Supporting Remote Manipulation with an Ecological Augmented Virtuality Interface

J. Alan Atherton and **Michael A. Goodrich** [1]

Abstract.

User interfaces for remote robotic manipulation widely lack sufficient support for situation awareness and, consequently, can induce high mental workload. With poor situation awareness, operators may fail to notice task-relevant features in the environment often leading the robot to collide with the environment. With high workload, operators may not perform well over long periods of time and may feel stressed. We present an ecological visualization that improves operator situation awareness. Our user study shows that operators using the ecological interface collided with the environment on average half as many times compared with a typical interface even with a poorly calibrated 3D sensor; however, users performed more quickly with the typical interface possibly because of the poor calibration.

1 INTRODUCTION

A manipulator is a machine that can operate on its surrounding environment. Manipulators are commonly modeled after a human arm and hand, and aim to perform actions similar to what humans can do with their arms and hands. A *remote* manipulator is located away from the operator, so that the operator depends on sensors near the robot to provide information about the remote environment. Figure 1 shows an example of a remote manipulator and supporting sensors. A *mobile* manipulator is manipulator mounted to a mobile base, and is often also a remote manipulator.

Mobile manipulators are useful tools in areas that are dangerous or inaccessible to humans. For example, they are used in planetary exploration (Mars rovers), urban search and rescue (USAR), and explosive ordnance disposal (EOD) [16, 2, 17]. Operators use the manipulator to grab, push, poke, operate tools, and generally try to do what people do with their hands. In order to safely benefit from the capabilities these robots provide, a human operator must control them remotely. Operators can use the mobile manipulator to perform a variety of tasks from a safe location. Although supporting mobile manipulation is our end goal, the complexity involved is more than we can test in one study, so we limit the scope of this work to remote manipulation as a step toward supporting mobile manipulation.

One difficulty in remote manipulation is for the operator to understand the situation given limited data. In remote situations, the only connection the operator has with the robot is the user interface, and the operator is limited to whatever information the robot or the interface can provide. This is sometimes referred to as looking at the world through a "soda straw" [23]. Problems that come from this limited view of the world include missed events, difficult navigation in new situations, and an incomplete understanding of the explored world [23]. For example, depth perception is a very important aspect of understanding the environment for manipulation, but single-camera video displays provide limited depth perception in unstructured environments.

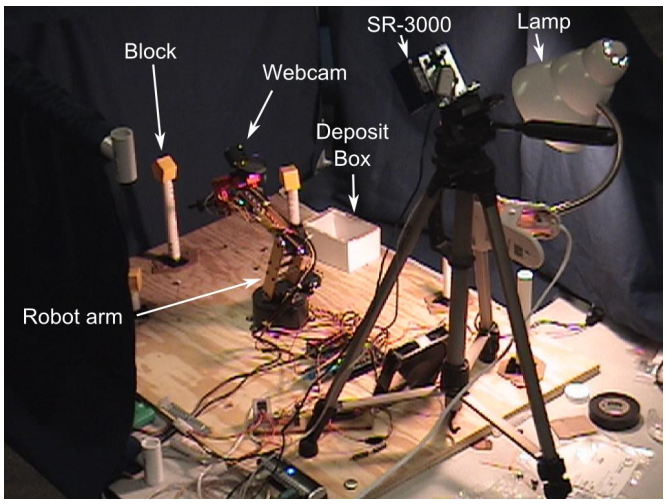

Figure 1. Remote manipulator robot, surrounding environment, and supporting sensors.

Factors that affect performance in remote manipulation tasks include the physical capabilities of the robot, the autonomous behaviors of the robot, the user interface, and the capabilities and skill of the human operator. Although all of these factors affect performance, we can focus on a few factors that a system designer can control. Assuming that we have a skilled operator, the design variables that we can freely manipulate include the following key elements: robot capabilities, the robot behaviors, and the user interface. When the robot does not have the required capabilities for a task, then the usability of the user interface or autonomous behaviors make little difference for accomplishing the task. On the other hand, when the robot does have the necessary capabilities, then the user interface and autonomous behaviors can significantly affect performance.

Typical robotic user interfaces present several sets of information in multiple windows with each window showing a distinct set of information (see, for example, Figure 2). When information is displayed disjointly in this manner, operators must mentally integrate the displays to understand relationships between different sets. While such interfaces can be very useful, including for testing and debugging a system, such a design can increase the mental workload on operators. By splitting the information into multiple windows, oper-

[1] Department of Computer Science, Brigham Young University, USA

ators have poorer manipulation or task-specific situation awareness as they may miss an important event from one display while focused on a different display.

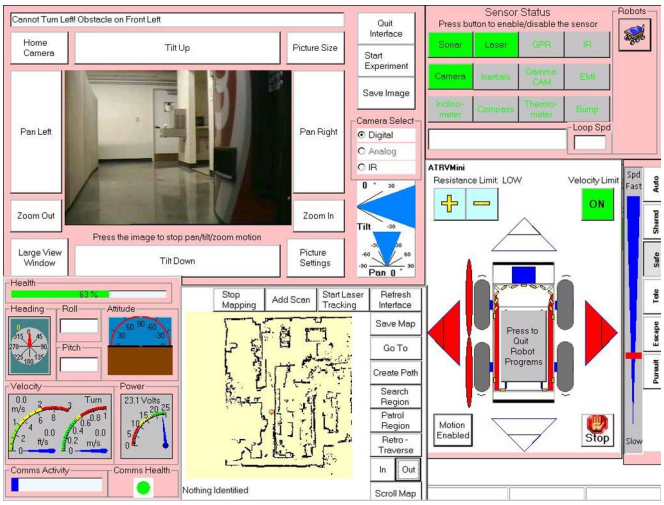

Figure 2. Idaho National Laboratory (INL) mobile robot control interface with multiple windows. Adopted from [1].

We present an ecological augmented virtuality (AV) user interface. As discussed by Vicente in [20], the term *ecological interface* originates from psychologist James J. Gibson who defines *ecological psychology* as study of human-environment relationships in a natural setting, as opposed to a laboratory setting. Ecological interface design aims to make relationships in the environment perceptually evident to the user in order to minimize workload for understanding those relationships. The term *augmented virtuality*, as defined by Paul Milgram, means the interface presents real-world information (e.g. camera video) inside a virtual environment [11]. Augmented *reality*, on the other hand, displays virtual information on top of camera video.

The ecological AV interface is inspired by work done by Milgram [12], Ricks [15], Nielsen [14], and Michaud [10]. The primary component of the interface is an ecological visualization for obstacles and objects around the robot. The visualization displays a 3D range scan in context with a virtual robot arm to show the spatial relationship between the robot and its surroundings; see Figure 3. We claim that the visualization gives the operator more viewpoints than are typically afforded by video camera systems and provides a grounded frame of reference, potentially improving performance.

The primary contribution of this research is the analysis of the benefits of using an ecological AV interface to support robotic manipulation. Another contribution is the interface design itself; in particular, the ecological 3D range scan visualization. We examine the effects of such an approach on overall performance and situation awareness.

For the remainder of the paper we will show how our approach improves situation awareness in a remote manipulation task. We discuss the motivating factors in the design of our user interface. We describe our experiment design, explain how the results show improved situation awareness, and present conclusions.

2 RELATED LITERATURE

Milgram and Drascic et al. show that stereo vision displays improve performance and accuracy in manipulation tasks, and that virtual tools augmenting the display further improve performance and accuracy [12]. People were able to more accurately position a robot arm with virtual tools including tethers, tape measures, pointers, landmarks, and object overlays. In this paper we look at a method that gives some of the benefits of stereo vision without the high bandwidth requirements and viewing hardware.

Nguyen et al. present several NASA virtual reality interfaces for scientific and space robotics, the most similar to ours called Viz [13]. Viz is a highly flexible, modular interface system that can display virtual models of robots in context with data from various sources, such as automatically generated 3D terrain models. The interface and visualization we present is similar in many ways to Viz; however, we have designed with real-time operation in mind, while Viz seems to be designed more for planning due to the long delays associated with extraplanetary robotics.

Nielsen shows that an augmented virtuality interface improves performance for mobile robot navigation and exploration tasks [14]. The interface displays a live video feed from a pan-tilt-zoom camera in context with a 2D map built with a laser rangefinder. This paper explores whether there are similar improvements when applying ecological augmented virtuality design to a manipulation task.

Tsui and Yanco et al. designed an interface for tasking a wheelchair-mounted manipulator arm [19]. Unlike other related work, the operator is in close proximity with the robot, so there is less need for situation awareness. The interface facilitates interaction between the robot arm and the operator. The operator indicates an object of interest with a joystick or touchscreen, then the robot arm autonomously moves close to the object.

3 METHODS

Our method of improving user interfaces for remote manipulation has four steps: choose a relevant problem, identify interface goals and requirements for the problem, design interfaces using different approaches, then test to see which approach is better. Thus, we break up the discussion of methods into four respective sections.

3.1 Application

The real-world application most similar to what we will use for experimentation is remote sample acquisition. The objective in sample acquisition is to navigate the robot through an environment and use a manipulator to collect samples, such as rocks, and analyze them on-site or return them to a lab for further analysis [16]. We will mimic this task and analyze the performance of operators using a variety of interface designs.

3.2 Interface Goals and Requirements

To guide the design of our interface, we choose some goals and requirements inspired by Goodrich and Olsen [5]. The primary goals are to:

1. maintain a manageable workload and
2. support situation awareness.

Requirements for the interface are to:

1. integrate information from multiple data sources,
2. provide views of the environment from multiple perspectives,
3. act as a grounding frame of reference for interaction with the robot, and
4. externalize memory.

3.2.1 *Goals*

Maintain a manageable workload.

A manageable workload means that the operator can perform the tasks without feeling overly stressed, operate in a situation where there are multiple competing tasks, and sustain operation for some period of time [18, p. 301]. Reducing the operator workload is especially important in EOD and USAR, where operators may have to work while physically and mentally fatigued [2]. To reduce the workload, we must (a) think about information presentation that leverages human perception and (b) devise control methods so that users can build correct, simple mental models [21, p. 132].

Support situation awareness. Endsley defines situation awareness as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [4]. In the context of remote manipulation, situation awareness involves understanding the environment surrounding the robot to avoid obstacles and to position the manipulator with sufficient precision to accomplish a task.

3.2.2 *Requirements*

Integrate information. When information coming from different sources is presented in separate contexts, it can be difficult to see and understand relationships that may exist between different types of data. For example, if robot position is displayed as text on one computer screen, and a map of the environment around the robot is displayed as an image on a separate computer screen, it may be difficult to quickly determine where the robot is located on the map. Integrating information can reduce mental workload by reducing the number of mental transformations that must take place to establish relationships between different sets of data [22, p. 189].

View the environment from multiple perspectives. Providing the ability to view multiple perspectives can help operators to acquire situation awareness, especially since they must understand a 3D environment on a 2D display. Several views can help the operator understand the relationship between the robot and its surroundings. This is particularly useful in manipulation tasks where depth perception plays a large role. Although a single camera only provides a single perspective, creating a virtual environment can be a way to provide multiple perspectives [14, 3]. With a virtual environment, the operator can move a virtual camera around to change perspectives.

Grounded frame of reference. Macedo et al. [9] describe how people perform better when information display is grounded with control. This means that the robot moves in a way the operator expects. When the display is rotated it may not immediately be clear what direction the robot arm should move when given a command. Grounding simplifies the operator's mental model of how the controls affect the robot arm by connecting the motion of the arm to the display [21, p. 135]. A simpler mental model means that the mental workload is lower and performance is higher. More recent work by Hiatt et al. suggests that a task-centric frame of reference yields better performance than robot-centric or view-centric frames of reference [6].

Externalize memory. When the interface includes real-time data, such as robot telemetry or a live video stream, it can be difficult to remember what happened in the past. For example, if an obstacle comes into view on the camera, then leaves the view, it is difficult to remember where the obstacle is after a while [5]. Relieving burden on short-term memory can help reduce workload and improve performance. Externalizing memory means we place information in the interface so the operator no longer needs to keep what happened in working memory [22, p. 191].

3.3 Interface Design

We designed a new interface with these requirements in mind. The interface is a type of 3D ecological AV display similar to the work done in [12, 14, 15, 10]. Our interface differs from [12] in that they use stereo vision with a fixed camera viewpoint and overlay virtual elements as in traditional augmented reality, whereas we create a virtual environment to display a 3D model that can be seen from several viewpoints. While [14, 15, 10] apply ecological AV interfaces to real-time robot navigation, we look at this type of interface for real-time robotic manipulation.

In order to reduce workload on the operator and improve situation awareness, we implemented a display that: (1) integrates information from different sources, (2) provides views from multiple perspectives, (3) gives a grounded frame of reference, and (4) externalizes memory. We explain the design of each of these parts and how each meets the requirements in the previous section.

Integrating information into one display may reduce the number of mental transformations the operator must perform to establish the relationships between different data sources, such as robot state, map, manipulator arm state, and camera video. We integrate information by rendering virtual representations of the data sources such that their spatial relationships are directly perceivable. We render a virtual graphic of the manipulator arm that reflects the position and orientation of all the robot's parts. We also display a 3D scan generated from the range image data that we explain in greater detail later in this section. See Figure 3 for a screen-shot that depicts these representations, and also shows the view the user has while using the interface. By integrating these sources into a single display, the mental workload on the operator is potentially reduced [15].

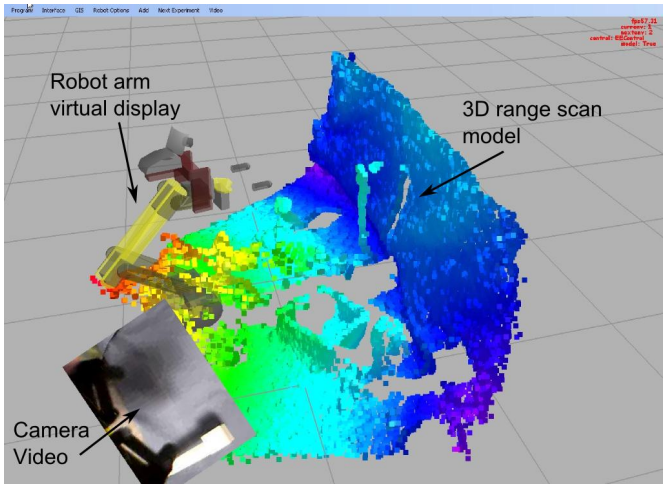

Figure 3. Augmented virtuality user interface for remote manipulation. This is the user's view in the interface.

Since we display information in a virtual environment, we can provide **virtual viewpoints** from anywhere in the environment. The user is enabled to change the "look-from" point of the virtual camera, and the focal point of the camera is set on the center of the workspace of the robot. These adjustments give multiple perspectives to the user and enhance understanding of the world surrounding the robot arm.

While multiple perspectives are nice, they introduce a problem with controlling the robot arm. When the user indicates a direction for the robot arm to move, should the robot move according to its own frame of reference or from the virtual perspective's frame of reference? In essence, this is a problem of **grounding**. We implemented two control methods: the *robot frame-of-reference control* method is direct, independent control of each joint, and the *grounded control* method is a "flying the end effector" control method where the user controls an invisible target point to which the arm reaches. In other words, the user moves the target position for the gripper, and the robot decides how to move each joint to reach that position. Controlling the end effector has been shown to reduce workload dramatically compared to controlling the joints, although situation awareness is negatively impacted when controlling the end effector [7]. We hypothesize that the negative situation awareness effects will be reduced as a result of the integrated nature of our display. With end-effector control, the orientation of the virtual view affects the direction the target point moves (see Figure 4). In other words, commanding the target point to move right will cause the target to move right in the computer screen frame of reference, and not necessarily in the robot's frame of reference. This can cause confusion while focused on the camera video (because the video is in the robot's frame of reference), so we rotate the camera video to give a sense for what direction the robot will move relative to the camera video. The grounded control connects the control of the robot arm to the virtual display of the world.

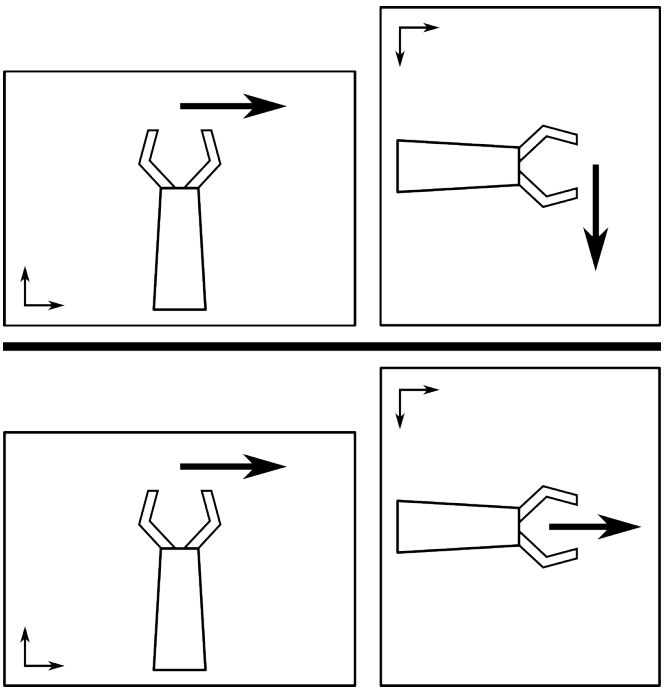

Figure 4. Two frames of reference for controlling a robot arm. All boxes show the robot arm from a top-down perspective. The robot arm is in the same position in each box, but the view is rotated. The top row shows the direction the end effector would move when commanded to move right when using a robot-centric reference frame. The bottom row shows the direction when using a view-centric reference frame.

Part of the world around the robot is represented by a 3D scan. The scan is captured from a ranging camera (but only when the user requests the scan) and then displayed as a set of colored splats. The color assigned to each splat is a function of depth in a "heat map" gradient. We align the 3D scan with the graphic of the robot arm simply by manual measurements that result in a degree of misalignment. The misalignment is not constant and in some cases appears to be perfect, but in some cases is off by as much as 2 cm. We leave improved alignment to future work.

The 3D scan presents a perceptual display of information about the world and thus potentially improves situation awareness while reducing workload. Such a display also provides **externalized memory**, as the 3D scan provides a greater field-of-view than the video camera.

4 User Study

In the user study we evaluate the usefulness of the 3D scan compared with the camera video as well as two different control methods: (1) joint control and (2) end-effector control. We hypothesized that users would perform tasks most quickly and with the fewest collisions while using the 3D scan, camera video, and end-effector control together. We measure the performance of users as they collect blocks with a robot arm in two control-type conditions (joint and end-effector) and three visualization conditions (scan-only, video-only, and scan+video).

4.1 Experiment Setup

For our experiment, we designed a simple object collection task. Participants in the experiment were required to pick up 3 small yellow blocks (2.8 cm wide) and drop them into a cardboard box (8W x 10L x 6H cm). These target blocks are placed on top of posts with heights 6, 10, and 18 cm. Seven different layouts specify the position for each post and the deposit box to prevent users from memorizing the course. In most of the layouts, the posts are placed within relatively easy reach of the robot arm, but a few configurations require the operator to almost fully extend the arm to reach the block. Curtains obscure the robot from the operator's view, although audio cues are present (primarily when a block falls from its post or is deposited).

The robot system is comprised of a robotic arm, sensors, computers, and a joystick controller. See Figure 1 for a photograph of the experiment setup.

The robot arm used in the experiment is a modified Lynxmotion 6 degrees-of-freedom (DOF) arm (5 DOF + gripper). The modified arm can reach approximately 30 cm and lift approximately 0.5 kg. The gripper measures 5 cm when fully open, so there is a 2 cm clearance between the gripper "fingers" and the target blocks. The robot servos provide no position or force feedback. We modified the arm with higher-power servos and added position feedback for the gripper. Feedback on the gripper allows us to display the gripper correctly when an object is grasped; otherwise, the gripper would appear completely closed when grasping an object even though it should appear only partially closed.

The experiment used a ranging image sensor and a webcam. The ranging image sensor is a Swiss Ranger SR-3000 mounted on a stationary tripod above and behind the robot arm. This positioning simulates a configuration where the sensor is mounted above and behind the arm on a mobile robot base. In a real mobile manipulator situation, the base is often stationary during manipulation. Future work will consider the arm mounted to a mobile base. Upon user request, the SR-3000 provides a 3D point for each pixel in the image (dimension 176 by 144 pixels). A generic USB color webcam, mounted to the arm, has a resolution of 320 by 240 pixels and a frame rate

of approximately 8 frames per second in our environment. There is approximately a 0.25-second delay on the camera video and robot commands, and about a 1-second delay on the 3D range data.

The controller is a Logitech WingMan RumblePad (see Figure 5), and of its features we use two analog thumb joysticks, a digital directional pad, six thumb buttons, and two finger "shoulder" buttons. This controller is very similar to present-day common video game controllers. The thumb joysticks control the motion of the robot arm, the directional pad controls the virtual view orientation, and the buttons control the virtual view zoom, operate the gripper, send the arm to stow position, and request a 3D snapshot.

Figure 5. Logitech WingMan RumblePad controller used to control the robot arm and user interface in the experiment.

4.2 Procedure

Before running the actual tests, each participant watched a tutorial video and performed a small training exercise to become familiar with the system. The tutorial included instructions to work quickly while trying to avoid collisions with the environment. The training exercises allowed the participant to explore how to change the view, how to control the robot arm in two different modes, and how to interpret the 3D range scan. After the tutorial, each participant collected at least one block with the whole system running the same as a normal experiment run, except the data was not recorded. Most participants spent about 15 minutes for the entire training portion. Participants were allowed to ask questions freely during the training. During the actual tests, however, only questions related to how to use the system were answered and not questions regarding the state of the world. Once finished with training, the users transitioned into the actual testing.

For the actual tests, each participant performed the task with 3 out of the 6 possible interfaces. The six variations are listed in Table 1. The participants progressed through the tests in pseudo-random, counterbalanced order to reduce order-dependent effects.

At the end of each test, a small survey had the users rate the interface on a scale of 1 to 10 with the following questions: 1. How much effort was required to use this interface effectively? 2. How difficult was it to learn this interface effectively? 3. How much confidence did you have in the robots actions? At the end of all tests, the users completed an additional survey with the following questions: 1. How much time in a week do you spend playing video games? 2. Rank the interfaces in order of your preference (selection from a list). 3. Any comments? At this point the experiment procedure was finished.

The independent variables in the experiment structure are (a) user interface display and (b) manipulator control mode. The user interface displays and manipulator control modes are explained in greater detail in Section 3.3.

Table 1. Variations in the user study. Parentheses show acronyms for each variant.

Interface/Control	*Joint control*	*End-effector control*
Video only	Variant 1 (VJ)	Variant 2 (VE)
3D scan only	Variant 3 (SJ)	Variant 4 (SE)
Video and 3D scan	Variant 5 (SVJ)	Variant 6 (SVE)

As shown in Table 1, variant 1 corresponds to the typical interface and control model. While not every traditional interface looks like this, we tried to capture the most common elements between these interfaces, namely the live video stream and current robot arm position. We left out some elements common to interfaces that are not relevant to the task, such as battery level indicators or robot "health status" information. The remaining variants of the interface use different arm control modes and the AV interface previously shown in Figure 3.

We measured operator performance primarily by time to complete tasks and number of collisions. To make these measurements, we recorded video footage from a separate camera behind the robot arm, then analyzed the video post-hoc to measure timings and collisions. See Table 2 for a listing of metrics and their measurement methods.

Table 2. Metrics and their respective methods for measurement in the user study.

Metric	Measurement
Manipulation time	Time from first arm movement to block deposit
Operator workload	Time performance, subjective survey
Situation awareness	Collisions
Quality of interactions	# interactions, time spent interacting
Operator preferences	Subjective survey

5 RESULTS

We present results for 30 people that participated in the user study. Twenty-four male and six female students at Brigham Young University participated. Participants were primarily young adults near age 18. Although we present results for 30 people, 34 participants actually attempted the experiment. One participant simply was not able to perform the task under any conditions, two participants were highly distracted, and one performed the task when a bug in the system had significantly slowed things down. Because we want as much as possible to avoid measuring such effects, we do not include data from these participants in the analysis. On three runs the system crashed after an hour of usage, though in each of these cases the participant was nearly finished, so we kept this data in the analysis. The primary results of interest that we found relate to (1) manipulation time, (2) collisions, and (3) subjective measures.

5.1 Manipulation Time

Participants performed the tasks fastest in the video-only conditions followed closely by the scan+video conditions, with scan-only slowest by a significant margin. People worked faster with joint control compared to end-effector control. See Figure 6 for the distributions of manipulation times, and Table 3 for significance between conditions. For the statistical significance analysis, we first use log correction to warp the data to a normal distribution. The D'Agostino-Pearson normality test shows that the warped data is significantly different from normal unless we remove statistical outliers (measured as 1.5 times the distance between quartiles 1 and 3), so we remove these outliers for significance analysis. The results in Table 3 show p-values for the corrected data.

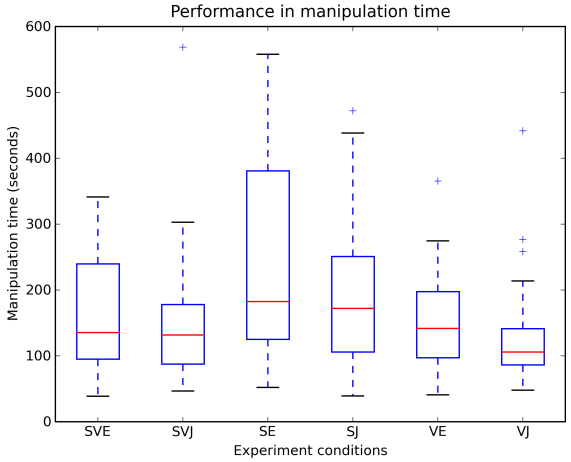

Figure 6. Performance measured as time required to perform a manipulation task. Note: To make this plot more readable, we do not show outliers above 600 seconds.

Table 3. Statistical significance analysis with p-values for two sample two-tailed t-tests. Bold numbers are significant at $\alpha = 0.05$. Note: S denotes 3D scan display, V denotes camera video display, E denotes end-effector control, and J denotes joint control.

Conditions	*SVE*	*SVJ*	*SE*	*SJ*	*VE*
VJ	**0.014**	0.080	**< 0.001**	**< 0.001**	**0.021**
VE	0.550	0.575	**< 0.001**	0.185	
SJ	0.574	0.065	**0.023**		
SE	**0.010**	**< 0.001**			
SVJ	0.287				

Several things might have caused the difference in time performance but we will discuss only a few. The likely largest factor in the slowness of the scan-only interfaces is misalignment between the 3D range scan and the virtual representation of the robot arm. In some cases, the misalignment was as much as 2 cm, which is substantial considering the clearance between the gripper opening and the block is about 1 cm on either side. Participants generally trusted the 3D range scan at first, but as they realized that the alignment was imperfect, they often grew frustrated in scan-only conditions. Because of the misalignment, users had to simply guess where and how the alignment deviates. On the other hand, with video conditions, users had undistorted visual feedback on the position of the gripper in relation to a target block, although depth perception is decreased.

Another factor in time performance involves view adjustments. When the 3D range scan was present, users spent some of the time adjusting the virtual perspective, while with the video-only interfaces almost no adjustments were made. Adjusting the view does not affect the viewpoint of the video camera, so people did not need to adjust the view much. One exception to this is in the end effector control mode, where the video is oriented corresponding to the rotation of the robot arm and the virtual viewpoint to give a sense for what direction the robot arm will move in the camera view. As such, users adjust the view somewhat in end effector control mode, although very little compared to when the 3D scan was present.

Another factor in time performance is the speed of the manipulator arm between control modes. In end-effector control the gripper moves at constant velocity in cartesian space regardless of position in the environment. On the other hand, joint control moves at constant angular velocity, which means that the farther the arm is extended, the faster the gripper moves in cartesian space. Participants tended to operate the arm in a more extended configuration, so end-effector control ended up moving somewhat slower overall.

5.2 Collisions

We separate collision types into two classes: (1) environment collisions and (2) target block collisions. Target block collisions are contact between the robot arm and the target blocks, except when a grasp action is underway. Environment collisions are contact between the robot arm and any part of the environment except target blocks. We report the total number of collisions for each condition across all participants for environment collisions in Figure 7 and for target block collisions in Figure 8. Because of high variability and a small sample size, this data has low statistical significance. We will look at statistical significance for collisions in future work.

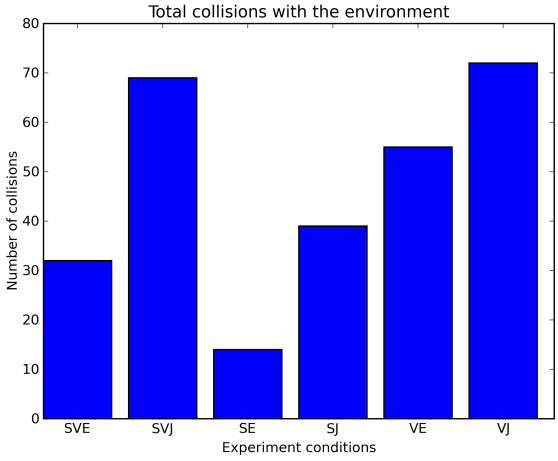

Figure 7. Total collisions with the environment.

In terms of displays, more environment collisions occurred with the video-only and scan+video interfaces than with the scan-only interfaces, while block collisions happened more with the scan-only interface than with any of the video interfaces. The alignment of the 3D scan is poor and coarse, so fine-grained, precise manipulation is difficult with scan-only interfaces. However, the granularity is suf-

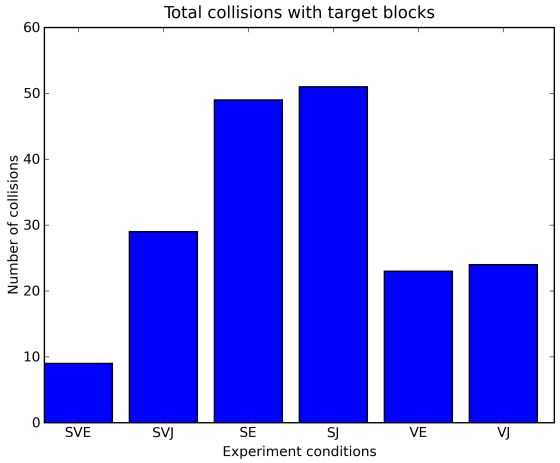

Figure 8. Total collisions with target blocks.

ficient for overall situation awareness and avoiding obstacles on a larger scale.

Even though the 3D scan is present in the scan+video interfaces, the video draws users' attention away from the 3D scan and so users appear to fail to notice imminent collisions [8].

Another factor in the number of collisions has to do with sensors. The webcam is closer to the environment than the SR-3000, so less of the environment is visible through the webcam. Such a view gives no information about what is to the sides of the camera, so users bump into things before they realize an obstacle is in the way. The webcam gives no depth information directly, so the user must use visual cues (for example shadows, relative sizes, and motion parallax) to determine depth.

When depositing a block with a 3D scan present, users often would drop the block well above the deposit box, while with video-only interfaces they usually had to make sure the block was inside the box before releasing the block. This reach-in approach generally caused several collisions with the deposit box, whereas the drop-from-above approach kept the arm clear of obstacles.

When we consider control type, more collisions occurred with joint control than with end-effector control. Joint control has a fixed angular resolution; that is, moving the control stick to the left rotates the robot arm a fixed number of degrees. On the other hand, end-effector control moves in cartesian space, so the resolution is finer as the arm extends as compared to joint control. Control resolution impacts the results because as the gripper nears an object, the user must finely adjust the position of the gripper for the final alignment.

We will present ANOVA for pairwise effects in future work.

5.3 Subjective Measures

In addition to objective measures for manipulation time and collisions, we also recorded a few subjective measures. Users answered the questions mentioned in Section 4.2. The results have sufficiently high variance that no conclusions can be made based on statistical significance, but we discuss trends anyway. For each of the subjective measures, we look at how many times a particular condition was rated better than the other conditions. Then we sum each of the condition votes for a global ranking. We look at subjective results for preference in overall interface, workload, learning required, and confidence in the robot's actions.

More users preferred joint control over end-effector control, except with the SVE condition. Users preferred variants with the 3D scan and video both present (SVE, SVJ). The preference ordering for all interfaces is as follows ($\sim$ denotes indifferent to and $\succ$ denotes preferred to): SVE $\sim$ SVJ $\sim$ VJ $\succ$ SJ $\succ$ SE $\succ$ VE. One possible reason for the low ranking of the scan-only interfaces (SE, SJ) is the misalignment between the virtual robot graphic and the 3D scan display. This disconnect led to frustration in the users, and that is reflected in this ranking. Video rotation (see Section 3.3) is probably the cause of dislike in the case of the video-only end-effector (VE) condition.

No obvious classification appears in the results for workload between the conditions. The preference ordering for workload is: SVE $\succ$ SJ $\succ$ VJ $\succ$ SVJ $\sim$ VE $\succ$ SE. Note that although performance was not the highest for the SVE condition, users apparently feel that the workload is lowest for the SVE condition. Strangely, SJ also ranks well for workload keeping in mind the misalignment between the 3D scan and robot arm graphic.

In terms of learning required to effectively operate the interface, joint control was generally easier to learn than end-effector control. The preference ordering for learning is: VJ $\succ$ SJ $\succ$ SVJ $\sim$ VE $\succ$ SVE $\succ$ SE. Joint control is most similar to many common types of control that people may have already been exposed to, such as driving a remote control car. End effector control is a new concept for most people, so more training is required to understand and use it effectively.

When it comes to understanding how the robot will move when given commands, users tended to feel more confident with joint control, except for the SVE condition. The preference ordering for confidence is: SVE $\succ$ VJ $\succ$ SJ $\succ$ SVJ $\succ$ VE $\succ$ SE. Part of the lack of confidence with end-effector control is likely due to the constraint for the target point. Normally it is best to constrain the target point to the robot arm's workspace, but due to implementation error we constrain it to a box that covers the robot's workspace. The side effect of such a constraint is that when the target point moves outside of the workspace, the robot's actions are confusing until the target point moves back inside the workspace. In the future it would be better to constrain the target point to the robot's workspace.

5.4 Discussion

The results seem to indicate that the 3D scan may actually increase workload, although it also improves situation awareness. Multiple perspectives with the 3D scan seem to provide better situation awareness, as fewer environment collisions happened when the 3D scan was present. In addition, it appeared from subjective observation that users who changed their view to check alignment had fewer collisions with the environment; future work should test this claim. Only a coarse understanding of the environment was possible with the resolution and calibration of the 3D scan, as indicated by a higher number of object collisions when video was absent. Without a good calibration for the robot arm and ranging camera (both intrinsic and exterior calibration), operators cannot trust the 3D scan to be accurate, and it may encourage them to depend more on camera video.

End-effector (grounded) control was poorly rated except when combined with the 3D scan and camera video even though fewer environment collisions happened with this control. This may be due in part to the view-dependent nature of the control. As an example case, we can consider when a user moved the gripper close to an object, and then changed the virtual viewpoint to gather more understanding. After the view change, the controls also change, and so

when the user indicates the same direction on the controller as before, the robot moves differently. Apparently users think more from the robot's frame of reference than the computer screen's frame of reference. Perhaps more training with this control method would improve understanding, appeal, and performance.

Since the scan-only display with end-effector control exhibits the fewest world collisions and the scan+video display with end-effector control exhibits the fewest block collisions and has reasonable time performance, we predict that a better camera video integration with the 3D scan would reduce collisions while maintaining reasonable performance. This better integration could be done by virtually projecting the video in the 3D space near the virtual display of the robot arm, as if the virtual depiction of the camera was projecting the video.

6 CONCLUSIONS AND FUTURE WORK

Remote manipulation requires operators to control robots in complex environments with limited information. Typical user interfaces present information to the operator in the form of several camera video feeds and data displayed in several other channels. Although information required to perform tasks is available, the mental workload seems to be high enough that some of the information is forgotten or missed.

Evidence from the experiment suggests that although the ecological AV user interface for remote manipulation slows performance, it can also increase situation awareness and lessen mental workload. We believe that users were able to better understand the spatial relationship between the robot arm and its environment with the ecological interface compared to traditional interfaces. This represents a small step toward providing support for *mobile* manipulation. As this is an exploratory study, we have identified several problem areas with room for improvement. Future work needs to further test and refine these ideas.

In the future, we plan to make the display more integrated and ecological. We can add more autonomy to the robot arm, such as a "move close to that object" behavior. Because users spend much time adjusting the virtual perspective, we can improve the interaction method for adjusting the view with, for example, head tracking. Although we looked at operator workload with subjective measures and based on time performance, we can also use formalized methods such as NASA TLX or more objective measures such as secondary task performance.

ACKNOWLEDGEMENTS

We give thanks to Idaho National Laboratory and the employees there who supported this work. We also give thanks to ARL for support.

REFERENCES

[1] D. J. Bruemmer, D. A. Few, R. L. Boring, J. L. Marble, M. C. Walton, and C. W. Nielsen, 'Shared understanding for collaborative control', *IEEE Transactions on Systems, Man and Cybernetics, Part A*, **35**(4), 494–504, (July 2005).

[2] J. Casper and R.R. Murphy, 'Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center', *IEEE Transactions on Systems, Man and Cybernetics, Part B*, **33**(3), 367–385, (2003).

[3] Joseph Cooper and Michael A. Goodrich, 'Towards combining uav and sensor operator roles in uav-enabled visual search', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 351–358, Amsterdam, The Netherlands, (2008). ACM.

[4] M. R. Endsley, 'Design and evaluation for situation awareness', in *In Proceedings of the Human Factors Society 32nd Annual Meeting*, pp. 97–101, Santa Monica, CA, (1988).

[5] M. A. Goodrich and Jr. Olsen, D. R., 'Seven principles of efficient human robot interaction', in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pp. 3942–3948 vol.4, (2003).

[6] L. M. Hiatt and R. Simmons, 'Coordinate frames in robotic teleoperation', in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1712–1719, (9–15 Oct. 2006).

[7] D.B. Kaber, J. Riley, R. Zhou, and J.V. Draper, 'Effects of visual interface design, control interface type, and control latency on performance, telepresence, and workload in a teleoperation task.', in *In Proc. XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*, (2000).

[8] R. Kubey and 79-86. M. Csikszentmihalyi, 286(2), 'Television addiction is no mere metaphor', *Scientific American*, **286**(2), 79–86, (2002).

[9] Jose A. Macedo, David B. Kaber, Mica R. Endsley, Preecha Powanusorn, and Seonwan Myung, 'The effect of automated compensation for incongruent axes on teleoperator performance', *Human Factors*, **40**(4), 541–553, (December 1998).

[10] Francois Michaud, Patrick Boissy, Helene Corriveau, Andrew Grant, Michel Lauria, Daniel Labonte, Richard Cloutier, Marc A. Roux, and David Iannuzzi, 'Telepresence robot for home care assistance', in *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, (March 2007).

[11] P. Milgram and F. Kishino, 'A taxonomy of mixed reality visual displays', *IEICE Transactions on Information Systems*, **E77-D**(12), 1321–1329, (1994).

[12] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, 'Applications of augmented reality for human-robot communication', in *Intelligent Robots and Systems '93, IROS '93. Proceedings of the 1993 IEEE/RSJ International Conference on*, volume 3, pp. 1467–1472 vol.3, (1993).

[13] L. A. Nguyen, M. Bualat, L. J. Edwards, L. Flueckiger, C. Neveu, K . Schwehr, M. D. Wagner, and E. Zbinden, 'Virtual reality interfaces for visualization and control of remote vehicles', *Autonomous Robots*, **11**(1), 59–68, (2001).

[14] Curtis W. Nielsen, *Using augmented virtuality to improve human-robot interactions*, Ph.D. dissertation, Brigham Young University, 2006.

[15] Bob Ricks, Curtis W. Nielsen, and Michael A. Goodrich, 'Ecological displays for robot interaction: A new perspective', in *Proceedings of IROS 2004*, Sendai, Japan, (2004).

[16] Paul S. Schenker, Eric T. Baumgartner, Sukhan Lee, Hrand Aghazarian, Michael S. Garrett, Randall A. Lindemann, D. K. Brown, Yoseph Bar-Cohen, Shyh-Shiuh Lih, Benjamin Joffe, Soon S. Kim, B. D. Hoffman, and Terrance L. Huntsberger, 'Dexterous robotic sampling for mars in-situ science', in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ed., David P. Casasent, volume 3208 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 170–185. SPIE, (September 1997).

[17] Jean Scholtz, Mary Theofanos, and Brian Antonishek, 'Development of a test bed for evaluating human-robot performance for explosive ordnance disposal robots', in *HRI '06: Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 10–17, New York, NY, USA, (2006). ACM Press.

[18] Thomas B. Sheridan, *Telerobotics, automation, and human supervisory control*, MIT Press, Cambridge, MA, USA, 1992.

[19] Katherine Tsui, Holly Yanco, David Kontak, and Linda Beliveau, 'Development and evaluation of a flexible interface for a wheelchair mounted robotic arm', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 105–112, Amsterdam, The Netherlands, (March 2008). ACM.

[20] K.J. Vicente and J. Rasmussen, 'Ecological interface design: theoretical foundations', *Systems, Man and Cybernetics, IEEE Transactions on*, **22**(4), 589–606, (1992).

[21] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, Prentice Hall, Upper Saddle River, NJ, 3rd edn., 2000.

[22] Christopher D. Wickens, John Lee, Yili D. Liu, and Sallie Gordon-Becker, *Introduction to Human Factors Engineering (2nd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.

[23] D.D. Woods, J. Tittle, M. Feil, and A. Roesler, 'Envisioning human-robot coordination in future operations', *IEEE Transactions on Systems, Man and Cybernetics, Part A: Applications and Reviews*, **34**(6), 749–756, (November 2004).