions are currently simplified by introducing only a single operator with 12 homologous robots. However, our MrCS is a distributed system that can be easily extended to support multiple operators (human agents with a user interface). Moreover, the Event-Lens approach provides a beneficial way to divide tasks between operators and for operators to share information or recall situation awareness.

Once we face the problem of managing teams of 100 or more robots (or, in other words, large-scale systems), the call center design of the control architecture with embedded prioritized spatiotemporal information will be proposed to solve these problems. At that point, the information sharing problem due to the operators' loss of situational awareness will be one of the priorities of our concern. Just imagine three operators controlling more than 100 UVs with some UV-originated requests, such as verifying or marking targets, or alerting operators about low fuel and other UV needs, such as veering off path or becoming bogged down by terrain (trees in a forest, for example) that operators must monitor to detect. As a result, the Event-Lens interaction mode may be a good choice for people to share information or recall situational awareness. The Event-Lens procedure provides a beneficial way to divide tasks between operators, in which exploration and perceptual search (identifying targets) tasks could be neatly deconstructed and handled by 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, a call center approach must be proposed, in which some operators address independent control needs for monitoring and exploration of UVs, whereas other operators address independent spatiotemporal information in a queue for target marking and other perceived tasks.

APPENDIX A

[DEMOGRAPHIC QUESTIONNAIRE]

1. What's your experiment ID?
2. Gender: Female Male
3. Age:
4. Education:
    a. Currently Enrolled as Undergraduate
    b. Completed Undergraduate
    c. Some Graduate School
    d. Completed Graduate School
5. Major Field(s) of Study:
6. In a typical working DAY, how much time do you spend using a computer?
    a. <1 hour b. 1-3 hours c. 3-5 hours d. 5+ hours
7. In a typical WEEK, how much time do you spend playing video games? (Including smart-phone game apps)
    a. <1 hour b. 1-5 hours c. 5-10 hours d. 10+ hours
8. Please check all types of video games that you play regularly.

a. Strategy/Tectical b. Adventure c. Sport Simulations

d. First Person Shooter e.Arcade-style Games f.Vehicle Simulations

9. How often do you play video games?

a. Frequently b. Occasionally c. Never

10. Please select the photo sharing websites that you are using:

a. Flickr b. Pinterest c. Photobucket

d. Snapfish e. Picasa f. Others

11. How often do you use the photo sharing websites:

a. Frequently b. Occasionally c. Never

APPENDIX B

[TASK QUESTIONS FOR BEHAVIOR ANALYSIS USER STUDY]

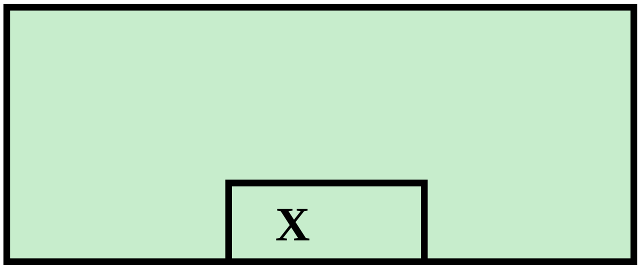
B.1 INTRODUCTIONS

- All the task questions are independent, which are not related to each other.
- When you got stuck, please read out the question number that you are working on.
- Take as long as you wish and try to get a confident answer.
- If you don't know the answer to a question, please try your best or write down "I don't know" and EXPLAIN HOW/WHY YOU GOT STUCK.

B.2 LIST OF TASK QUESTIONS

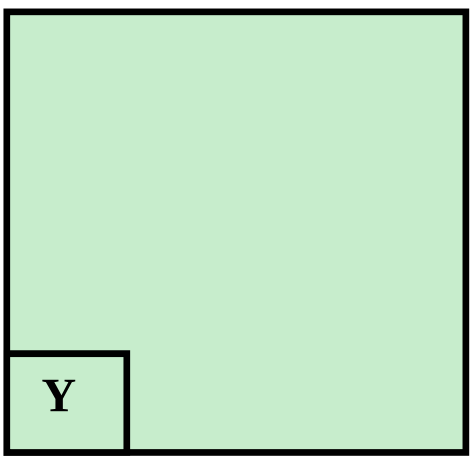
Training Task

1. Approximately when did target Green enter room X ?
2. Approximately when did room X been fully explored?
3. Approximately how long did target Green stay in room X?



Real Task:

4. Where is the yellow target located at 2’-3’?
5. Approximately when did target Green enter room Y?
6. Approximately when did room Y been fully explored?
7. Approximately how long did target Green stay in room Y?



8. What was the status of the Green target after 10’ ?
9. What progress was made for exploring the upper left part of the map in the first 5 min?

APPENDIX C

[ATTENTIONAL CONTROL SCALE]

Developed by Derryberry & Reed (2002), presented as an on-line questionnaire in this study.

Items are scored on a 4-point scale (1. almost never; 2. sometimes; 3. often; 4. always).

R: reverse-scored item.

1. It's very hard for me to concentrate on a difficult task when there are noises around. (R)
2. When I need to concentrate and solve a problem, I have trouble focusing my attention. (R)
3. When I am working hard on something, I still get distracted by events around me. (R)
4. My concentration is good even if there is music in the room around me.
5. When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.
6. When I am reading or studying, I am easily distracted if there are people talking in the same room. (R)
7. When trying to focus my attention on something, I have difficulty blocking out distracting thoughts. (R)
8. I have a hard time concentrating when I'm excited about something. (R)

9. When concentrating I ignore feelings of hunger or thirst.
10. I can quickly switch from one task to another.
11. It takes me a while to get really involved in a new task. (R)
12. It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures. (R)
13. I can become interested in a new topic very quickly when I need to.
14. It is easy for me to read or write while I'm also talking on the phone.
15. I have trouble carrying on two conversations at once. (R)
16. I have a hard time coming up with new ideas quickly. (R)
17. After being interrupted or distracted, I can easily shift my attention back to what I was doing before.
18. When a distracting thought comes to mind, it is easy for me to shift my attention away from it.
19. It is easy for me to alternate between two different tasks.
20. It is hard for me to break from one way of thinking about something and look at it from another point of view. (R)

APPENDIX D

[POST-EXPERIMENT INTERVIEW]

1. What do you think about the tasks? Were there any tasks that you found hard? If yes, why did you find it hard?
2. Have you ever got stuck? If yes, which question? How and why did you get stuck?
3. Was there any function in the interface that confused you?
4. Which function do you consider as the most useful? Explain.
5. If you want to know the overall exploration status at any point, what would you do using the interface?
6. If you are asked to find the state of X (e.g. room X) at some time period, what would you do?
7. If you were asked to find out how long it took to fully explore room Y, what would you do?
8. What do you think about the prototype in terms of viewing temporal distributions of images?
9. Please give us one or more suggestions that can help improve the interface.

BIBLIOGRAPHY

Alberts, D., Garstka, J., & Stein, F. (1999). Network-centric Warfare: Developing and leveraging information superiority. *The Command and Control Research Program.*

Anabuki, M., & Ishii, H. (2006). 3D and Sequential Representations of Spatial Relationships among Photos. *Proceedings of CHI 2006, ACM Press*, 472-477.

Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., & Wrobel, S. (2011). A Conceptual Framework and Taxonomy of Techniques for Analyzing Movement. *Journal of Visual Languages and Computing 2011, 22(3),* 213-232.

Andrienko, N. and Andrienko, G. (2011). Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics, 17(2),* 205-219

Andrienko, N., and Andrienko, G. (2007). Designing visual analytics methods for massive collections of movement data, *Cartographica, 42(2),* 117-138

Aretz, A. J. (1991). The Design of Electronic Map Displays. *Human Factors*, 33, pp. 85-101.

Astle, D. E., & Scerif, G. (2009). Using Developmental Cognitive Neuroscience to Study Behavioral and Attentional Control. *Developmental Psychobiology,* 51(2): 107-118

Aubert, T., Corjon, J., Gautreault, F., Laurent, M., Causse, M., & Dehais, F. (2010). Improving situation awareness of a single human operator interacting with multiple unmanned vehicles: first results. *Proceedings of the HUMOUS 2010 conference - HUMans Operating Unmanned Systems*.

Bak, P.(2009). Spatiotemporal analysis of sensor logs using growth ring maps. *IEEE TVCG, 15(6),* 913–920.

Balakirsky, S., Carpin, S., Kleiner, A., Lewis, M., Visser, A., Wang, J., & Zipara, V. (2007). Toward hetereogeneous robot teams for disaster mitigation: Results and performance metrics from RoboCup Rescue. *Journal of Field Robotics, 24(11-12),* 943-967.

Baldwin, C. & Reagan, I. (2009). Individual differences in route-learning strategy and associated working memory resources. *Human Factors,* 51, 368-377.

Becker, R., Cleveland, W., & Wilks, A. (1987). Dynamic graphics for data analysis. *Statistical Science, 2(4),* 355-395.

Bertin, J. (1967). *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison, 1983.

Birch, J. (1997) Efficiency of the Ishihara test for identifying red-green colour deficiency. *Ophthalmic Physiol Opt*, 17, 403–408.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychol (Amst)*, 29, 387–398.

Cai, G., Sharma, R., MacEachren, A. M., & Brewer, I. (2006). Human-GIS interaction issues in crisis response. *International Journal of Risk Assessment and Management, 6(4/5/6),* 388 - 407. doi: 10.1504/IJRAM.2006.009538.

Carpin, S., Lewis, M., Wang, J., Balakirsky, S., & Scrapper, C. (2006). Bridging the gap between simulation and reality in urban search and rescue. *Robocup 2006: Robot Soccer World Cup X, Springer, Lecture Notes in Artificial Intelligence*, 1–12.

Carpin, S., Stoyanov, T., Nevatia, Lewis, M., & Wang, J. (2006b). Quantitative assessments of USARSim accuracy, *Proceedings of PerMIS*.

Carpin, S., Wang, J. & Lewis, M. (2005). High fidelity tools for rescue robotics: Results and perspectives. *Robocup 2005 Symposium.*

Casner, S. M. (2005). The Effect of GPS and Moving Map Displays on Navigational Awareness While Flying Under VFR. *International Journal of Applied Aviation Studies*, 5 (1), pp. 153-165.

Cassenti, D., Kelley, T. Swoboda, J. & Patton, D. (2009). The effects of communication style on robot navigation performance. *Human Factors & Ergonomics Soc Meeting*, San Antonio, TX.

Chappelle, W., McMillan, K., Novy, P., & McDonald, K. (2010). Psychological profile of USAF unmanned aerial systems Predator & Reaper pilots. Aviation, Space, and Environmental Medicine, 81, 339.

Chen, B., Neubert, B., Ofek, E., Deyssen, O., & Cohen, M. (2009). Integrated videos and maps for driving directions. *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. 223-232, ISBN: 978-1-60558-745-5

Chen, H., Atabakhsh, H., Tseng, C., Marshall, B., Kaza, S., Eggers, S., Gowda, H. Shah, A., Petersen, T., & Violette, C. (2005). Visualization in Law Enforcement. *Posters in CHI 2005, ACM Press*, 1268- 1271.

Chen, J. Y. C. & Barnes, M. J. (2012) Supervisory Control of Multiple Robots : Effects of Imperfect Automation and Individual Differences. Human Factors: 54(157)

Chen, J. Y. C. & Clark, B. R. (2008). Unmanned Aerial Vehicle (UAV)-Unmanned Ground Vehicle Teaming: UAV Guided Navigation. *Army Research Laboratory Report.*

Chen, J. Y. C. (2010). Effects of Operator Spatial Ability on UAV-Guided Ground Navigation. *IEEE*.

Chien, S. Y., Wang, H, & Lewis, M. (2010). Human vs. algorithmic path planning for search and rescue by robot teams, *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society* (HFES'10), 379-383.

Chien, S., Wang, H., & Lewis, M. (2011). Effects of Spatial Ability on Multi-Robot Control Tasks. *Proceedings of the 55th Annual Meeting of the Human Factors and Ergonomics Society*, (HFES 11)

Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1990). *Introduction to Algorithms*. MIT Press and McGraw- Hill, 1990, 887-897.

Darken, R. P. & Cevik, H. (1999). Map Usage in Virtual Environments: Orientation Issues. *Proceedings of the IEEE Virtual Reality Conference*. pp. 133-140.

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*, 111, 225–236.

Draper, M. H., Calhoun, G. L., Nelson, J. T., & Ruff, H. A. (2006). Evaluation of synthetic vision overlay concepts for UAV sensor operations: landmark cues and picture-in-picture. *Proceedings of the Human Factors and Ergonomic Society 50th Annual Meeting,* 121-125.

Drury, J., Scholtz, J., & Yanco, H. (2003). Awareness in human-robot interactions. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics,* 111–119.

Dykes, J. A., & Mountain, D. M. (2003). Seeking structure in records of spatio-temporal behaviour. *Computational Statistics & Data Analysis, 43(4),* 581–603.

Feldman Barrett, L., Tugade, M., & Engle, R. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130, 553–573.

GeoTime. (2012). 2012 Olympics Torch Relay Route. *Retrieved from Geotime(s) blog web site: http://www.geotime.com/GeoTime(s)/May-2012/2012-Olympic-Torch-Relay-Route-in-CSV.aspx*

Graham, A., Garcia-Molilna, H., Paepcke, A., & Winograd, T. (2002). Time as Essence for Photo Browsing Through Personal Digital Libraries. *Proceedings of Joint Conference on Digital Libraries (JCDL '02).*

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423, 534–537

Green, C. S., & Bavelier, D. (2006a). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance,* 1465–1468.

Green, C. S., & Bavelier, D. (2006b). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101, 217–245.

Gugerty, L., & Brooks, J. (2001). Seeing Where You Are Heading: Integrating Environmental and Egocentric Reference Frames in Cardinal Direction Judgements. *Journal of Experimental Psychology: Applied, 7 (3),* 251-66.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in avsynthetic work paradigm. *Applied Cognitive Psychology*, 24, 1149–1167.

Harada, S., Naaman, M., Song, Y.J., Wang, Q., & Paepcke, A. (2004). Lost in Memories: Interacting with Photo Collections on PDAs. *Proceedings of Joint Conference on Digital Libraries (JCDL '04).*

Hart, S., & Staveland, L. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In Hancock, P. A., & Meshkati, N. (Eds.), *Human mental workload,* 139-183. Amsterdam, Netherlands: Elsevier.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32, 175–191.

Hendy, K., Lao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors, 30,* 30–47.

Hirtle, S. C. (2012). Models of Spatial Cognition. In Waller, D. & Nadel, L. (Eds). *Handbook of Spatial Cognition*. 211-226. Washington, DC: APA

Hubert-Wallander, B., Green, C., & Bavelier, D. (2010). Stretching the limits of visual attention: The case of action video games. *WIREs Cognitive Science,* 2, 222–230. doi:10.1002/wcs.116

Humphrey, C. M. (2000). *Information abstraction visualization for human-robot interaction. Dissertation*

Humphrey, C. M., & Adams, J. A. (2009). General visualization abstraction algorithm for geographic map-based directable interfaces. *Proceedings of the 53th Annual Meeting of the Human Factors and Ergonomics Society, 53(15),* 970–974.

Hunn, B. (2005). The Human Challenges of Command and Control with Multiple Unmanned Aerial Vehicles. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, 20-24(5).

Hurter, C., Tissoires, B., & Conversy, S. (2009). FromDaDy: Spreading aircraft trajectories across views to support iterative queries. *IEEE TVCG, 15(6)*, 1017–1024.

Huynh, D. F., Drucker, S. M., Baudisch, P., & Wong, C. (2005). Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. *Proceedings of CHI 2005 Extended Abstracts*, 1937-1940.

Jul, S., & Furnas, G. W. (1998). Critical zones in desert fog: aids to multiscale navigation. *Proceedings of the 11th annual ACM Symposium on User Interface Software and Technology*, 97-106. San Francisco, California, United States: ACM. doi: 10.1145/288392.288578.

Kandel, S., Abelson, E., Garcia-Molina, H., Paepcke, A., & TheoBald, M. (2008). PhotoSpread: A Spreadsheet for Managing Photos. *Proceedings of CHI 2008, ACM Press,* 1749-1758.

Kapler, T., & Wright, W. (2004). GeoTime Information Visualization. *IEEE Symposium on Information Visualization 2004*, 25-32.

Kapler, T., & Wright, W. (2005). GeoTime information visualization, *Information Visualization, 4(2)*, 136-146

Kapler, T., Harper, R., &Wright, W. (2005).Correlating Events with Tracked Movements in Time and Space: A GeoTime Case Study. *Proceedings of Intelligence Analysis Conference, Washington, DC.*

Keyes, B., Micire, M., Drury, J., & Yanco, H.(2010). Improving human-robot interaction through interface evolution. In Chugo. D. (Eds.) *Human-Robot Interaction,* InTech, February 2010, ch. 14, 183–202.

Kincaid, R., & Lam, H. (2006). Line graph explorer: scalable display of line graphs using Focus+Context, *Proceedings of Int. Working Conf. Advanced Visual Interfaces,* 404-411.

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. *International the 21st International Cartographic Conference*, 1988-1995. Durban, South-Africa.

Lewis, M., Hughes, S., Wang, J., Koes, M., & Carpin, S. (2005). Validating USARsim for use in HRI research. *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, 457-461.

Lewis, M., Wang, H., Chien, S., Ma, Z., Velagapudi, P., Scerri, P., & Sycara, K. (2011). Process and performance in human-robot teams. *Journal of Cognitive Engineering and Decision Making, 5 (2),* 186-208.

Lewis, M., Wang, H., Chien, S., Velagapudi, P., Scerri, P, & Sycara, K. (2010). Choosing autonomy modes for multirobot search. *Human Factors, 52(2),* 225-233.

Lloyd, J., Persaud, N.V., & Powell, T.E. (2008). Equivalence of real-world and virtual-reality route learning: a pilot study. *Cyber Psychology & Behavior, 12(4),* doi: 10.1089/cpb.2008.0326.

Lohrenz, M. C., Gendron, M. L., Edwards, S. S., Myrick, S. A. & Trenchard, M. E. (2004). Demonstration of a Moving-Map System for Improved Precise Lane Navigation of Amphibious Vehicles and Landing Craft. *Sea Technology* pp. 39-43.

Massart, D., Smeyers-Verbeke, J., Capron, X. & Schlesier, K. (2005) Visual presentation of data by means of box plots. *Practical Data Handling, LCGC Europe,* 18(4), 215-218

Mehler, A., Bao, Y., Li, X., Wang, Y., & Skiena, S. (2006). Spatial Analysis of News Sources. *IEEE Trans. On Visualization and Computer Graphics, 12(5),* 765-772.

Micire, M., Drury, J., Keyes, B., & Yanco, H.(2009). Multi-touch interaction for robot control. *Proceedings of the 14th international conference on Intelligent user interfaces, ser. IUI '09. ACM,* 425–428.

Morison, A. (2010). *Perspective Control: Technology to Solve the Multiple Feeds Problem in Sensor Systems.* Unpublished Dissertation. The Ohio State University. August 2010.

Mosier, K. L., Sethi, N., McCauley, S., Khoo, L., & Orasanu, J. M. (2007). What you don't know can hurt you: Factors impacting diagnosis in the automated cockpit. *Human Factors, 49(2),* 300–310.

Nielsen, C., & Goodrich, M. (2006). Comparing the usefulness of video and map information in navigation tasks. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, ACM,* 95–101.

Nielsen, C., Goodrich, M., & Ricks, R. (2007). "Ecological interfaces for improving mobile robot teleoperation." *IEEE Transactions on Robotics, 23(5),* 927–941.

Nielsen, C., Ricks, B., Goodrich, M., Bruemmer, D., Few, D., & Walton, M.(2004). "Snapshots for semantic maps," *Proceedings of International Conference on Systems, Man, and Cybernetics (SMC'04),* 2853–2858.

Pepper, C., Balakirsky, S., & Scrapper, C. (2007). Robot Simulation Physics Validation. *Proceedings of PerMIS'07.*

Peuquet, D. J. (1994). It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers, 84(3),* 441-461.

Peuquet, D. J. (2002). Representations of Space and Time, Guilford, New York.

Pongnumkul, S., Wang, J., & Cohen, M. (2008). Creating map-based storyboards for browsing tour videos. *Proceedings of User interface Software and Technology (UIST).*

Riveiro, M., Falkman, G., & Ziemke, T. (2002). Visual analytics for the detection of anomalous maritime behavior. *Proceedings of International Conference Information Visualization*, 273–279.

Scerri, P., Farinelli, A., Okamoto, S., & Tambe, M. (2005). Allocating tasks in extreme teams. *Proceedings of 4th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '05),* New York, NY, USA, ACM.

Shanbhag, P., Rheingans, P., & DesJardins, M. (2005). Temporal Visualization of Planning Polygons for Efficient Partitioning of Geo-Spatial Data. *IEEE Symposium on Information Visualization*, 211-218.

Snavely, N., Seitz, S. M., & Szeliski, R. (2005). Photo tourism: exploring photo collections in 3D. *Proceedings of SIGGRAPH '06, ACM,* 835-846

Snavely, N., Seitz, S., & Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3)*, 835-846, http://photosynth.net/

Spence, R. (2007). *Information Visualization: Design for Interaction (2nd Edition).* Publisher: Prentice Hall, 304,ISBN: 0132065509.

Takahashi, O., & Schilling, R. J. (1989). Motion planning in a plane using generalized voronoi diagrams. *IEEE Transactions on Robotics and Automation, 5(2),* 143–150.

Tanaka, H., Arikawa, M., & Shibasaki, R. (2002).*A 3-D Photo Collage System for Spatial Navigations.* Springer-Verlag Berlin Heidelberg , 305-316.

Taylor, B., Balakirsky, S., Messina, E., & Quinn, R. (2007). Design and Validation of a Whegs Robot in USARSim. *Proceedings of PerMIS'07.*

Thomas, L. C. & Wickens, C. D. (2001). Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection. *Proceedings of Human Factors and Ergonomics Society 45th Annual Meeting*, pp. 336-340.

UE3 Available from: http://www.unreal.com/technology.php, accessed September 15, 2010.

Valero, A., de la Puente, P., & Rodr ǵuez-Losada, D. (2010). Evaluating a human-robot interface for exploration missions. *Proceedings of the 7th IFAC Symposium on Intelligent Autonomous Vehicles.*

Velagapudi, P., Wang, H., Scerri, P., Lewis, M., & Sycara, K. (2009). Scaling effects for streaming video vs. static panorama in multirobot search. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems.*

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),* 224-229.

Wang, H., Kolling, A., Abedin, S., Lee, P., Chien, S., Lewis, M., Brooks, N., Owens, S., Scerri, P., & Sycara, K. (2011a). Scalable Target Detection for Large Robot Teams. *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, (HRI2011)

Wang, H., Kolling, Brooks, N., Lewis, M. & Sycara, K. (2011b) Synchronous vs. Asynchronous Control for Large Robot Teams. *Proceedings of 14th International Conference on Human-Computer Interaction,* (HCII2011)

Wang, H., Lewis, M., Chien, S., Scerri, P., Velagapudi, P., Sycara, K. & Kane, B. (2010). Teams organization and performance in multi-human/multi-robot teams. *2010 IEEE International Conference on Systems, Man, and Cybernetics,* (SMC 10), 1617-1623.

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

Wang, W. (2004). *Human Navigation Performance Using 6 DOF Dynamic Viewpoint Tethering in Virtual Environments*. Unpublished Ph.D. dissertation, University of Toronto, Canada.

Wang, W. (2004). Human Navigation Performance Using 6 DOF Dynamic Viewpoint Tethering in Virtual Environments. *Unpublished Ph.D. dissertation*, University of Toronto, Canada.

Wang, J. (2007). *Human Control of Cooperating Robots* (Doctoral dissertation). Available from University of Pittsburgh Electronic Theses and Dissertations (ETD) Project. URN (etd-01072008-135804)

Ware, C. (2000). *Information Visualization: Perception for Design,* Academic Press, 2000.

Ware, C. (2004). *Information Visualization: Perception for Design.* Second Edition. Elsevier Inc., 129.

Werner, S. (2002). Cognitive Reference Systems and Their Role in Designing Spatial Information Displays. *Künstliche Intelligenz. 16 (4),* 10-13.

Werner, S., Krieg-Brückner, B., Mallot, H., Schweizer, K. & Freksa, C. (1997). Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation. *Informatik aktuell,* 41-50.

Wood, J., Dykes, J., Slingsby, A., Clarke, K., & Ware, C. (2007). Interactive Visual Exploration of a Large Spatio- Temporal Dataset: Reflections on a Geovisualization Mashup. In IEEE Trans. *Visualization and Computer Graphics, 13 (6),* 1176-1183.

Woods, D. D. & Sarter, N. B. (2010). Capturing the dynamics of attention control from individual to distributed systems. *Theoretical Issues in Ergonomics, 11(1),* 7-28.

Woods, D. D. & Watts, J. C. (1997). How not to have to navigate through too many displays. In Helander, M. G., Landauer, T. K., & Prabhu, P. (Eds). *Handbook of Human-Computer Interaction.* Elsevier Science, Amsterdam, Netherlands, 2nd edition.

Woods, D. D. (1984). Visual momentum: a concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies, 21(3):229244.*

Woods, D. D., & Sarter, N. B. (2010). Capturing the dynamics of attention control from individual to distributed systems. *Theoretical Issues in Ergonomics, 11(1),* 7-28.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002). Can we ever escape from data overload? a cognitive systems diagnosis. *Cognition, Technology & Work, 4,* 2236.

Woods, D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning human-robot coordination in future operations. *IEEE Transactions on Systems, Man, and Cybernetics: Part C Applications and Reviews, 34(2),* 210-218.

Wright, W., & Kapler T. (2002). Visualization of Blue Forces Using Blobology, *Proceedings of Command and Control Research and Technology Symposium.*

Yanco, H., & Drury, J. (2007). Rescuing interfaces: A multi-year study of human-robot interaction at the AAAI robot rescue competition. *Autonomous Robots, 22(4),* 333–352.

Yuri, A., Wren, C., Sorokin, A., & Kaur, I. (2007).Visualizing the History of Living Spaces. *Proceedings of Information Visualization 2007,* 1153-1160.

Zaratti, M., Fratarcangeli, M., & Iocchi, L. (2006). A 3D Simulator of Multiple Legged Robots based on USARSim. *Robocup 2006: Robot Soccer World Cup X.* Springer, LNAI, 13–24.