iVR Control

An immersive VR Simulation for Rover Navigation and Control

by

Hudson Pridham

A thesis presented to OCAD University
in partial fulfilment of the thesis requirements for the degree of
Master of Design in the
Digital Futures Program

Toronto, Ontario, Canada, April 2014

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I authorize OCAD University to lend this thesis to other institutions or individuals for the purpose of scholarly research. I understand that my thesis may be made electronically available to the public. I further authorize OCAD University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature ______________________________________

Abstract

Tackling issues endemic to traditional remote rover control systems, such as cognitive workload arising from unfamiliarity with the robot, its characteristics, and its context, this thesis explores the use of a VR headset and immersive simulation to enable robot operators to view the robot being operated in its context. Needs, insights, and solutions were gathered through a study of relevant literature, rover operator interviews and job shadows, and three prototype development and testing sprints. These provide evidence that gestural controls coupled with immersive VR interfaces can improve rover operator's abilities to establish situational awareness and complete traditionally complex tasks, such as robot arm repositioning and task switching. This thesis concludes with six key insights concerning the creation of VR control systems for rover operation: affordance, consistency, communicability, feedback loop, spatial memory, and simulation sickness.

Acknowledgements

To my wife, Kate, you're the smartest person I know. I couldn't have done this without you.

My thanks goes out to the great people at Ontario Drive and Gear. You're doing great things and I was honoured to be included in your work.

Additional thanks go to Nick Puckett for technical support and use of his equipment.

Table of Contents

List of Figures and Tables vii

Chapter 1: Introduction

Background and Motivation 2
Research Question..5
Scope of Research 6
Overview and Organization.....................................7

Chapter 2: Literature Review

Control and Embodiment......................................11
Metaphor and 3D Interfaces.................................18
Feedback and Affordance 22

Chapter 3: Methodology and Research Design

Chapter 4: Needs Assessment and Ideation

Needs Assessment...33
Initial Solution Ideation 41
DIY Components Assessment..............................44

Chapter 5: Prototype Development and Testing

Testing and Insight Collection Method 51
Test Equipment Design...53
Test 1 55
Test 2 and 3 64
Insights 77

Chapter 6: Conclusion

Discussion..84
Future directions 86
Bibliography...88
Appendix..91

List of Figures and Tables

Page	Fig.	Title/Source
		Literature Review
12	2.1	Exocentric mental model
12	2.2	Egocentric mental model
12	2.3	Hybrid map and camera feed display (Sanquino et al., 2012)
15	2.4	Operating room image navigation through gesture tracking (O'Hara, 2014)
16	2.5	Six DOF (Six degrees of freedom, n.d.)
20	2.6	2D icon layout (Ark et al. 1998)
20	2.7	2D/3D hybrid icon layout (Ark et al. 1998)
20	2.8	3D icon layout (Ark et al. 1998)
24	2.9	ISS robot workstation (Lapointe & Massicotte, 2003)
24	2.10	ISS 3D robot control overview (Lapointe & Massicotte, 2003)
		Needs Assessment and Ideation
33	4.1	Rover navigation up stairs (Source: Guizzo, 2012)
33	4.2	Fukushima rover control system (Source: Guizzo, 2012)
34	4.3	ODG mobile platform (rover)
34	4.4	ODG handheld controller
37	4.5	Example and coverage of the CSA's rover 3D imaging tool. Image credit: Canadian Space Agency
39	4.6	Two ODG rover operators working in tandem
39	4.7	CSA test setup
43	4.8	Rover maps its context via depth cameras
43	4.9	A 3D terrain model of the rover's context is generated from the data provided by the depth cameras

43 4.10 A 3D model reflecting the state of the rover is placed in the 3D terrain model

43 4.11 The operator's gestures are tracked via depth camera

43 4.12 the operator experiences the 3D model of the rover's context via a VR headset

43 4.13 Combined gesture tracking and VR headset enable natural interaction with the rover

45 4.14 Oculus Rift VR Headset

47 4.15 Kinect tripod setup

Development

51 5.1 Testing timeline

53 5.2 Kinect and Oculus Rift setup

53 5.3 View of the tester by the Kinect (Skeleton Tracking)

54 5.4 3D avatar in Unity

54 5.5 Unity, having mapped the Kinect's skeleton tracking to the avatar, displaying stereoscopic images seen by the operator

55 5.6 Rover prototype and initial control system

56 5.7 Physical obstacle course

56 5.8 Virtual obstacle course

57 5.8 Remote desktop interface test

57 5.10 Immersive virtual interface test

57 5.11 Physically present handheld interface test

58 5.12 Tester avatar being dragged by the virtual rover

60 5.13 Cable imposed movement constraints and tangling

61 5.14 Red safety ring

62	5.15	Diagram of the rover turning without dragging the operator
62	5.16	Diagram of operator walking back behind the rover
64	5.17	Rover equipped with the robot arm
64	5.18	Rover at the Fukushima cleanup site using its onboard robot arm to open a door
65	5.19	Test 2 and 3 target course
66	5.20	Test 2 - direct manipulation control method
67	5.21	Test 3 - menu based control method
68	5.22	Initial VR Glove prototype
72	5.23	Tester ducking while repositioning the robot arm into a floating target
72	5.24	Tester standing inside of a floating target while repositioning the robot arm
75	5.25	Test 3 menu based control method showed in context with the rover and operator's avatar
78	5.26	Ghosting and ducking to explore the environment
78	5.27	Reaching outside the Kinect's field of view and putting tension on the Oculus Rift's cable
79	5.28	Tactile cues via vibration motors indicating grab
80	5.29	3 axis robot arm manipulation using gestures
80	5.30	2 axis rover navigation using handheld controls
81	5.31	Camera view
81	5.32	VR view
82	5.33	Gamepad locomotion versus physical locomotion

Chapter 1: Introduction

Background and Motivation

Recent exploration efforts on Mars coupled with the tragic events and subsequent cleanup efforts at the Fukushima nuclear reactor have shown the pressing, if not urgent need to develop and advance remote telepresence and rover control systems. Science and industry are increasingly deploying remote robot telepresence in a wide variety of areas, including exploration, search and rescue, and construction. As these machines increase in complexity so do also their interfaces and methods of control. Existing technologies require their operators to undergo extensive training to properly comprehend and affect their controls properly (Guizzo, 2012; Lapointe & Massicotte, 2003). If training is insufficient or the system too complex the user can fail to grasp the meaning of key interface elements and the metaphors they employ by appearance alone.

In their research paper covering the education of computing principles to teenage student Syslo & Kwiatkowska (2005) present similar lines of argument, stressing the inability of students in their study to develop the mental models and the set of cognitive processes required to properly leverage and navigate these technologies directly from the software alone. Standard interface metaphors breakdown in complex applications when employed in a discovery based learning environment. As exemplar of this point, Blackwell (2006) describes how, at the 2003 User Interface Design workshop he attended, interface metaphors were described as a visual communication channel via which the designer of the interface achieves the rapid transfer of an effective mental model into the user’s head. This description implies the false presupposition that users are passive recipients of new understanding. Active, subjective, contextualized, or embodied interpretations by users of interface metaphor are problematic because users might discover or construct new interpretations, rather than understanding the metaphor and receiving the expected interpretation of the interface element.

Lastly, Don Norman states in the Design of Everyday Things, that it is lack of visibility and communication of active processes that makes so many computer-controlled devices, such as tele-operated rovers, so difficult to operate. Likewise, it is an excess of visibility that makes feature-laden products so intimidating because it obscures what is currently taking place (Norman, 1988/2002).

In the context of this thesis I will argue that rovers used in remote and hazardous environments are not exempt from these assertions. Despite being tasked with highly important scientific and humanitarian missions (I would argue because of it) it is important for these robots to be intuitive to operate. To do so interface designers must provide more methods of visualizing active processes relevant to facilitating human computer interactions. Yet, as Blackwell and Syslo & Kwiatkowska warn, these methods, which take shape through the use of interface metaphors, can be interpreted incorrectly and lead to confusion. As such, a system which introduces redundancy in the communication channel by using 3D visualization in addition to metaphor can reinforce meaning and provide less opportunity for operator error.

Prior research projects concerning rover operation have explored this problem through the representation of the interface on a 2D/3D display hybrid. This took the form of a camera feed from a remote rover placed directly above a 2D map of the rover's surroundings with the aim of representing the rover's current condition more naturally to the operator. This display method was found to improve operator performance across all measures, including keyboard mistakes, map disorientation, and collision with objects (Sanguino et al, 2012).

My research builds on this approach, providing additional cues and information by moving the interface into a fully 3D environment as seen through a virtual reality headset. Such a system provides additional benefits beyond ensuring the correct interpretation of control metaphors and seamless visualization of active processes: it has been understood for centuries that human memory works best when attached to particular spaces and locations. Ancient Romans and Greeks used the ‘memory palace’ mnemonic system to recall large treatises by mentally associating key phrases with specific locations in a fictional palace. (Yates, 2011) This thesis explores the creation of a virtual reality simulation to house a natural interface which leverages this innate spatial memory ability. This can be done by ensuring the interface maintains a consistent proximity and relationship to the user’s body and fixed objects such as the rover. Much like a watch on a wrist, inside this simulation controls are located in virtual space around the user's body. There are only so many places that input/output devices (I/O) can be located on the human body, hence the need to extend the body and I/O into virtual space.

Research Question

This thesis aims to explore the different experiences of robot operators in viewing a simulation of a robot in its context through a VR headset and viewing the robot's context through its onboard camera. The main question I will be addressing in this study is how VR headsets and the afforded view of a robot in context aid in building an operator's spatial model of that robot's context. In addition, I will also be questioning what cognitive outcomes are derived by locating interface elements in relation to an operator's body and the rover, rather than a third party (ie. keyboard, mouse, or gamepad).

I Hypothesize that using a VR headset and simulation to enable robot operators to view the robot being operated in context will allow those operators to overcome some of the issues endemic to controlling robots, such as unfamiliarity of the rover, its context, and the rover's relationship to it. Furthermore, unlike traditional point of view control systems, using a VR simulation as a control method and subsequent 3D interface will encourage the formation of spatial memory and recall. This is achieved by allowing for a view of the rover in context and for the locating of interface elements in relation to operators' bodies and key locations on the rover.

Scope of Research

Rather than quantitative metrics concerning key performance gains, this thesis is primarily concerned with the qualitative study and analysis of the user experience with the virtual reality system under development and outlined in the body of this text. Furthermore, the scope of research has been limited to one particular application of this technology, extra-planetary exploration. While this excludes possible insights derived from other applications of tele-robotics it allows for the targeted assessment and service of those needs found in application of extra-planetary exploration. Lastly, acknowledging my limited time and resources during the development of this research, I will not be addressing the time delay inherent to many telepresence systems, particularly in extra-planetary exploration.

Overview and Organization

The remainder of this thesis is structured as follows: chapter 2, ‘Literature Review’, outlines and analyzes the theories and developments around interfaces and control systems in the context of my thesis. These are organized around the themes of control and embodiment, metaphor, and feedback and affordance as have or can be applied to robotics, virtual reality, and analogous technologies. Chapter 3, ‘Methodology and Research Design’, describes and justifies the methodologies, supporting research methods, and the research process used in this thesis. Chapter 4, ‘Needs Assessment and Ideation’, describes my development process, from interviewing, job shadowing, to initial prototype ideation. Chapter 5, ‘Development’, discusses my prototyping process, testing, iteration, and subsequent insights. Chapter 5, ‘Conclusion’, reflects upon the outcomes of this project and identifies possible future directions.

Chapter 2: Literature Review

An examination of the literature has revealed three key themes around rovers and methods for their control. The first regards methods of controlling rovers and other robotic devices and how embodiment and natural interaction can improve these control methods. This section indicates that providing embodiment in the rover's environment can significantly improve navigation abilities for operators. It also emphasizes that natural gestures can improve input fidelity and operational outcomes in manipulation tasks. Secondly, the literatures speaks to the challenges of comprehension with traditional 2D interfaces, how metaphor has traditionally been used to overcome these challenges, and how 3D interfaces can provide additional support for providing meaning. The last theme is feedback, affordance, and the mechanisms by which possible actions and active processes are communicated to rover and robot operators. This section reveals how traditional control systems can burden operators with excessive amounts of data and information and the mechanisms by which it can be simplified using better representation.

Control and Embodiment

Rover operation is a complex, cognitively demanding task. When a rover is deployed in a new environment operators can face a host of factors such as low visibility, disorientation, and obstacles that pose challenges to the rover's safety. These include difficulties in judging pass-ability, damaging or getting the rover stuck, and a high cognitive load due to spatial transformations. Any one of these factors can result in the loss of expensive equipment or slowing down time-critical work (Guizzo, 2012; Helton et al, 2003).

As a result, a great deal of research has been undertaken to improve rover control systems and operational performance. Traditionally, such work has focussed on improving methods for operators and scientists to visualize the rover's environment. This has involved a number of technologies, ranging from the VEVI's panospheric camera, which enabled the capturing of 360 degree-panoramas, and the Pathfinder's stereo pipeline's abilities to generate 3D terrain models from stereoscopic images. As apposed to traditional imaging systems, these provided operators and scientists the ability to see continuously around the rover, allowing a more natural sense of position, contributing to situational awareness of the rover's environment (Nguyen et al., 2001; Sanguino et al., 2012; Stoker et al., 1999).

When the rover is also situated in the panorama or 3D model, such systems can be particularly useful for maneuvering. They do so by assisting in rapidly building up an exocentric view of the rover in its context, circumventing the traditional process of egocentric to exocentric conversion. This is important, for it is exocentric information (survey knowledge) that we mentally navigate by (see figure 2.1). When moving through an environment for the first time we rely on our own vantage point, which provides egocentric information, to maneuver. This egocentric information is made up of knowledge of our

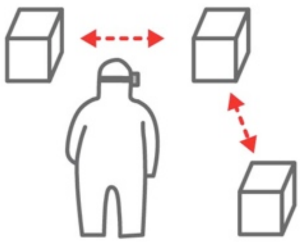


Figure 2.1: Exocentric mental model

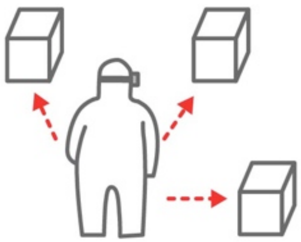


Figure 2.2: Egocentric mental model

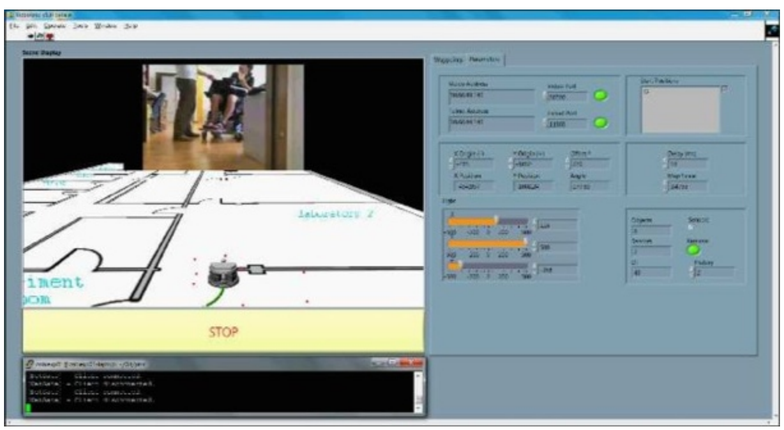


Figure 2.3: Hybrid map and camera feed display (Sanguino et al., 2012)

own position in relation to landmarks and the routes between (see figure 2.2). As we become more familiar with an environment we gradually generalize that egocentric knowledge into exocentric knowledge (Bowman et al., 2004).

Combining both egocentric and exocentric viewpoints, Sanguino et al. (2012), focused on improving the navigational abilities of rover operators through the consolidation of map, camera and rover sensor data into one virtual display. This consisted of a real-time video feed from the rover atop a map of its environment which was overlaid with proximity sensor information (see figure 2.3). Unlike camera only solutions, which have several limitations, including a limited field of view and lack of information from the rover's sensors, this aggregated approach to data representation improved orientation and maneuvering abilities of the operator, allowing for navigation through otherwise complex environments (Sanguino et al., 2012).

This setup required the existence of accurate maps of the environment to exist before the mission. Yet, the ability to rapidly acquire and assess knowledge about an environment where none exists is critical to mission success in hostile environments were preliminary reconnaissance and mission planning is impossible. Such is the case in extra Earth rover operations where high resolution orbital images can only provide data in the order of meters. In these cases operators must rely on data collected as the mission unfolds to form an accurate understanding of the site (Nguyen et al., 2001).

To enhance this ability for operators to assess environments in-situ, researchers and practitioners have looked to the use of immersive VR (virtual reality) environments (Bowman et al., 2004; Bowman et al., 2012). Though the advantages derived through the use of virtual systems in the operation of remote rovers have not yet been fully researched (Sanguino et al., 2012), initial studies into its potential have found that even simple tracking of an operator's head movements, moving their viewpoint in the VR environment accordingly, produced greater comprehension of the source material. These systems also afforded faster and more precise camera repositioning by the operator (Bowman et al., 2004; Bowman et al., 2012).

Enabling interaction with this data in a more natural fashion can further facilitate the comprehension of the imagery and data sets. Studies into the effects of increasing natural, 3D interactions, have shown positive gains in performance and user experience. By contrast, it can be detrimental to performance in cases where natural interactions are only moderately afforded to users (Bowman et al., 2012). This can partly be attributed to our perception of an environment and sense of presence being directly tied to our perception of the quality of interaction afforded in it. “Manipulation is one of the most fundamental tasks for both physical and virtual environments” (Bowman et al., 2004).

Furthermore, according to Mantovani and Riva (1999), in order to be present or embodied in a given environment, rather than requiring a high quality visualization, we must be afforded freedoms of agency and movement. Presence is always mediated by both physical and conceptual tools. Reality is not out in the world but is co-constructed in the relationship between us and our environment; the human mind is actively creating a reality to embody. As such, the sense of presence in a VR environment, like improved mechanisms for visualization and affordance of interaction, has a positive effect on spatial knowledge acquisition and usage (Bowman et al., 2004).

While a traditional control system involving a three-button mouse was used to navigate around the 3D model generated by Pathfinder's stereo pipeline (Stoker et al., 1999), researchers have looked to other technologies to enable natural interactions with an eye to allow unprecedented levels of interaction fidelity and enhanced performance. For example, the research findings of Bowman et al. (2012), indicate that increased interaction fidelity has a positive affect on difficult 3D tasks such as manipulating both position and orientation at the same time.

As the aforementioned has illustrated, natural gestures and travel techniques result in more precise control as well as improved spatial understanding. These gains are attributed to natural gestures affording intuitive control over all 6 degrees of movement and rotation at once in 3D space. In addition, these methods of movement provide proprioceptive cues, such as leaning, turning, and walking (Bowman et al., 2004; Bowman et al., 2012; Chance, 1998; Ware & Franck, 1996). That said, gestural control can be exhausting after long durations. Dan Saffer notes, “Human beings aren't meant to hold their arms out in front of their bodies making gestures for long periods of time. It creates a condition called Gorilla Arm (aching muscles, stiffness, a swollen feeling) because it violates basic human ergonomics” (Saffer, 2011, location 690 of 730). In his opinion, rather than act as a goal to strive towards or a solution to all our interface problems, gestural interfaces should act as a indicator of how little of the body we use when interacting with our current devices and how they could be more engaging and physical.

Rather than being used in all situations, gestural controls can be employed when other methods of input are impossible. This is the case in complex medical operations. To help plan and carry out these operations, surgeons must navigate through CT and MIR (O'Hara, 2014). Yet, once prepped and ready for surgery a strict boundary must be maintained between what is sterile and what is not. As a result, direct interaction with a keyboard and mouse are not an option, as is rescrubbing or removing gloves. Working with these constraints, surgeons at the Sunnybrook Hospital in Toronto have turned to Kinect gesture tracking to navigate through MRI and CT images (O'Hara, 2014) (see figure 2.4).

Figure 2.4: Operating room image navigation through gesture tracking (O'Hara, 2014)

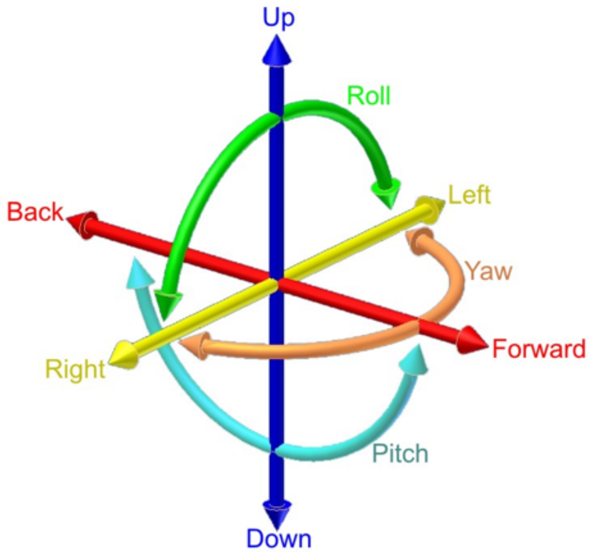


Figure 2.5: six DOF
(Six degrees of freedom, n.d.)

There are also cases in which full body gestural controls are the best option among many. In their comparative study into the benefits of natural (gestural) control systems, Bowman et al (2012), asked users to perform a task of selecting, repositioning, and rotating 3D letter-shaped objects until they were aligned with similar shapes. The more natural gestural techniques allowed for simultaneous control over all six degrees of freedom (DOF), that is: forward and back, left and right, up and down, roll, yaw, and pitch (see figure 2.5). As such, they significantly outperformed the less natural technique, which utilized traditional keyboard an mouse manipulations and limit control to one DOF at a time.

By contrast, gestural control can be less efficient than traditional methods in some instances. In the same study, Bowman et al (2012), examined a video game, Mario Kart Wii, that offered multiple interaction techniques. In this game the primary task is steering a vehicle around a race track. After comparing driving performance across four different steering techniques: two traditional techniques based on gamepad input and two natural techniques utilizing gestural control akin to using a real steering wheel, it was seen that the less natural techniques were faster and more accurate. Among the many reasons the author speculates could be attributed to this, include the possibility that small muscles such as those in the hand can be faster and more precise than the large muscles used to turn a steering wheel. Some actions can also be a poor match for natural gestures, such as teleportation through a VE (virtual environment). As a result, rather than gestural input providing a more intuitive or natural method of discovery, these actions are often mapped by designers to button presses, arbitrary gestures, or abstract interface metaphors (Bowman et al., 2012).

In these studies, it is evident that embodiment and natural gestural input can contribute significantly to control methods for rovers, robots, and similar complex interfaces. By gaining an understanding of the way in which these techniques can and should be employed, this research provides the basis for the development of more tailored, precise, and situationally aware rover control systems.

Metaphor and 3D Interfaces

Metaphor, as described by Lakoff and Johnson (1980) in their book "The Metaphors We Live By", is a fundamental mechanism of mind, one that allows us to use what we know about our physical and social experience to provide understanding of countless other subjects. From the earliest days of the modern desktop computer revolution, functions and capabilities of computing systems have been communicated through symbolic representations of interface elements (metaphor) (Blackwell, 2006).

Yet, despite the implementation of metaphor in computer interfaces, problems of communicability were seen to be prevalent even in the early days of the desktop computer's development; tests conducted to establish the benefits of using the Apple Lisa computer for office professionals found that most users experienced confusion and frustration (Blackwell, 2006). Windows, icons, and direct manipulation of on screen elements all proved frustrating to a sample of intended target users (Carroll and Mazur 1986, as cited in Blackwell 2006, p. 498). Rather than a comparative model of metaphor, "understanding and experiencing one kind of thing in terms of another" (Lakoff & Johnson, 1980, p. 4), as the designers of the Apple Lisa had intended, users were constructing new meaning out of the differences between literal and figurative interpretations, as apposed to their similarities. This is likely because when one thing is described in terms of another, a third thing has been created: the relationship between the two (Blackwell, 2006).

For this reason, systems that rely solely on metaphor for training can be problematic. In their 2005 paper, Syslo & Kwiatkowska stress that students are unable to develop mental models and the set of cognitive processes required to properly leverage and navigate digital technologies directly from the software alone. Blackwell (2006) describes how, at the 2003 workshop he attended, interface metaphors were described as a visual communication channel via which the designer achieves the rapid transfer of an effective mental model into the user's head. This description implies that users are passive recipients of new understanding. Active, subjective, contextualized, or embodied interpretations by users of interface metaphor are problematic because users might discover or construct new interpretations, rather than understanding the metaphor and receiving the expected conceptual model.

In addition to discovering new meaning, creating interface metaphors with the expectation for users to function as passive recipients of an intended meaning can prove problematic when the objects or cultural practices being referenced by a metaphor are unfamiliar to the user. This was clear at least as far back as the development of the Xerox Star, the pseudo precursor to the Apple Lisa. In the development of the Xerox Star, icons comparing disk storage on a computer to a filing cabinet would confuse users rather than inform them of the disk storage's function. As Lakoff and Johnson (1980) notes, "metaphors are rooted in physical and cultural experience; they are not randomly assigned. A metaphor can serve as a vehicle for understanding a concept only by virtue of its experiential basis" (p.18). In this instance, users lacking any exposure to the physical object being referenced (the filing cabinet) will experience confusion.

Conversely, VR interfaces provide a common experience around which to frame the interfaces, the 3D environment we already inhabit (Shedroff & Noessel, 2012). This provides more information, such as depth, landmarks, and connectivity, than is available in 2D interfaces (see figures 2.6 - 2.8). Such information can support traditional metaphors or provide a fallback reference system when, as previously mentioned, a user lacks exposure to the cultural practice or physical object being referenced. According to Ark et al. (1998) "The more redundant dimensions available to the user, the greater the chance of the user being able to choose an attribute to which to relate." (p. 209).

Research has also shown that the methods of operation in VR environments can be learned much easier than traditional systems, thereby enabling faster training. (Bowman et al., 2012) This is because VR environments

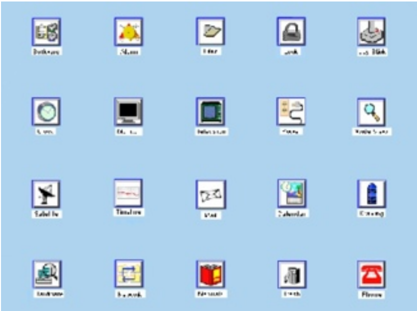


Figure 2.6: 2D icon layout (Ark et al., 1998)

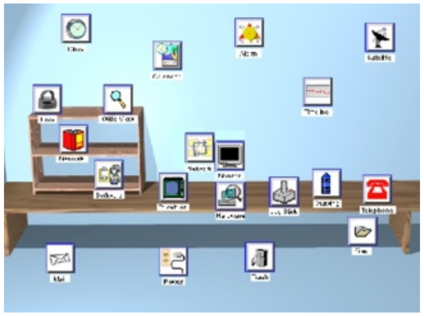


Figure 2.7: 2D/3D hybrid icon layout (Ark et al., 1998)

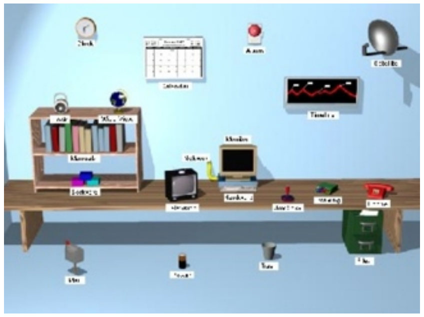


Figure 2.8: 3D icon layout (Ark et al., 1998)

conform more naturally to our existing schemas, as we are already accustomed to operating in a 3D environment (Shedroff & Noessel, 2012). As Shedroff and Noessel puts it, “new interfaces are most understandable when they build on what users already know. If an interface is too foreign, it’s easy for users to get lost trying to understand what the interface is or how it works” (Shedroff & Noessel, 2012, location 18 of 348). Similarly, “If we already know something about a topic, then learning new information is easier. If the new information conflicts with what we know, then learning can be harder.” (Errey et al., 2006)

In addition, our learning, or memory, also functions better when associated to physical locations or places (Yates, 2011). We draw important clues from our context and our interactions with it (Ark et al., 1998): The layout of objects in a 3D environment form connectivities that make physical sense.

As illustrated by the literature, metaphor is a mechanism for communicating meaning or understanding of an interface element's function. This method of communication is susceptible to misinterpretation. Therefor additional communication channels should be introduced to reinforce these metaphors and meanings. As noted by Ark et al. (1998) 3D interfaces can provide additional information than what is possible on traditional 2D interfaces with metaphor alone. Additional benefits can also be seen from including 3D in interfaces, as the cited literature also provides evidence that 3D in interfaces can improve learning and speed training.

Feedback and Affordance

Norman (2002) states that it is a lack of visibility that makes any computer system difficult to operate. This can be overcome by providing a user a good conceptual model of how a system works, the possible actions that can be taken, and a feedback system to indicate what is happening in the machine at any given moment. As such, in order to ensure the ease of operation of VR environments designers are tasked with creating feedback and affordances to indicate what is happening in the system at any given moment.

Affordance, coined by JJ Gibson, was originally intended to describe the range of activities that could be carried out on an object. Whether they were perceived or not was irrelevant. Don Norman argues that affordances must be discoverable to be useful and truly afforded to the user. Under Norman's definition of affordance VR environments are much more useful than traditional rover control systems as they communicate affordances in a more obvious fashion, through natural signalling (Norman, 2007; Shedroff & Noessel, 2012).

Beyond communicating information such as active processes as noted by Norman, researchers and practitioners have observed that the inclusion of feedback mechanisms in VR environments, such as haptics and sound, can significantly improve a user's sense of presence. In their mixed reality system, test subjects asserted that their sense of presence was improved by the inclusion of tactile feedback (Borst & Volz, 2005). Similarly, Bowman et al. (2004) notes that the inclusion of three-dimensional sound in VR environments can add important depth cues which assist in localization and way finding through the environment.

These feedback mechanisms allow for tacit and implicit signalling. Interpreting these signals does not require specific training or learning, it simply exploits existing perceptual patterns and their recognition. They are particularly important in the design of interfaces because they inform without interrupting or need for conscious attention (Norman, 2007). Through these tacit and implicate signalling 3D VR environments can better communicate affordances to operators: can the rover fit, can the arm move, will I tip the rover if i start to drive up that incline. Data signalling whether these things are possible is not easily communicated through traditional rover control interfaces.

It is often up to the operator to remember discrete data such as rover direction, location, tilt, and state of operation, then combined and compared the data to form a judgement call on the next course of action (Errey et al., 2006). This can be a very cognitively demanding task as the number of elements imposed on working memory is a major contributor to cognitive load. Furthermore, in cases where cognitive demand is high working memory will be reduced, making learning of the information being conveyed more difficult (Errey et al., 2006). As such, interfaces should strive to combine and present data in aggregated or natural methods.

In their paper, Lapointe & Massicotte (2003), discusses the control system for working with the Canada Arm aboard the International Space Station (ISS). This involves operators toggling between camera and arm control (see figure 2.9), ensuring that a good camera is selected for the task at hand. This requires the operators remember the locations of all the cameras aboard the ISS. As, Lapointe & Massicotte (2003) points out, “the building of such a cognitive model (mental map) of the system requires a lot of training for the operators and is prone to errors”. As an alternative, less cognitively demanding control system Lapointe & Massicotte (2003) proposed a VR system which simultaneously displays the ISS and Canada Arm, along with the positions, orientation, and field of view of all the cameras (see figure 2.10) This eliminates the need for operators to remember the location and states of all the cameras, also reducing training time.

As the previous example illustrates, there are measurable gains to be found from employing virtual environments for control systems. Yet, transitioning

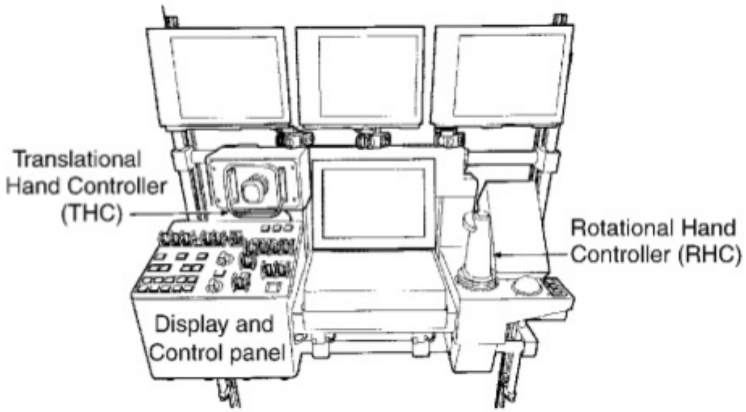


Figure 2.9: ISS robot workstation (Lapointe & Massicotte, 2003)

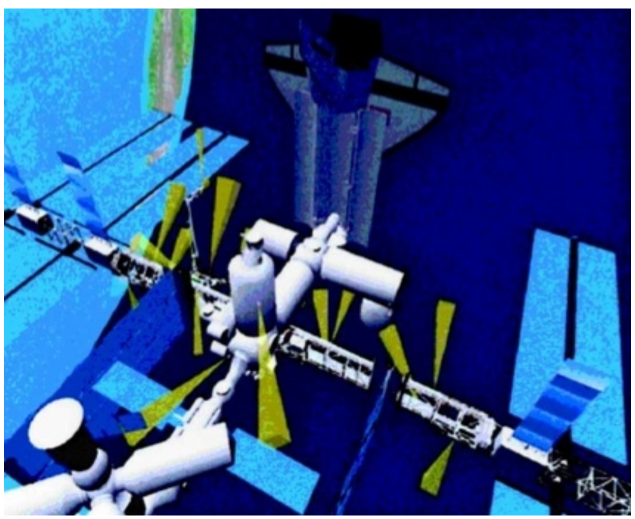


Figure 2.10: ISS 3D robot control overview (Lapointe & Massicotte, 2003)

traditional control systems or any other software over to this approach is not simply a matter of transferring conventional 2D interaction styles into a 3D environment. "In immersive VEs, users have to deal with 6 DOF (degrees of freedom) input as apposed to 2 DOF" (Bowman et al., 2004, location 4468 of 9592). This can lead to challenges with discovering and selecting crucial interface elements. As such, researchers have taken to adapting existing 2D metaphors, which revolve around GUI (graphical user interfaces) and WIMP (windows, icons, menus, and pointers), to 3D VR environments. One such adaptation involves the mapping of these interface elements to a physical surface. In this way the menu system is accessed simply by bringing the surface into view. Furthermore, the physical object upon which the menu is mapped allows for accurate selection of menu elements in an otherwise arbitrary 3D space (Bowman et al., 2004).

While useful, navigation through such systems requires the user's full attention. By contrast, body referenced menus, attached to the hand, head, or chest, provide for spatial associations to be formed with not only the environment but the interface's menus and windows as well (Bowman et al., 2004). Shedroff & Noessel note this is the case because, "with physical objects, it's common for people to remember where things are spatially, using the surroundings as reference" (Shedroff & Noessel, 2012, location 61 of 348). This is only the case when spatial arrangement is consistent, as is the case when menus are associated with the user's body.

As discussed in this and previous sections of the literature review, 3D and VR environments provide methods for imparting feedback and improving the communication of information to operators. These provide tacit signaling and communicate affordances, such as potential commands for rovers or similar robots, in a more obvious fashion to operators than is typically possible with traditional 2D control systems. Moreover, as noted in the metaphor and 3D interfaces section, adding 3D interfaces can provide additional benefits above and beyond one single purpose: 3D cues reinforce meaning and, as discussed in both the control and feedback sections, improve situational awareness of the rover or robot's context.

Chapter 3: Methodology and Research Design

In the previous chapter, the many benefits of 3D and VR interfaces as well as natural interaction were discussed. These include improved situational awareness, precision, and comprehension of complex data sets. As such, with the goal of improving rover control systems, this project involved the creation of a number of robot control prototypes employing VR interfaces and natural interactions. In doing so, I employed a comparative prototype research methodology. This involved picking a design element to vary and building prototypes to embody multiple design alternatives around that design element. Physical controls could be replaced with gestures but I had to compare it to a physical control interface. This goes beyond designing for use, rather, it promotes designing for evaluation. This methodology produces observations across situations of use that helped me think about rover control design in a new light.

To understand the data derived from my comparative prototype research methodology I used framework analysis this is because framework analysis, which assists in comparing themes across many cases, also understands each case in it's own context. Essentially, data is grouped according to theme and by interview, into a thematic matrix. It is suited for qualitative research like this, which has specific questions, and a pre-defined sample. Though I looked for specific areas that I identified from preliminary studies, this method allowed me to be open to unexpected themes that emerged from participant experiences.

In addition to these overarching methodologies I used research methods such as user-centred design and participatory design. User-centred design helped my prototypes respond to the needs and wants of robot operators as found in my initial interviews. The assumptions which contributed to those prototypes were then tested and optimized around how they were actually used. This testing process also leveraged participatory design methods, involving test subjects in the process of creating new prototypes. This ensured these prototypes met the needs of tester’s while also providing me new, yet unheard insights.

Using these methodologies I undertook the following research process, which consisted of three steps:

1. Interviews with rover operators
2. Observation of an active rover mission
3. Development and testing of prototype rover control systems by novice users

The first step focused on needs assessment and deriving insights into challenges with rover operating as asserted by operators themselves. Conducted one on one, these interviews involve leading questions about current operational practices and difficulties. Interviewees were also questioned about possible solutions to their own challenges with current rover controls. In this step I interviewed three individuals that actively engage in rover operation in their jobs.

The second step focused on the direct process and experience of controlling rovers. In this step I observed an active rover mission. This facilitated insight gains above and beyond those derived from the interviews. Furthermore, this step allowed me to gain first hand experience of a rover control interface in a complex mission setup.

The third step focused on developing and testing prototype control systems intended to address the issues identified in steps one and two. These were radically different from traditional control systems, as such, I tested them with novice users, possessing no experience with rover operation. This ensured each prototype was evaluated by the testers solely on its own performance and characteristics rather than existing systems. During three test sessions, research participants were asked to test a number of robot interface prototypes, then provide feedback on their experiences with each during test debriefing interviews. This feedback guided development by identifying opportunities and challenges as well as assisting in the creation of experience narratives and further insights that fed back into the creation of future interface prototypes.

Chapter 4: Needs Assessment and Ideation

In this chapter I review my process of rover operator needs assessment and ideation of a VR interface for rover control. In setting out on this project, possessing little knowledge about rover operation, I began by researching the process of rover operation and identifying needs and areas for improvement for rover operators. This began with a study of the Fukushima nuclear disaster cleanup effort as portrayed by one rover operator. I then interviewed and performed a job shadow of a number of operators. The insights gained from this study, the interviews, and job shadow, in addition to the findings from the literature review, contributed to the ideation of an initial VR interface for rover operation. This interface was meant to address issues with current control systems using VR and natural interaction. This led to an evaluation of current consumer and DIY (do it yourself) technologies which could be utilized to create the prototype.

Needs Assessment

My process of needs assessment for rover operators involved investigations into current challenges these individuals face through a literature review as outlined in the previous chapters and also an informal study of the operations being undertaken by rover operators in remediation and cleanup after the Fukushima nuclear disaster. Much of this involved reading first hand accounts on the blog of one such operator.

The challenges and issues this rover operator faced ranged from poor visibility when in radiation gear, difficulty with complex terrain (such as stairs)(figure 4.1), and difficulty with simultaneous operations on the existing control system (a dual joystick handheld controller)(figure 4.2)(Guizzo, 2012). Subsequent to these initial inquiry, I possessed enough knowledge regarding rover operation to formulate further, more targeted, explorations through firsthand observations and interviews.

Figure 4.1: Rover navigation up stairs (Guizzo, 2012)

Figure 4.2: Fukushima rover control system (Guizzo, 2012)

Figure 4.3: ODG mobile platform (rover)

Figure 4.4: ODG handheld controller

This came through a connection with Ontario Drive & Gear Limited (ODG), a business that I had become familiar with during my undergraduate degree at the University of Waterloo. Best known for manufacturing of the ARGO, an amphibious, multi-purpose, all terrain vehicle, ODG also manufactures platforms for various robotic vehicles including lunar rover prototypes for the Canadian Space Agency (CSA) and the National Space Agency (NASA) (figure 4.3). While engineers first, with a collective 11 years at ODG, these individuals are also tasked with testing new rover designs out in the field, possessing 8 years of relevant experience.

My research with ODG involved conducting interviews with members of Ontario Drive and Gear's rover engineering team to discover pain points with current control systems and their ideal solutions to those problems. It also involved performing observations of a rover testing exercise with members of that same team for the Canadian Space Agency.

While my initial research had drawn me to focus primarily on using VR to improve rover control methods, my interviews with ODG engineers made it clear that, while not perfect, their current method of rover control is mostly adequate for their current work. What is truly needed by these operators is a better way to visualize the world around the robot as well as the data coming from it.

For issuing rover command ODG employs a handheld control system using two joysticks, one controls the operation of the rover's right tires, the other controls the left tires (figure 4.4). In this system pushing two joysticks forward commands the rover forward, likewise, one forward one back commands a turn in place. The ODG engineers noted that this setup is very simple and intuitive for most rover operators. Yet, performing multiple control operations at a time is a challenge with handheld controllers (Guizzo, 2012). Such operations include setting rover speed, direction, and pitch, all of which require continuous modification during operation. To provide such functionality, in ODG's setup the pitch of the rover (set using active suspension) is mapped to the X axis of the same joysticks used for tire speed control. This setup can pose problems with accidental triggering of functions. Particularly when operating at high speeds.

In interviews it was expressed that rover operation can be easy initially as long as the rover isn't reversed to their own point of view. It is this reversal of direction, driving at oneself, that is the hardest challenge for operators. To overcome this issue they often use a technique of flipping the control around in their hands, matching the controller to the orientation of the rover and thereby returning right and left joystick input to correct relative directions. It was noted during an interview that while helpful, this technique is only used in tight spots and only within the first hour of rover operation. This is perhaps because it takes a period of time for rover operators to adapt their cognitive model to that of the rover. However, the challenge posed by the reversal of direction is only applicable to direct line of sight operation, where the operator's perception of left and right are decoupled from that of the rover.

While most of their tests are conducted via direct line of sight control, ODG engineers have had some experience with camera controls. In the case of a camera feed control test, the ODG engineers insisted that such setups have to be very well calibrated in order to prevent distortion of the camera feed. Even when fully optimized though, operating via camera does not provide enough contextual information for the robot, lacking resolution and depth perception. As such, operators typically prepare before missions or tests by studying topography maps from satellite or surveys.

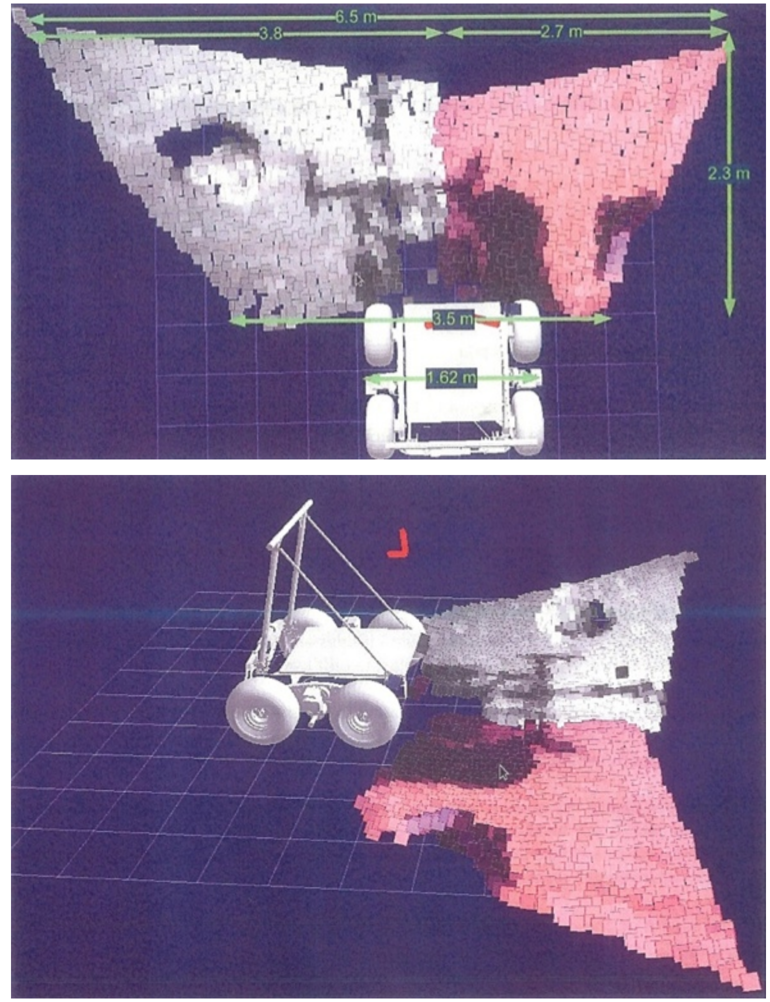
The optimum condition for ODG engineers is standing beside the rover, this lets them assess the landscape to a greater degree. This is extremely important in driving a rover as there are a lot of judgement calls to make; will a rock roll away when driven over or will a maneuver create a rock slide? As such, the process for driving involves progressing a small bit, analyzing the terrain, then repeating.

These observations suggest that what rover operators most need is a method to relate the terrain to the rover. That could be a side on view as apposed to a view from an onboard camera. Such a view would communicate a greater deal of information regarding rover mobility through the terrain.

As previously mentioned, in addition to conducting interviews with ODG engineers, I was able to observe two of those same engineers drive a rover of their own design during testing of a rover control interface for the Canadian Space Agency (CSA). These tests were conducted at the CSA offices in Saint-Hubert, Quebec, while controlling a rover in a sandpit 40km away. This ensured that test operators had only been able to see the test course through the rover's onboard camera.

In this test the CSA wanted to discover how rover operators would use three methods of controlling the rover, how they would use the information available to them, and in what situations. The three control methods included: a gamepad with two joysticks, setting destination waypoints by clicking points on a video feed, and keying in the angle and distance of travel into a menu. In addition, operators were provided with a number of visualization tools, including topography maps, onboard cameras, panorama creation tools, rover tilt readouts, and a method for generating 3D models of the area directly in front of the rover extending out approximately two and a half meters (figures 4.5).

To perform the test rover operators were tasked with four exercises encompassing navigation, obstacle avoidance, and environment assessment. In the test the CSA used a two operator system, one operator drove, while the other navigated and controlled the main camera (figure 4.6). To work effectively it was necessary for the operators to give orders to each other for commands such as camera adjustment or panorama generation. This created a fallback system wherein operators second



Figures 4.5 - Example of the CSA's rover 3D imaging tool.
Image Credit: Canadian Space Agency

guessed and ran decisions past each other. This system wasn't without its flaws though, communication was not always adequate, therefor operators were frequently surprised by actions taken by their co-pilot.

Regarding communication, one tool that was particularly problematic was the zoom function. It often resulted in confusion or fear that the rover was dangerously close to obstacles. This is likely due to a shift in the camera's function (zoom) without communicating a shift in the perceptual context (control panel) of the operator. Everything remained the same on the control panel save for the camera display; for all intents and purposes it appeared as though the rover head moved forward by a number of feet. The operator in charge of camera control should have communicated his intent to zoom the camera but, just as importantly, the UI of the control panel should have communicated the zoom. This communication breakdown undermined trust in the tool. As a result, the operators became less inclined to use it.

Another tool that suffered from under-communication was the 3D model generation function. The operators frequently cited a need for more information from the 3D data such as measurements and angles. In this regard the 3D model differed from the rest of the control interface which tended to provide this data; the main camera had a overlay depicting distances away from the rover in meters. These hard numbers proved helpful when keying in distances into the rover's keyboard control method.

The primary issue for the CSA's test control interface was that it required testers to look between a number of displays and input devices to perform tasks. The rover tilt readout was on a secondary display from the main camera view. Some of these displays were extremely far away from the main area of focus, over one meter in many cases (figure 4.7). This was perhaps because of the non-specific setup of the test room, which likely primarily functions as an office or lab, being reserve to testing on rare occasions. To provide an optimal control setup these screens should have been closer together or positioned with some thought towards their use. Secondary to the issue of placement was clarity and saturation of iconography. Having both “break on” and “break off” beside each other didn’t make it clear which state was active. In the opinion of one ODG operator, one button should have toggled instead. Similarly, many readouts were duplicated across displays.

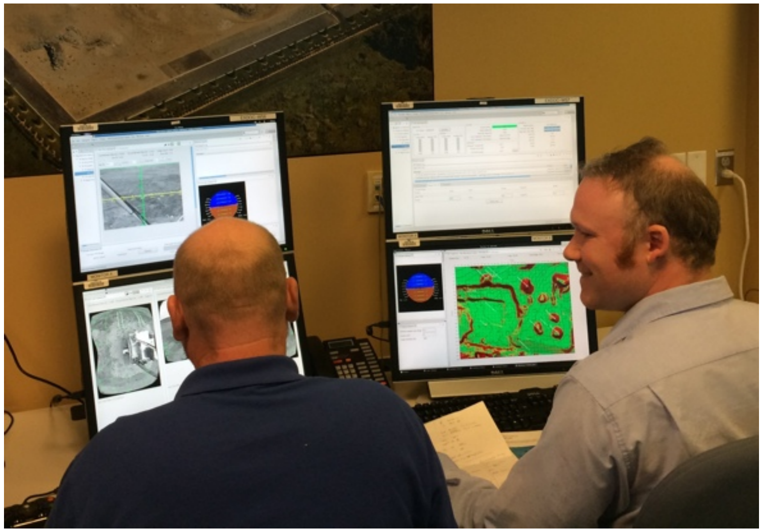


Figure 4.6: Two ODG rover operators working in tandem

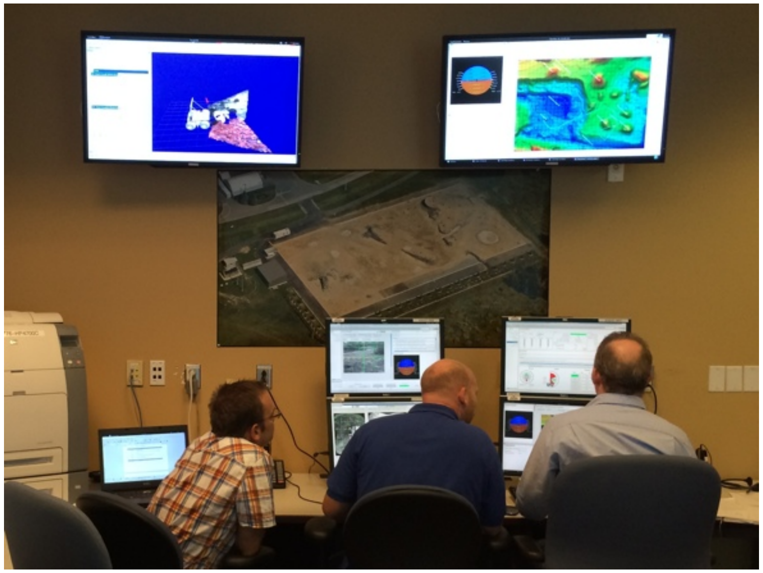


Figure 4.7: CSA test setup

Summary of insights

After conducting a first round of interviews of rover operators, initial literature review, and field observations, it is clear that rover operators have a number of needs, likes, and dislikes for current rover control devices. Primary among these is difficulty in visualizing and understanding of the data collected by the rover's onboard cameras and sensors. Being aware of the rover's tilt and distance from a hazard, or judging its ability to navigate a particular terrain is essential but very challenging.

Secondarily, operators have difficulty with the traditional mapping of multiple commands to one joystick. Tapping forward and back on the two joysticks with current controls set wheel speed for that associated side of the rover. These same joysticks, when tapped left or right, effect the active suspension of the rover, tilting it rather than driving.

Third, operator's expressed a universal dislike of video game controls used as a rover controller. Game controllers lack the resolution of control that high end and dedicated devices have. This is particularly important when navigating tight or obstacle filled terrain, in which fine adjustments are required.

In addition to these insights, there are number of rover control interface design best practices I learned from my observations of the CSA's test. First, it is clear that I must provide relevant data to the correct contextual locations or, better still, to the area of the user's attention when required. In the CSA's test the 3D model should have been positioned beside the rover's main camera feed in order to facilitate comparison between the feed and the model. Second, it is necessary to ensure communication is facilitated between the user and the interface. Active functions should be clearly indicated. Likewise, the action taken by a function should be clearly communicated.

Initial Solution Ideation

As illuminated through my previous literature analysis, interviews and observations of rover operators, the most pressing issue faced by rover operators is that of visualization of the rover's context. This can frequently lead to confusion or error in issuing commands. As such, my initial design was focused on visualization.

At this ideation stage, rather than involve rover operators, I chose to develop my initial concepts without user consultation. As Baker (2009) notes, while user feedback is helpful in checking the assumptions being made, it is less useful for entirely new types of products. This is because it requires users to rely on their imagination to understand the design (Baker, 2009).

As an initial foray into a solution to the problem of visualizing the rover's context I intended to create an interface which would remove the barrier of abstraction imposed by traditional control and visualization devices. The core of this interface would have relied on creating virtual worlds in which both the operator and rover cohabitate, thus negating traditional interfaces (keyboard, mouse, screen) altogether. It was through this radical re-contextualization of rover control systems that I hoped to create a rover interface of the future. Similar efforts are underway in the US NAVY, by casting aside limitations of current built technology and working nearly entirely in VR they intend to envision the command and control centre for battleships 15 years in the future (Hollister, 2004).

In this first prototype, the human body (that of the operator) is the scale by which the virtual world will be measured and understood.. The primary intent of this system was to bring implicit communication and signalling into an the otherwise complicated process of controlling a remote robot; we know how big we are innately. As Don Norman points out in 'The Design of Future Things', traditional systems, lacking implicit communication channels, fail or frustrate; users must read and check in on every setting and measurement and can quickly become overwhelmed (Norman, 2007).

The control system operates through a multitude of elements feeding into one central digital experience (see figures 4.8 - 4.13). As apposed to traditional rover control systems that limit the operator's perception of the rover's environment to what is visible through its onboard cameras, this VR control system creates a complete virtual representation of the rover's surroundings. This would help rover operators quickly overcome their unfamiliarity with the robot and its context, as well as encouraging the formation of spatial memory, thereby improving recall of interface and control locations (Bowman et al., 2004; Lapointe & Massicotte, 2003; Sanguino et al., 2012).

In addition to allowing for easier visualization of the rover's context, this interface would also allow for control of the rover through multimodal inputs such as natural body gestures, ideally allowing for more intuitive command execution in complex missions ((Bowman et al., 2004; Bowman et al, 2012).

Figure 4.8: Rover maps its context via depth cameras

Figure 4.9: A 3D terrain model of the rover's context is generated from the data provided by the depth cameras

Figure 4.10: A 3D model reflecting the state of the rover is placed in the 3D terrain model

Figure 4.11: The operator's gestures are tracked via depth camera

Figure 4.12: The operator experiences the 3D model of the rover's context via a VR headset

Figure 4.13: Combined gesture tracking and VR headset enable natural interaction with the rover

Chapter 5

DIY Components Assessment

Though ultimately virtual, my initial design for the rover control environment would be created and interacted with through the use of a number of hardware and software components feeding in to one central experience. As such, ensuring I could build this system required an extensive study of the current solutions available and their ability to interconnect and communicate.

Though discrete in this document, this section formed a symbiotic relationship with my initial prototype ideation, leading to much research into hardware and software capabilities of current solutions offered in the market and by the DIY (do it yourself) community. It was at this point I selected the Kinect and the Oculus Rift VR headset as the best components to use in implementing my theory, both because of their price and availability but also because of the readily available resources for working with them.

For clarity, I will first outline the technologies I intended to use in the first prototype and how they interact together. I will then describe the nuances of the major components and why they were chosen.

1. Real time 3D scanner > Microsoft Kinect
2. Virtual reality headset > Oculus Rift
3. Interactive 3D environment > Unity game engine

In this initial prototype concept the rover itself is equipped with a Kinect which is used to create the 3D virtual reality simulation of the world around the robot. The Kinect is also used to track the motion of the operator in the remote control centre. The Unity video game engine is used to present the 3D model of the world around the robot to the operator and the Oculus Rift is used to immerse the operator visually into the simulation as though he or she were there in person.

Figure 4.14: Oculus Rift VR Headset

The Oculus Rift

The Oculus Rift is an advanced virtual reality headset with immersive, 100 degree field of view, stereoscopic 3D viewer. This viewer presents a unique images for each eye, reproduces the way human eyes perceive images in the real world, creating a much more natural experience.

The Oculus Rift also has built in head tracking. This is extremely significant in improving immersion in virtual environments because it removes gimbal lock, one of the biggest issues with first person perspective in virtual environments such as modern first person shooter (FPS) computer games and simulations software. Gimbal lock directly relates to the limitation of the user's 3D orientation and perspective, represented by rotations on X, Y, and Z axes, commonly referred to as pitch, yaw, and roll. In theory the user is free to look around freely while using a mouse and keyboard or dual joystick gamepad for control. In practice it is much more limited. Using dual joystick input, one joystick controls walking around, and the other controls changes in orientation (vantage point). Pushing left and right on this second joystick will generally rotate the vantage point about the vertical axis (yaw), while pressing up and down on the same joystick will tilt you around the horizontal axis (pitch).

The shortcoming with this control setup is best illustrated by imagining an object moving towards you and passing over your head. Using a two joystick setup you would push up on the second joystick to increase the pitch as it passes over you. Yet, when the object is directly over your head, pushing up doesn't adjust the pitch any more. Up on this joystick simply means "look up", not "pitch toward this direction". In order to continue tracking the object, you would need to rotate around 180 degrees by pressing horizontal on the same stick, and proceed to adjust your pitch downward by pressing down on the joystick.

This is not an issue with the Oculus Rift, the user inherently understands which way is down. As a result, it affords continuous pitch adjustment. As apposed to the previous example, when using the Oculus Rift there would be no need to rotate 180 degrees to continue tracking the object beyond the vertical; pitch can be adjusted continuously. Moreover, using the Oculus Rift, there is no longer need for manual, cognizant controls for yaw, pitch, and roll. The control simply becomes "look at the target". This fundamentally changes embodiment in Virtual Environments.

Though the Oculus Rift greatly improves orientation control, it lacks translational tracking. That is, it doesn't know where your head is in 3D space, only what direction it is looking in. This results in issues while moving through 3D environments. Without translation tracking the wearer of the Oculus Rift is unable to duck or move his head left and right, up and down, forward and backward, without the use of a keyboard or gamepad. This results in an unrealistic feel and detracts from the sense of immersion in the 3D environment. Fortunately, tracking the user's body via the Kinect will add the missing translation tracking and improve the sense of realism of the simulation.

Figure 4.15: Kinect Tripod setup

The Kinect

In addition to the Oculus Rift the Microsoft Kinect is integral to making my VR interface a reality. The Kinect is a motion sensing input device for computers and the Xbox 360 video game console. It uses a built in infrared depth camera to track people and objects in order to create natural (gestural) input for games and software. I had intended to utilize two functions of the Kinect in this project but throughout development and in the final prototype I was limited to one. The first, which I couldn't use, was the Kinect's ability to generate 3D models of the world in front of it using its onboard depth camera. The second was the Kinect's skeleton tracking (body tracking) ability using the same depth camera.

3D model generation, which was to be used to build up a 3D representation of the world around the rover, did not make it into this project. The largest issue with it was, rather than generating a polygon model, which would be accepted by most 3D software, the Kinect generates point cloud data which has to be post processed in order to be useful. This is typically achieved using the KinectFusion algorithm. In addition to not being realtime, these models are traditionally limited in size to two square meters due to the processes undertaken to generate them.

An alternative to KinectFusion is the Kintinuous software, which relies on various odometry estimation techniques to increase the robustness of the Kinect camera tracking in addition to outputting and updating polygon models of the world around the Kinect (Whelan et al, 2012). Unfortunately, KinectFusion is still under development and I was unable to procure a distribution of the software. As a result, over the course of development and testing, the 3D model of the area around the rover consisted of pre-constructed 3D models.

By contrast, skeleton tracking is easily achieved using a number of open source tools. In the case of this project I used Zigfu, an implementation of OpenNI for the Kinect. This was easily integrated with the Unity game engine, the environment I chose to build my 3D interface due to the Oculus Rift readily supporting it. Finding a software environment which readily supported both hardware tools I aimed to employ was extremely important; it allowed me to focus my attention on the issues relevant to improving rover control and iterate my prototype designs quickly. Iteration was key because, as noted previously, the first design had not been tested or subjected to user feedback. This design would go on to change substantially over the course of this project.

Chapter 5: Prototype Development and Testing

Prototype development and testing consisted of a three month period in which designed were constructed, tested, and revised or set aside. Three tests, each with the same six participants, were conducted on six major, eight minor, prototypes. These tests were evaluated using the methods, such as framework analysis, outlined in the research methods chapter. Insights and best practices for VR, 3D, and natural interfaces were then derived and subsequently explained.

Testing and Insight Collection Method

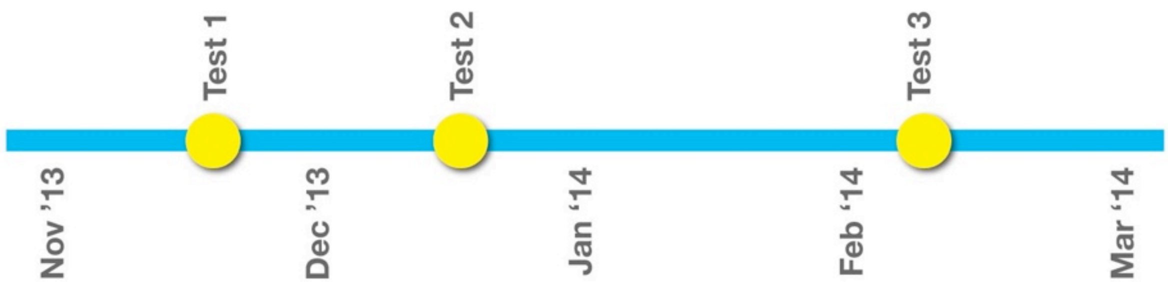


Figure 5.1: Testing timeline

Over a period of three months prototypes were developed and subjected to three test sessions in a large multi room lab at OCAD. Each test session consisted of six individual one hour test slots, these slots were filled by the same six volunteers over the duration of testing. This produced a collected eighteen hours of testing distributed across six unique prototypes. I strove for these particular numbers because, as Jacob Nielsen suggests, it is better to conduct three, five person tests, rather than one fifteen person test. “You want to run multiple tests because the real goal of usability engineering is to improve the design and not just to document its weaknesses. After the first study with 5 users has found 85% of the usability problems, you will want to fix these problems in a redesign” (Nielsen, 2000).

During the one hour test sessions testers were asked to complete a number of predefined exercises with each of the then current prototypes, three in November, two in December, and one (consisting of eight minor prototypes) in January. This was in order to evaluate and understand there characteristics in categories identified and developed with the assistance of the literature review, operator interviews, and field observations. These categories, developed using framework analysis were:

1. affordance
2. consistency
3. communicability
4. feedback loop
5. spatial memory
6. cognitive load
7. simulation sickness
8. preconceived notions

This process of evaluation involved direct supervision and observation of the tests, post analysis of the tests via video recording, and debriefing interviews directly after each participant's test. More than anything, these debriefing interviews coupled with the experience of having just watched the tests, provided the greatest insights and drivers for the development of the interface prototypes over the three months.

In recruiting testers I drew upon family, friends, and peers. The intent was to find individuals with no formal experience controlling remote rovers. This allowed me to work with a blue sky approach. The testers provided me feedback around the experience of the prototypes being tested, not previous experience with existing solutions.

As previously stated, test participants were asked to complete a number of predefined exercises. In designing the test exercises I leveraged what I had seen during my observations of the Canadian Space Agencies tests and the training exercises of rover operators at the Fukushima disaster site. In test one participants were asked to complete multiple long treks across obstacle filled terrain following loosely defined paths. In test two and three, participants were asked to navigate tight spaces, evaluate obstacles, and sample (touch) identified targets.

Test Equipment Design

VR Environment

Prototype development began with the construction and coding of the VR environment, an essential aspect of all future prototypes and testing sessions. This began with connecting the Kinect to Unity, using the Zigfu software outlined in my DIY components and solutions feasibility assessment section (figure 5.2). This allowed me to track my body (skeleton) in unity (figure 5.3), thus enabling me to map my body movements onto a 3D avatar, creating an initial semblance of embodiment into the 3D environment; an environment that the rover and operator would eventually cohabitate (figure 5.4). I was then able to import the Oculus Rift headset script into the same Unity project, creating a stereoscopic window into the 3D world. By anchoring that camera onto the head on my 3D avatar the sense of embodiment in the 3D environment was complete: I could look down and see the 3D body of the avatar as if it were my own (see figure 5.5).

Figure 5.2: Kinect and Oculus Rift setup

Figure 5.3: View of the tester by the Kinect (Skeleton Tracking)

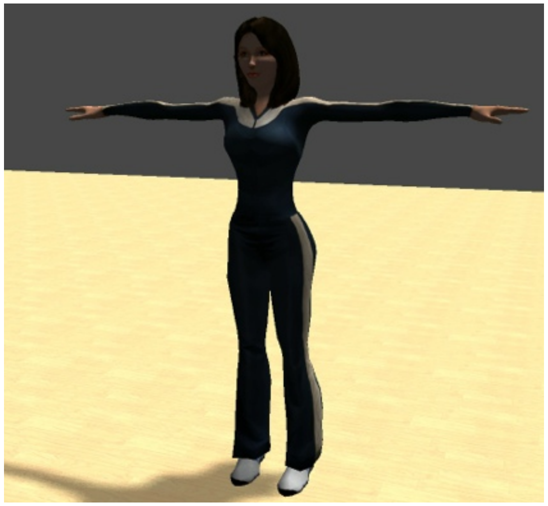


Figure 5.4: 3D avatar in Unity

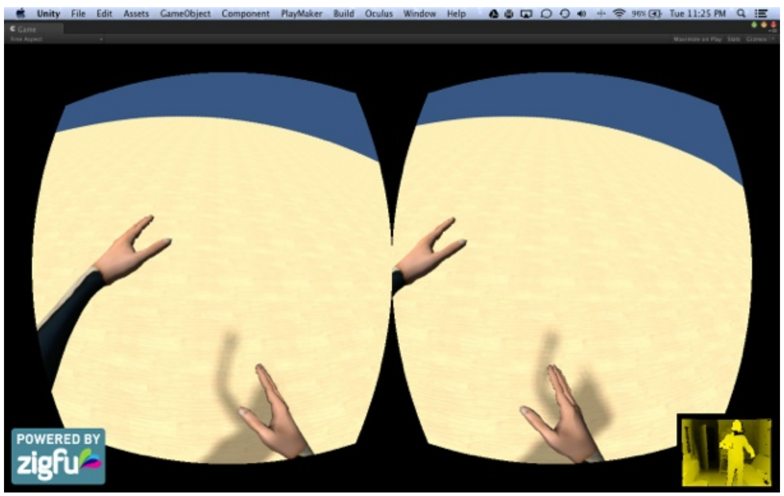


Figure 5.5: Unity, having mapped the Kinect's skeleton tracking to the avatar, displaying stereoscopic images seen by the operator (each image corresponds to an eyes)

Rover

To facilitate my comparative prototype methodology, it was necessary that I construct a physical rover to compare driving both physical and virtual rovers. This was achieved using the Rover 5, which is larger and more powerful than most of the other consumer grade rover chassises available, thereby affording expansions. These included additional hardware to provide control and navigation to the rover comparable to typical rovers such as cameras and wireless networking. Networking was provided through Xbee radios broadcasting and receiving command signals from a custom graphic user interface (GUI) in Processing which translated mouse clicks and button presses on GamePads into rover commands. Camera feed was provided by mounting an iPhone broadcasting video over Skype to my computer (figure 5.6). This video setup proved adequate though a method to tilt the camera would have been beneficial.

Test 1

For the first round of tests I developed three prototypes that built off feedback and insights I received from interviews with Ontario Drive and Gear rover operators and observations made during the Canadian Space Agency's rover control tests. Rather than discrete control solutions, these prototypes were merely a means to assess how different methods of visualizing the rover and its context would affect the test operators' experiences and operational outcomes. These three prototypes covered three control paradigms:

1. Immersive virtual interface (Figure 5.9)
2. Remote desktop interface (Figure 5.10)
3. Physically present handheld interface (Figure 5.11)

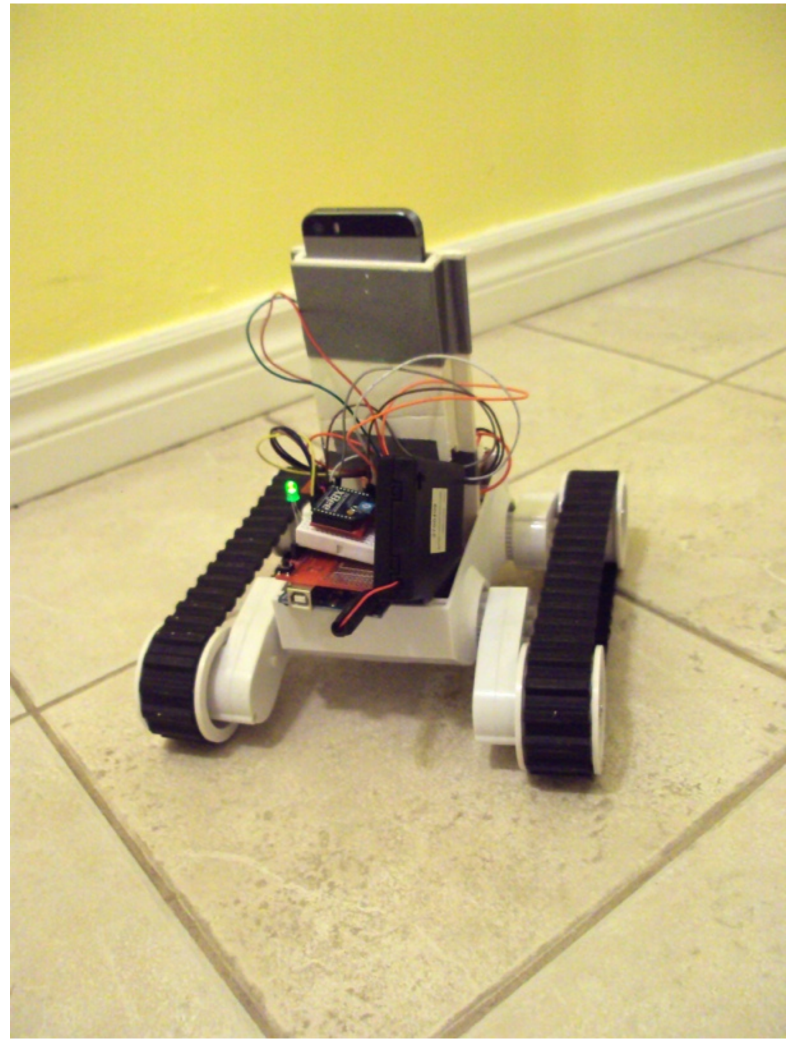


Figure 5.6: Rover Prototype and Initial Control System

Test 1 Environment

In test one participants were asked to complete multiple long treks across obstacle filled terrain following loosely defined paths. As this test was in both physical and virtual environments this involved creating the same obstacle course out of coloured cardboard and digital boxes (figures 5.7 and 5.8). In order to complete the test participants were asked to drive around blue and white obstacle cubes to reach and contact (nudge) the correct target cube, first red, then green, and then orange. As an added challenge, they were instructed not to hit any of the blue and white obstacle cubes.

Testing was carried out in the following order: remote desktop interface, immersive virtual interface, and physically present handheld interface. Upon arrival, testers were brought to a separate room from the test course, this ensured they only experienced the test course through the rover's onboard camera and, later, VR simulation before seeing it in person, the final stage of the test.

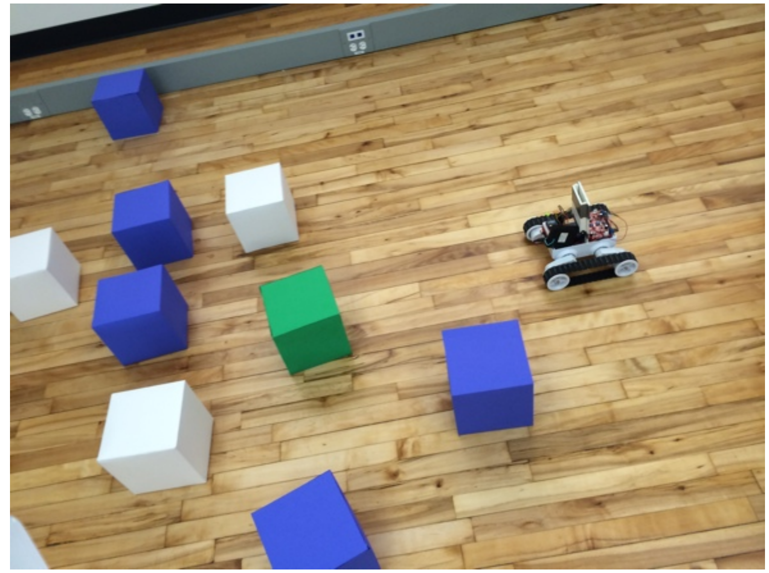


Figure 5.7 Physical obstacle course.

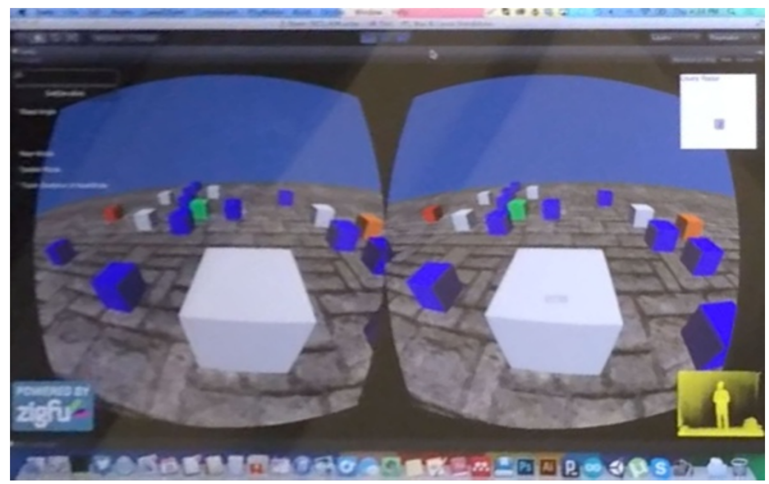


Figure 5.8 Virtual obstacle course.
(each image corresponds to an eyes)

Figure 5.9 Remote desktop interface

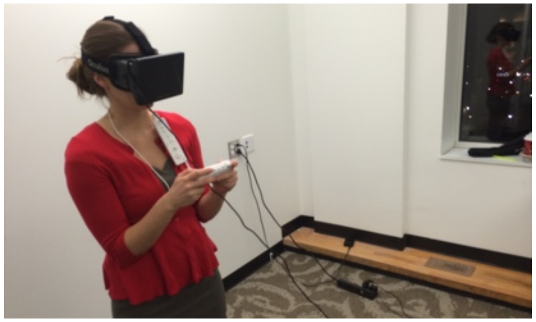


Figure 5.10: Immersive virtual interface

Figure 5.11: Physically present handheld interface

Test 1 Prototypes

The first test, revolving around the remote desktop interface prototype, involved the tester controlling the physical rover through the obstacle course using a handheld controller and a video feed displayed on a 13 inch portable computer (figure 5.9).

The second test, which revolved around the immersive virtual interface prototype, used the same handheld control but, rather than a traditional display, it relied on VR goggles to display the rover and its context, both in virtual. An additional component was freedom of mobility for the tester; afforded by Kinect skeleton tracking, the tester could walk around the virtual rover and its context (figure 5.10).

The last test did away with display methods altogether, bringing the test participant into the same room as the rover. The operator was allowed to freely move around the test course as he or she desired (figure 5.11).

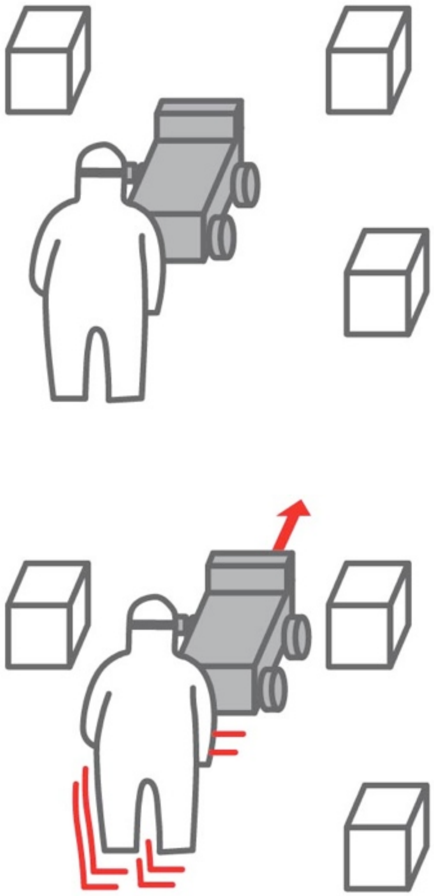


Figure 5.12: Tester avatar being dragged by the virtual rover.

Navigation and Control Methods

Three additional differences between the tests, aside from the method of visualization, applied to the virtual interface prototype. First, the rover was represented by a cube rather than a 3D model of a rover. This was the case because of limited development time available to create an accurate 3D model of the rover. As a result, it was difficult for testers to differentiate the rover from the cubes making up the obstacle course. It also made it hard to tell what was forward, back, left, and right for the rover; complicating the task of driving by failing to communicate any orientation information to the operator.

The second difference was the method of locomotion in the VR environment. In addition to affording testers freedom of mobility through Kinect skeleton tracking, driving the rover also moved them through the virtual environment: as the rover moved test subjects were dragged along with it through the virtual environment (figure 5.12). Implementing a requirement for test subjects to walk along with the rover would have quickly resulted in their stepping out of view of the Kinect. The drag system overcomes this issue by keeping the rover at the centre of the operational area (field of view of the Kinect) along with the operator.

The third difference was that the operator's head motions in the virtual interface prototype were mapped to control the rover's steering. This came out of one ODG operators interview in which he commented that it was complicated to perform multiple control operations at once. As a solution he suggested that the targeting system of the rover could be made to follow the gaze of the operator, just as is the case in some helicopter targeting systems.

This control system was not well received by the majority of the testers. The main issue was that it differed substantially from the other two comparison prototypes, in which all control revolved around the gamepad. This negated my comparative prototype development methodology. I should have changed one element, visualization, leaving everything else the same. Because I didn't it made it hard for test subjects to directly compare the three prototypes. One notable exception among the testers that disliked this method of control was a tester who self identified as dyslexic. Where she struggled to maintain a proper handle on what was the rover's orientation in the physical interface test, this control setup remove the issue, forward was always relative to her perspective. In her own words, “Wherever I'm facing is the right way”. Nevertheless, while the look based drive mechanic didn't go over well in this test I would apply it later to other control systems to much better success, for example, the menu navigation and selection system in test three.

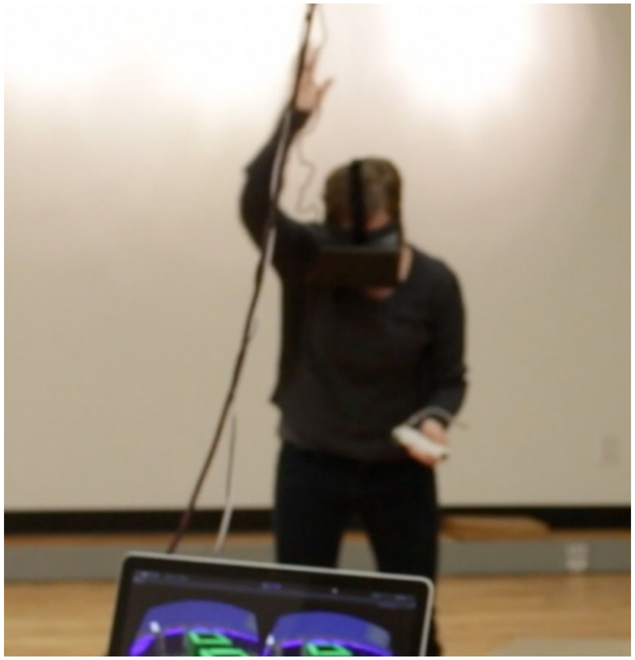


Figure 5.13 Cable imposed movement constraints and tangling

Inconsistency of Input

While I could not directly compare all aspects of the control system there were some commonalities, for example, I found inconsistency in aspects of the control system discouraged use. This applied to the joystick input on the gamepad. Users did not use the joysticks in any of the control tests because they were inconsistent with what their preconceived notion of what joystick control should do. Though joysticks can provide analog (variable speed) input, due to my inadequate implementation, they didn't provide that level of control over the rover. Exacerbating this was the 'glitchy' control output (inconsistent) of the rover due to my having inadequately implemented the technology.

Inconsistency also applied to the Oculus Rift. Users were initially reluctant to walk around with the Oculus Rift for fear of walking into things though the test room was clear of potential hazards. At this early stage in testing and exposure to the Oculus Rift testers had to re-conceptualize their expectation of walking around a room effectively blindfolded. In addition, the ever present tug of the Oculus Rift's video and power cables discouraged free movement through an otherwise open space (figure 5.13).

In anticipation of this, a red safety ring was included in the VR environment to communicate the safe operational area to the testers (figure 5.14). This ring encompassed both the area visible to the Kinect and area within which the length of the Oculus Rift's cables could be safely extended. Yet, the safety ring did not prove effective in keeping the users within the designated operational area. This was in part due to it being improperly calibrated, thus not accurately reflecting the extent of the safe operational area, causing testers to often ignored it altogether. Also, one user suggested that a real world ring could be used in order to provide some tactile feedback when moving outside the operational area. This would have made the user feel more willing to experiment in moving around the VR and physical environment.

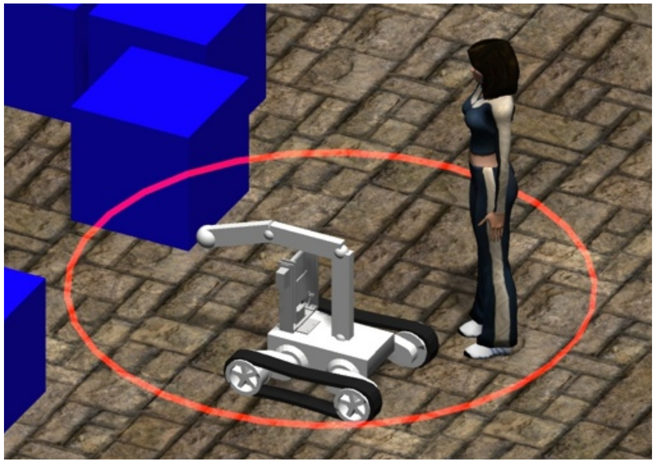


Figure 5.14 Red safety ring

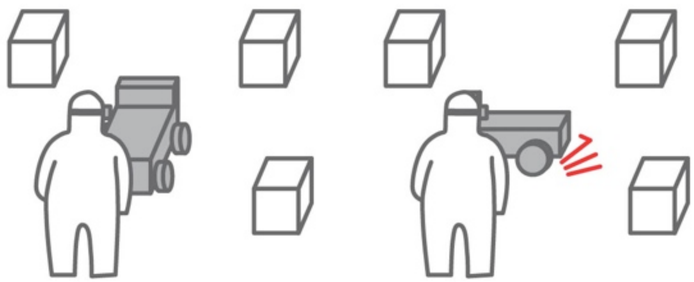


Figure 5.15 Diagram of rover turning without dragging the operator

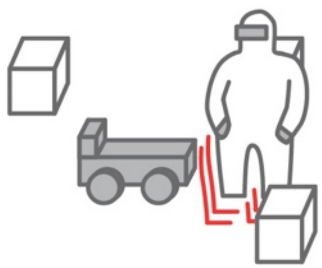


Figure 5.16 Diagram of operator walking back behind the rover

Exploration and Spatial Awareness

I had originally thought simulation sickness would be a big issue. Yet, to my surprise most users did not have issues with the Oculus Rift. From my anecdotal experience with simulation sickness while using the Oculus Rift, its root cause is when the operators vantage point changes independent of his or her own body movements. I inadvertently overcame this issue by implementing a dual locomotion system of Kinect motion tracking and rover drag system previously outlined in this section. Of note is that the drag system did not rotate the operator's avatar and frame of reference along with the rover (figure 5.15). It only moved it orthogonally in the horizontal X and Y plane (forward, backward, left, and right). This reduced unnecessary motion, thereby minimizing motion sickness, and also encouraged exploration of the VR environment using Kinect motion tracking, as operators felt it necessary to walk around in order to get back behind the rover after turning (figure 5.16).

Test subjects were seen to be more inclined to move through and explore the rover's context in the VR environment as apposed to the physical environment. One tester noted this was because, unlike the rover in all three tests or his own body in the physical test, in the VR environment the operator's avatar could move through objects without disturbing them. Therefor, rather than having to concern himself with stepping over and around objects around the rover, the tester could focus on exploration and driving.

As well as encouraging exploration, the Oculus Rift gave a good sense of how big the rover was relative to the obstacles. It also made it easy for testers to get a sense of the test course thanks to its big field of view. By comparison, the camera test had a much more limited field of view. Testers had to turn the rover around a full 360 degrees to get a sense of the test course. Despite this effort, because testers couldn't see the rover in context, they still could not grasp the width of the rover relative to the course obstacles.

One missing element in the VR simulation was sound feedback. Sound feedback gives cues as to when things are happening in the world around us; in the real world, when the rover drives it makes sounds. Addition, three-dimensional sound can provide depth cues and help build the sense of presence in a VR environment (Bowman et al, 2004). The test subjects noticed the lack of this sound in the VR world and asked for it to be included.

During her test debriefing interview, one operator stated, "Even though I see that it's moving, sound is a second signal that confirms that it is moving or stopped. It would make me feel more reassured." Reassured: that's a very apt choice of word for the is what communicability is all about, reassuring the user through communication channels.

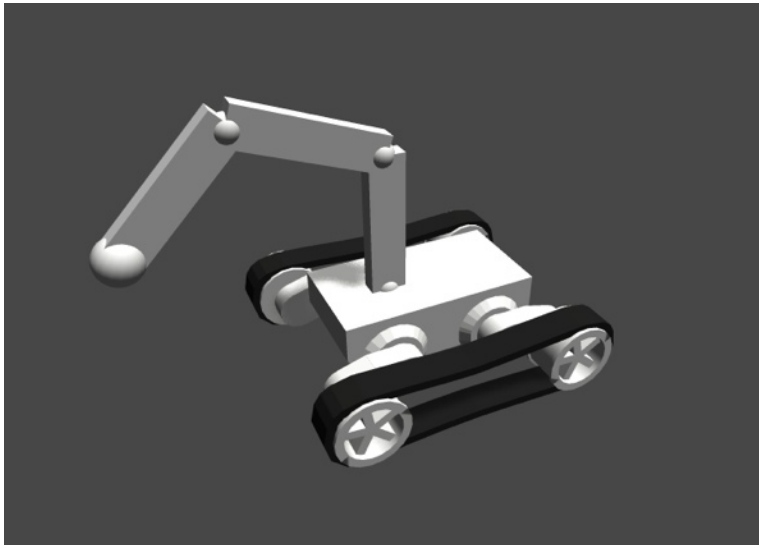


Figure 5.17: Rover equipped with the robot arm.

Figure 5.18: Rover at the Fukushima cleanup site using its onboard robot arm to open a door.

Test 2 and 3

Taking the insights from test one, I was able to move to a fully virtual setup for test 2 and 3, thus allowing me to iterate much more frequently. In these virtual tests I continued to follow my comparative prototype methodology, creating a number of prototypes that contrasted both physical control, via gamepads, and gestural control, via Kinect skeleton tracking. In order to understand the impact and nuances of these different approaches to VR control systems, the test environment was kept the same for both, as I'd done in test one.

To best challenge testers with tasks and complex control functions that rover operators undergo during real world missions, I designed and programmed a virtual target course. Beyond the course itself, this also required a virtual robot arm to contact those targets (Figure 5.17); to simulate the equipment on the rovers employed in the cleanup of the Fukushima nuclear disaster (Figure 5.18).

Test 2 and 3 Environment

Over the course of test 2 and 3 testers were asked to manipulate the robot arm aboard the rover into contact with a number of targets concealed by or within boxes (figure 5.19). These targets fell into three categories. First, there were targets concealed within boxes which were only accessible by reaching into the box's open top. Second were targets low to the ground, having stacks of blocks preventing access from above. These were only accessible by manipulating the robot arm in a very particular manor down to the ground. Lastly, there were targets inside floating boxes 1.5 meters off the ground, these were only accessible by reaching up with the robot arm from a position directly under the box. While this target setup was obviously unnatural in appearance, my intent was to prompt robot arm manipulation that would be commonplace in the real world: reaching into or around objects, reaching under objects, and reaching up to grasp objects.

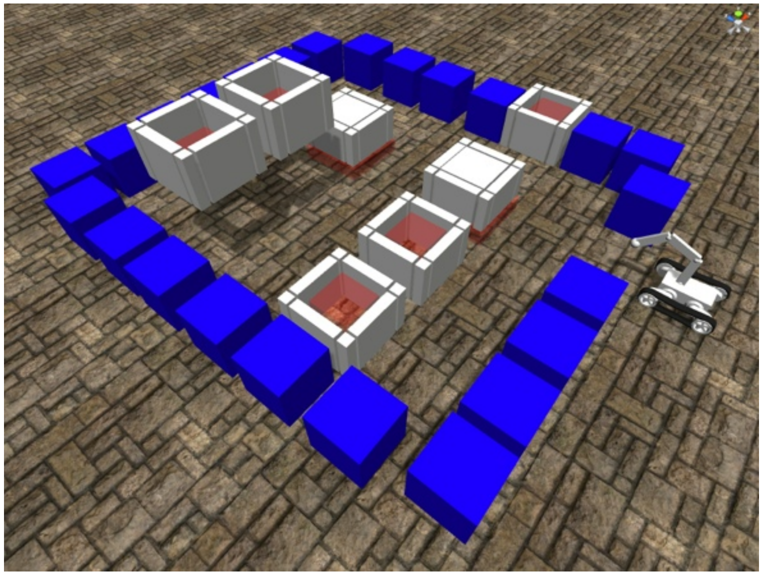


Figure 5.19: Test 2 and 3 target course.

In this course there are three open top boxes, two stacked boxes, and two floating boxes. The extent of the course is marked by blue boxes.

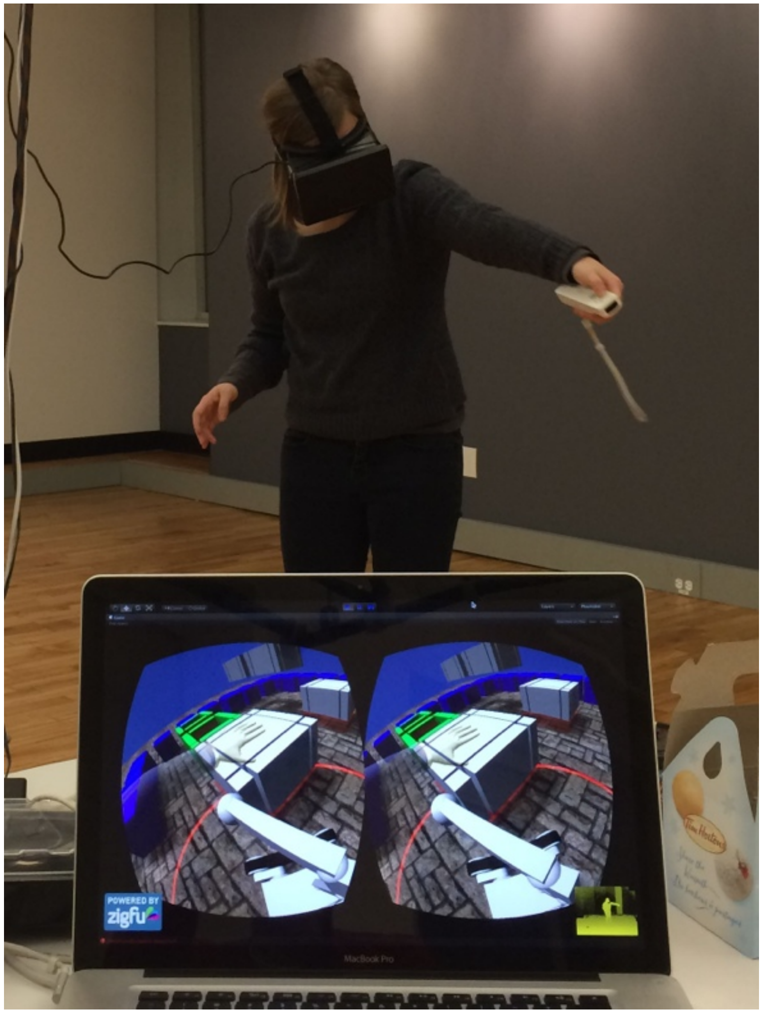


Figure 5.20: Test 2 direct manipulation control method. Tester reaching out to grasp and reposition the robot arm. (each onscreen image corresponds to an eyes)

Test 2 and 3 Prototype Design

Over the course of development and iteration spanning test 2 and 3 I created two variations on 3D interfaces to switch between control functions in the VR environment.

The first interface, used in test 2, required operators to manipulate virtual objects directly, in this way functions (repositioning the robot arm) were assigned by performing these functions directly on the object (robot arm or rover) to be controlled (Figure 5.20).

The second interface, used in test 3, assigned discrete functions for objects to each of the operator's hands through a virtual menu system (Figure 5.21). Once assigned these functions could then be performed at a distance. This provided access to eight functions:

1. robot arm joystick control
2. robot arm gesture control
3. robot arm hybrid control
4. rover drive control - 1 joystick
5. rover drive control - 2 joysticks
6. avatar scale control (joystick)
7. camera joystick control
8. camera gesture control

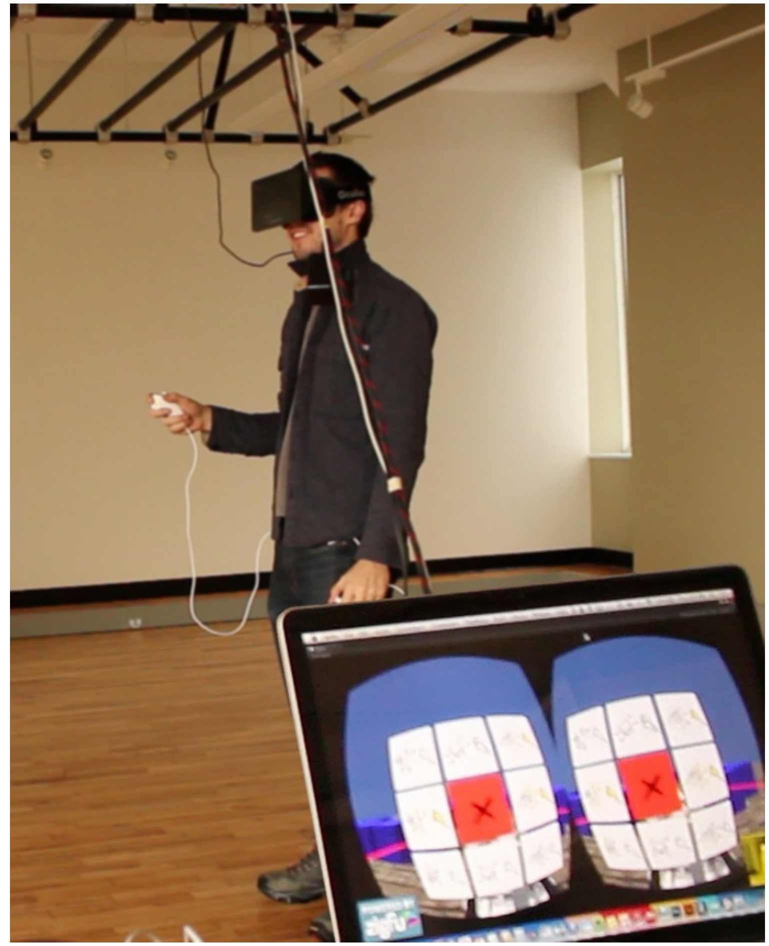


Figure 5.21: Test 3 menu based control method
Tester navigating the menu with his head motions..
(each onscreen image corresponds to an eyes)

Figure 5.22: Initial VR Glove prototype

Affecting repositioning of the robot arm through direct gestural manipulation required I overcome the inability of the Kinect to track fingers and subtle hand gestures, such as grasping, at any significant distance. Initially I had planned on overcoming this limitation by providing operators with custom made VR gloves (figure 5.22). These gloves could communicate the grasping gesture to the simulation software wirelessly (so not to be obtrusive) in addition to providing feedback cues, such as vibration through built in vibration motors, that is otherwise impossible with the Kinect alone. These seemingly inconsequential elements are extremely important for smooth and consistent operation of the rover in the virtual environment. Subtle vibrations in the gloves alert the operator to when his or her hand is over an actionable area.

Test 2 Prototype Evaluation

In addition to repositioning of the robot arm through direct gesture manipulation, my initial plan for implementing rover control and navigation was to create a method of driving control akin to pushing a cart; moving the rover would have necessitated pushing it's virtual avatar. This changed after the results of test 2.

The test subjects found it awkward to control the robot arm by direct manipulation. As well as being awkward, direct manipulation as a control method was seen as unnatural. One tester noted that we've been trained to use controls, not direct manipulation, "When you're a kid and you're playing with dump truck toys you always have the levers. You never directly work with the crane; that's not something you're used to doing."

Building off this insight, in test 3 I removed the necessity to directly manipulate the robot arm. Rather, by using a virtual menu system, users could assign the robot arm manipulation function to one of two hands. With the function literally in hand, operators had the ability to activate the reposition function at any distance from the robot arm.

Having been modified, robot arm control via gestures was very well received. One user speculated that perhaps it was because it was a robot arm being controlled via her own arm gestures, that made it easy to comprehend how her arm gestures would translate to the robot arm; it was a good cognitive fit through matching her existing mental model. Few testers noticed though that it was not a direct one to one mapping of their own arm motions; instead, the robot arm gesture control actually functioned incrementally, much like a mouse on a desktop. Each gesture was additive onto the robot arm's position rather than absolutely relative to their own arms position.

During robot arm gesture control, users had the ability to extend their own arm to some distance from their own body, thereby positioning the robot arm to a semi extended state, end the control gesture, reposition their own arm closer to their body, and begin the robot arm control gesture again, continuing to extend the robot arm further. Those users that did notice this ability did not enjoy it. Rather than being incremental, testers insisted that when using gesture control for moving the robot arm, their range of arm motion needs to be larger than that of the robot arm. This would provide some feedback, at least tacitly, that the robot arm had reached maximum extension.

Rather than being explicit, tacit communication channels inform without interrupting or need conscious attention (Norman, 2007). Joystick control inherently provides this tacit feedback because when the joysticks are maxed out and the arm no longer moved the robot arm is clearly fully extended. This lack of feedback was cited repeatedly as a source of discouragement toward using gesture control for the robot arm and a reason for using joystick input for the robot arm.

Test 3 Prototype Evaluation

As an attempt to bridge the gap between gestural an physical gamepad control I created single control functions for arm control and camera control which used hybrid control options. This required some sub functions, such as rotation or height adjustment, to be completed with gestures, and others with gamepad controls in a multimodal interplay.

Surprisingly, no test subjects like the hybrid control methods. It was easier intuitively to have it all one way or the other. One user said it was intuitively easy enough to grasp the concept but it was the shift in modalities that made it awkward: switching between small thumb movements and sweeping arm gestures. These findings match with the assertions made Sharon Oviatt in “Ten Myths of Multimodal Interactions”, in which she states that users do not always interact multimodally, even when they have the option available (Oviatt, 1999).

Another big assertion is that the flexibility of a multimodal interface can accommodate a wide range of users, tasks, and environments in which no single mode may suffice. Therefor, rapid input mode switching should be accommodated. Failing to acknowledge this will result in missing information that other modes can supply. Somewhat similarly, it is also asserted that multimodal systems can have greater reliability than unimodal systems. This is possible because, in a flexible multimodal interface, people will avoid using an input mode that they believe is error-prone for certain content (Oviatt, 1999).

There was clear evidence of these additional assertions taking place in my user testing. Testers found that certain tasks worked better with different input types. Rather than being mutually exclusive, joystick operation and gestural control both had unique merits that had to be leveraged in order to produce an optimal control system. For example, robot arm control worked better with the gesture control because it allowed for elaborate compound movements not easily input using joysticks. That said, it was extremely beneficial to have the joysticks as an alternative to gestural control. Camera controls for instance worked better with the joystick as it required very precise movement. In addition, joystick controls filled in on occasion when gestural control for the robot arm was impractical due to extreme distance from the Kinect resulting in jittery motion tracking, the operator's arm becoming occluded from the Kinect by his or her own body, or when gesture controls would result in the tester moving outside the operational area.

It was these realizations that led to my abandoning of the still untested VR glove. As development on the glove progressed I presented it to a number of test participants as a future prototype to be tested. While initially intrigued by the VR glove and the ability it could afford to naturally grasp objects in 3D space, these test subjects found the absence of an analog joystick to manipulate the rover to be a significant drawback.

Rather than use a VR glove for interaction, I would instead rely on Nintendo Wii controllers. They offered many of the same benefits as the VR glove, such as wireless

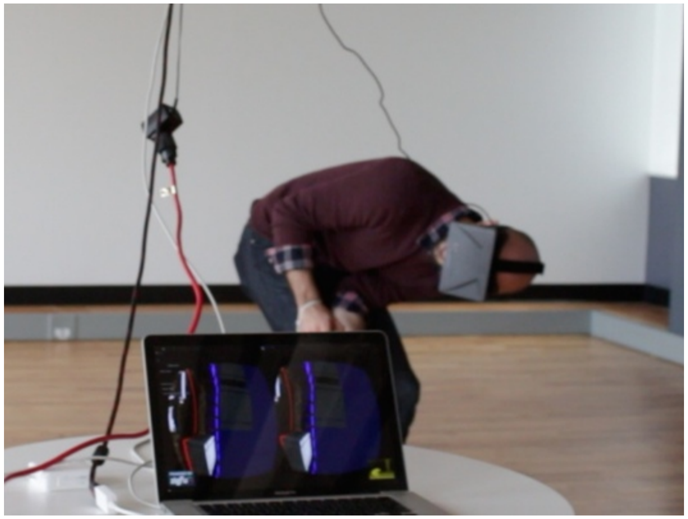


Figure 5.23: Tester ducking while repositioning the robot arm into a floating target.

Figure 5.24: Tester standing inside of a floating target while repositioning the robot arm.

communication and vibration feedback, with the added benefit of possessing digital and analog controls via d-pads and joysticks. Furthermore, unlike most gamepads, these Nintendo Wii controllers (Wiimote and Wii Nunchuck) were designed to be operated single handed, thereby allowing for independent right and left hand gestures.

Two handed operation was very popular with the users because of the afforded ability to control two functions at once through the mapping of separate functions to each hand. In the second test users hacked this functionality in by driving the rover with the robot arm in hand. This was not an ideal solution as it resulted in users standing inside the rover's 3D model for the majority of the test. I had perceived this to be a preferred vantage point for users but when it was no longer necessary, thanks to the new tools in test 3, user's were quick to note their preference to being external to the rover while performing multiple tasks.

This new freedom to move away from the rover while commanding two functions enabled testers to fully leverage the full extent of mobility offered by the VR environment as well as the etherial (ghostly) quality of their virtual avatar. As apposed to the rover, which cannot move through objects, the operator's avatar has no such limitation. This etherial quality encouraged exploration and comprehension of the rover's context through new (figure 5.23) and in some cases physically impossible vantage points, such as obtaining a better view of the robot arm during repositioning by standing within targets and obstacles (figure 5.24).

3D interface design, functions, and 'magic'

As previously mentioned, the move away from direct manipulation for assigning and carrying out functions on the rover necessitated the creation of a 3D interface. This interface would allow functions to be assigned to each of the operator's hands and be remotely triggered.

As noted by Bowman et al., (2004), creating a 3D interface is not simply a matter of transferring conventional 2D interaction styles into a 3D environment. As such, to create this interface I referenced designs from science fiction movies and video games. These media can provide many insights into possible methods in which a 3D, immersive, and embodied interface can take shape and be interacted with. Yet, I had to be mindful of using these media as design guides. In an interview with an ODG operator it was noted that video game mechanics don't translate to the real world. In her own words, "It's not a video game." I interpreted this to mean that interaction mechanisms shown in films and video games should not be directly transposed onto rover control. As such, rather than direct transposition from film and video games onto rover controls, or writing off apparently bad interaction methods altogether, I instead used a process of apologetic UI analysis as outlined by Shedroff & Noessel (2012).

Employing apologetics when analyzing bad interfaces from imagined futures, such as those found in film or video games, can help designers think more creatively about actual interfaces of the present. Apologetics is a term borrowed from religion. It's the practice of coming up with rational explanations to reconcile the apparent contradictions inside of a faith; making sense of the plot holes that surface when various religious stories are stacked on top of each other.

This can be applicable to Science Fiction; if you assume that everything in Science Fiction is there for a reason you can find some real interesting lessons in design. Take for example Star Wars. In particular, the scene with Luke and Han Solo in the Millennium Falcon blowing up Tie-Fighters. If we were really in space watching ships blow up we wouldn't hear anything; there is no air to propagate the sound waves. To understand this, one could simply assume that the sound was applied for the audience to have better enjoyment and understanding of what is going on. Yet, applying apologetics, it would be a miserable task to identify where in 3D space the Tie Fighters are without artificially generated sound.

A second display could be used to show their location around the ship but that would distract from the act of targeting them. Even a heads up display would be less efficient; sound is 360 degrees while our field of view is much less, at around 180 degrees. If we were to design the cockpit of the Millennium Falcon today we'd also have used artificially generated sound to communicate the location of Tie-Fighters around the ship. Using these artificial sounds, Luke can target a given Tie-Fighter while also being aware that another one is coming over his right shoulder. Artificial sound is vital to that interface working. You can only get to this place by trying to reconcile why a thing that seems broken may be that way for a good reason.

For my exercise with apologetics I looked at the video game Dead Space. In it the main character is controlled from a third person, over the shoulder, view. When the menu is activated it is projected in front of the character in view of both him and the player. While the player navigates the menu with a joystick the on screen character looks at the highlighted menu elements with his head position. The character does not reach out an touch the menus to select them, he simply looks at them to change his selection (figure 5.25).

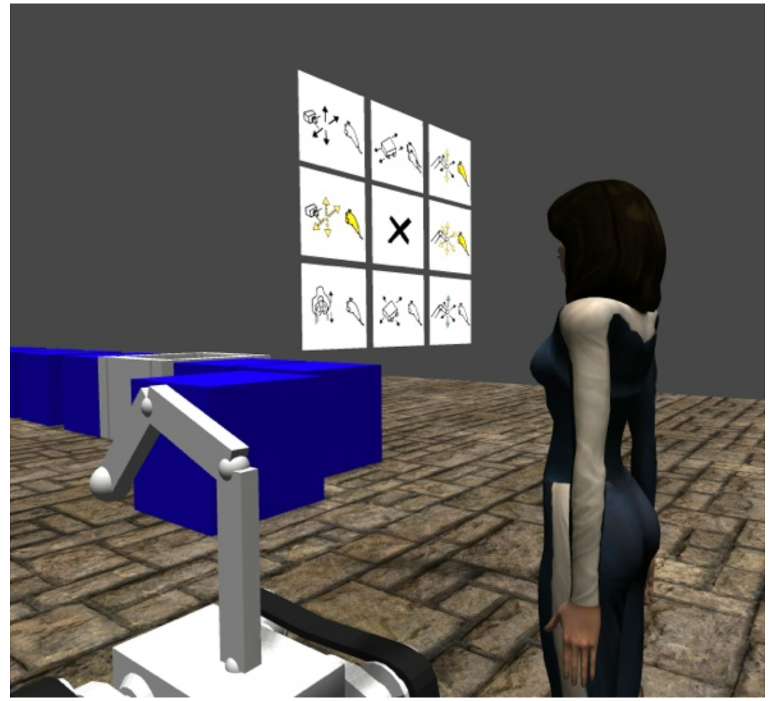


Figure 5.25: Test 3 menu based control method shown in context with the rover and operator's avatar.

From a game design standpoint this was probably implemented in order to keep the player's view of the menu unobstructed. Having the character's arm reach out to select each menu element in turn would have obfuscated much of the menu from the player. Furthermore, if implemented in the real world the user of the menu would have a very tired arm after a short period of use. Employing apologetics, from a UI standpoint, this process of using look as a selection method is the ideal setup. Looking to select menu elements is much easier than physically contacting them!

An added benefit with this method of look based menu navigation is that it decouples the menu's input channel from the main input channel, which is used for control and navigation. As Bowman et all, (2004) notes, this can decrease a users cognitive load by not requiring a switch between manipulation and system control actions issuing a given input modality, being decoupled (relying on separate modalities) allows users to perform both at once.

This selection method is akin to magic, which can be used to great effect in VR control systems, a space in which real world limitations have little to no bearing. Bowman et all, (2012), noted that magic or hyper-natural manipulation techniques outperform their more natural counterparts by making tasks easier to perform in the virtual world than in the real world.

A control prototype I created which used this principle to great effect was a mechanism to change the scale of the operator's virtual avatar, becoming a giant. This ability to become a giant and control the robot arm from afar was compared to being an omnipotent being by one user, "it's almost better than real life. You have magical powers." Beyond being entertaining, the vantage point while being a giant was extremely helpful for testers. Testers could perform some task much easier, such as positioning the robot arm into open top boxes. Conversely, remaining or returning to normal sizing was useful in positioning the arm into floating open bottom boxes. Were this project to continue, future tests should explore including the option to become smaller than normal size as a number of users requested this ability. This may have been helpful in the aforementioned floating box task.

Insights

Testing and prototype development concludes with six key insights concerning the creation of VR control systems for rover operation. These cross six areas:

1. Affordance
2. Consistency
3. Communicability
4. Feedback Loop
5. Spatial Memory
6. Simulation Sickness

In many cases, these six insights correlate to assertions and finds made in the initial literature review, ODG interviews, and observations of the CSA's testing.

Affordance

The VR simulation affords and encourages the exploration of the rover's context. On numerous occasions users were seen to be encouraged in their exploration of the 3D environment around the rover by its ethereal, ghostly, nature. Operators could move through obstacles without disturbing them. One operators said she didn't feel the need to duck to look into a floating boxed because she was a ghost, as such, she could see the space from a new perspective (figure 5.26).

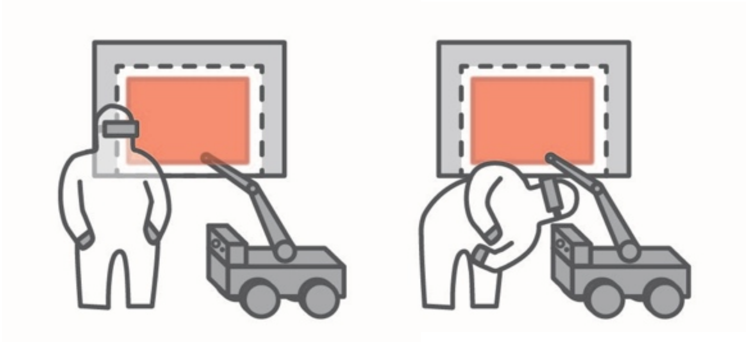


Figure 5.26: Ghosting and ducking to explore environment

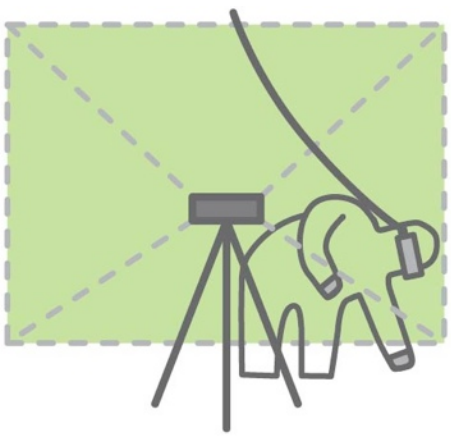


Figure 5.27: Reaching outside the Kinect's field of view and putting tension on the Oculus Rift's cable

Consistency

Current technologies inhibit operators' abilities to leverage the VR tools and space to their full extent. Many operators were unsure of the Kinect's ability to perform body tracking, particularly when reaching near the limits of the Kinect's field of view. This coupled with the tension in the Oculus Rift's cable and a fear of tangling in it produced a pragmatic effect on their desire to look around and explore freely (figure 5.27). The experience has to be consistent in order to be practical. One operator never let go of the robot arm once he first grabbed it. He said, because of the buggy nature of the test setup, he didn't trust himself or the system to be able to grab it again. This couldn't have been the most practical approach but it was consistent. This corroborates the findings of Bowman et al., (2012) who, in their research, observed that test subjects were less likely to use inconsistent natural gestures, favouring the use of traditional controls which were more reliable.

Communicability

Sensory feedback is necessary to build confidence in the performance of the VR controls. This feedback should cover auditory, visual, and tactile senses. Test subjects stated many times during test two that the avatar's hand had to acknowledge the that rover's arm was within range to be grasped. Think about your own experience of grabbing an object: as your hand moves closer to the object you rotate and open it in preparation for contact. The avatar in the test setup made no such preparation for the rover arm. Similarly, tactile cues using vibration in the controller should be provided to communicate that the grab has begun (figure 5.28). Otherwise the user can second guess their or the simulation's performance. As stated by Norman (2002), feedback systems assist in forming conceptual models of how a system works and what actions are possible with it by indicating what is happening in the machine at any given moment. Failing to provide feedback complicates a system by allowing false conceptual models to form.

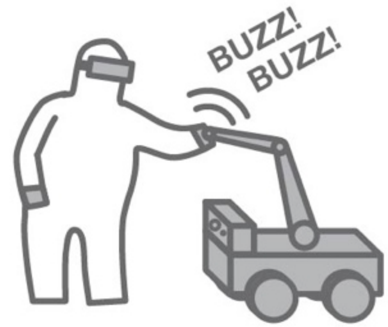


Figure 5.28: Tactile cues via vibration motors indicating grab

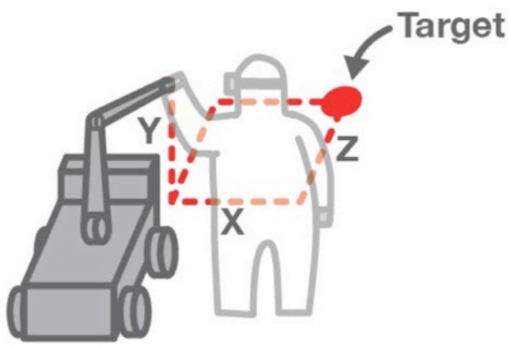


Figure 5.29: 3 axis robot arm manipulation using gestures

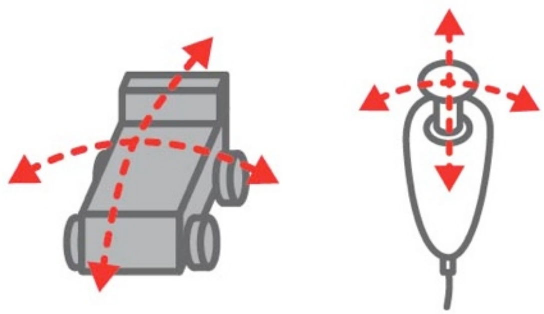


Figure 5.30: 2 axis rover navigation using handheld controls.

Feedback Loop

VR interfaces must strive to optimize a user's feedback loop, or, more specifically, the cycle between control input and output. The faster and more fluid it can be the more the user can get into the flow and concentrate on managing the system. Optimizing the user's feedback loop requires acknowledging that each control system, VR or hardware based, has its own virtues: the hardware gamepad controller is awkward for compound movement requiring manipulation in three axis but that's where VR gesture manipulation excels (figure 5.29). By contrast, gesture control is less useful for manipulations requiring repositioning in only one or two axis (figure 5.30).

Similar results were seen by Bowman et al (2012). In their study the experience of manipulating objects in 3D space was significantly easier with natural gestures. By contrast, for a driving video game, test subjects using traditional gamepad input methods had much greater performance than those using gestural controls.

Spatial Memory

Spatial Memory and familiarity with the rover's context is improved using VR controls. In the Desktop camera control test users had a limited field of view (figure 5.31). Test subjects had to keep turning the rover around to get a sense of the space around the rover and never did get a good feel for the width of the rover relative to the obstacles. By contrast, the VR simulation provided a large field of view, higher perspective, and a view of the rover in context (figure 5.32). This made it easy for test subjects to simply glance at the course and get a sense of the space, potential obstacles, and the rover's size in comparison.

The same findings were observed in the use of the VEVI's panospheric camera and Pathfinder's stereo imaging technology. These provided operators and scientists the ability to see continuously around the rover, allowing a more natural sense of position, contributing to situational awareness of the rover's environment (Nguyen et al., 2001; Sanguino et al., 2012; Stoker et al., 1999).

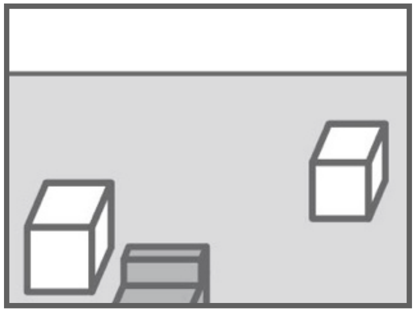


Figure 5.31: Camera View

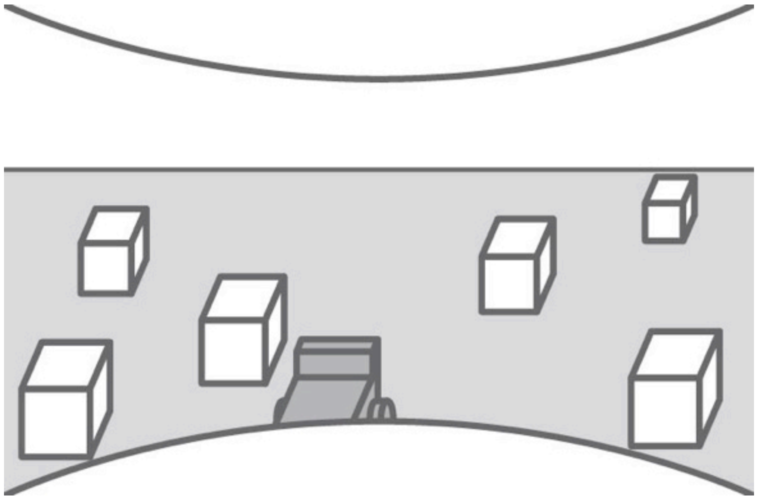


Figure 5.32: VR view

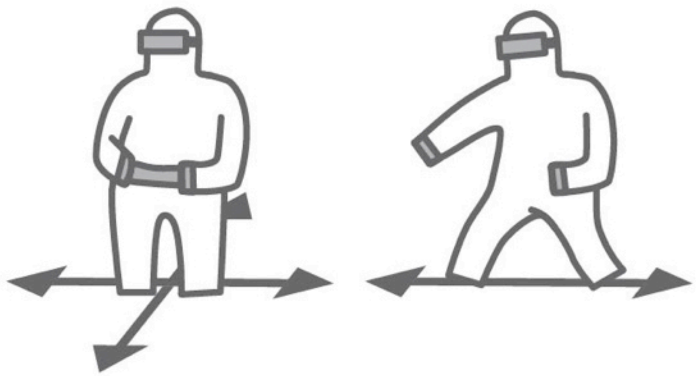


Figure 5.33: Gamepad locomotion versus physical locomotion

Simulation Sickness

Lastly, simulation sickness can be created by the Oculus Rift when the operator's vantage point is moved by a force other than his or her own physical movements, such as with a gamepad. Such movements of vantage point create a discrepancy between perception and expectation. It's akin to reading on a bus but in reverse, there is no sense of motion but the vantage point moves. Using the Kinect skeleton tracking ability to move the vantage point with the operators own physical movements, such as walking or ducking, eliminates this issue; the sense of motion and change in vantage point match (figure 5.33).

Chapter 6: Conclusion

Discussion

This thesis outlined the creation of an innovative control system for the operation of rovers in remote or potentially hazardous locations. The current standard for rover control systems is not that dissimilar to controlling a video game character in first person. As I observed in my research, this method of control breaks down when the rover's context or own characteristics are unfamiliar to the operator. As such, this thesis explored the abstraction of the traditional control structure for rovers and their menu systems; placing the operator external to the robot, controlling it in third person as seen from their own perspective inside of a virtual reality environment. Using motion tracking technology and virtual reality goggles the operator, along with a virtual representation of the rover, is located inside of a virtual reality environment simulating the contextual environment of the rover. Embodying the same environment as the rover, the operator is then able to interact with it through natural gestures and manipulation. This allows the operator to overcome some of the issues endemic to controlling robots, such as cognitive workload arising from unfamiliarity with the rover's size, its context, as well is the completion of complex tasks and manoeuvres.

In conclusion, I present a number of findings regarding the creation of natural, VR interfaces for rover operation and its affects on operators: first, development outcomes from this work imply that creating gesture control systems in isolation is not sufficient to create a productive interface. Optimizing gestural control systems requires collaboratively working with operators to understand how gestures, the operators, and the properties of the enabling technologies interact with each other in order to produce effective, straightforward, and effortless gestures and outcomes. This exercise can be challenging, as the application of fully immersive VR to the control of remote robots is in its infancy.

Second, this work and those cited in my literature review present evidence that shifting the perspective of a rover operator to third person, and enabling a view of the robot in context, increases rover operator task performance by providing a greater level of situational awareness to operators than traditional camera and display systems. Unfortunately, as evidenced in my prototype development section, consistency of the current technology makes the application of VR as a control system only viable for very select cases: where the operator does not or will not expect to move that far from the rover. In cases where the operator strays too far the VR environment will breakdown.

Third, in the application of VR, designers need not focus on visual fidelity of the simulation, rather, they must strive to provide the greatest fidelity and perception of feedback and affordance of action to the operator. As noted in my literature review, these perceptions, more than anything, contribute to a sense of realism as well as usefulness of a VR environment. These designers must also be aware of which input methods, gesture or traditional controls, are best suited to a given task or technology. From my own experience in designing the VR control systems outlined in this thesis, it is all too easy to feel the need to provide gestural control in all cases rather than question the viability of those gestures. Lastly, VR control system designers need to be aware of simulation sickness, its root causes, and the methods to mitigate it. This has the possibility of severely limiting the methods in which a VR control system may be developed or utilized.

While the system outlined in this thesis is not viable as a standalone method for operating a rover, it demonstrates the value of using full body immersive virtual reality environments in their operation. In doing this, I have shown how a virtual reality control systems can function and the challenges they face when applied in the operation of remote rovers. This includes highlighting design challenges such as the reliability of the skeleton tracking system and creating appropriate control methods for this new interface paradigm. As previously noted, these must be addressed further before such a system can move into real-world settings.

Future directions

As noted in my conclusion, the immersive VR control system outlined in this thesis is not currently viable as a standalone method for rover operation. Issues of reliability of skeleton tracking, refinement of control gestures, system portability, and training methods must be addressed before it can move into a real-world setting. Were this work to continue, improved tracking of the operator can be achieved by using the higher resolution Kinect for the Xbox One rather than the Xbox 360. Moreover, even greater improvements could be attained through technologies which can track points on an operator's body independent of fixed depth cameras such as the Kinect. These technologies are currently unavailable at the consumer level but will be arriving in the market shortly.

When issues with tracking the operator's body have been removed, further exploration into additional control gestures enabled by a full 360 degree virtual operational environment should be carried out. As noted in the prototype development section, many of the gestural control methods implemented in this work were limited by the gestural tracking technology I had employed. A particular focus should be placed on the method of driving the rover, which relied on one or two joysticks in the final prototype. An issue with the two joystick driving method was that it remove the ability to drive and control the robot arm at the same time. One tester, who self identified as a power user, wished to control as many independent functions as possible. In his opinion this would allow him to navigate the test environment with greater efficiency. This would likely require peddles, as are found in construction equipment, to control the treads in addition to joysticks and gestures for other functions. This would fundamentally change the nature of the VR environment, requiring the user to be sitting.

In addition to tracking, an additional area for future exploration should be the miniaturization of this control system. As was noted by one operator at Ontario Drive and Gear, many rover control systems need to be mobile. As a result, these don't have a camera feed and display because of the associated increase in size. It is easy to imagine a system relying solely on a goggle display, such as with the Oculus Rift, being quite a bit more portable than a traditional display setup. Yet, the control system I have outline in this thesis also requires a large controlled environment in which to setup the equipment involved in gesture tracking: Kinect on tripod and control computer. As such, explorations into miniaturizing the method for gesture tracking, as well as gleaning the possible benefits and future application potentials associated with doing so are in order.

Lastly, the process of introducing operators to the methods of operating in this VR control system need to be improved. As one tester noted, a tutorial level to introduce the UI concepts and core mechanics at each test would have been extremely valuable. Video games have been using this principle for decades. In games, before being able to progress users are required to move through a series of exercises or screens that introduces the core mechanics of the game. These provide very basic but essential information to performing successfully in the game. In conducting the tests, I observed that without an introductory script or a tutorial I lacked the ability to consistently introduce new UI functions. Again, a tester noted that the menus and functions that I implemented are good reminders of functionality but they aren't teaching tools.

Were these issues and lines of inquiry addressed VR control systems could begin to provide yet unimagined methods of interaction and conceptualization of data for numerous industries. As noted in my literature review and needs assessment chapters, the Canadian Space Agency and NASA are readily adopting 3D and VR systems to improve operational outcomes in space and terrestrial missions. Moreover, the US NAVY is exploring the use of VR systems to envision the command and control centre for battleships 15 years in the future. With the introduction of the Oculus Rift and the Kinect, such inquiry and development of VR control systems will accelerate; what the future holds will surely dazzle the imagination.

Bibliography

Ark, W., Christopher, D., Ted, D., & Zhai, S. S. (1998). *Representation matters: The effect of 3D objects and a spatial metaphor in a graphical user interface.* In In People and Computers XIII, Proceedings of HCI'98 (pp. 209–219). Springer-Verlag.

Baker, Phil. *From Concept to Consumer: How to Turn Ideas into Money.* A Stewart Emery Series Book. Upper Saddle River, N.J: FT Press, 2009.

Blackwell, A. F. (2006). *The reification of metaphor as a design tool*. ACM Transactions on Computer-Human Interaction *(TOCHI)*, *13*(4), 490–530.

Borst, C. W., & Volz, R. A. (2005). Evaluation of a Haptic Mixed Reality System for Interactions, 14(6), 677–696.

Bowman, D. A., Kruijff, E., LaViola, J. J., & Poupyrev, I. (2004). 3D User Interfaces: Theory and Practice. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.

Bowman, D. A., McMahan, R. P., & Ragan, E. D. (2012). Questioning naturalism in 3D user interfaces. Communications of the ACM, 55(9), 78. doi: 10.1145/2330667.2330687

Chance, S., Gaunet, F., Beall, A. and Loomis, J. *Locomotion mode affects the updating of objects encountered during travel*. Presence: Teleoperators and Virtual Environments 7, 2 (1998), 168–178.

Errey, C., Ginns, P., & Pitts, C. (2006). Cognitive load theory and user interface design: Making software easy to learn and use. PTG Global.

Guizzo, E. (2012). *Fukushima Robot Operator Writes Tell-All Blog*. IEEE Spectrum. Retrieved February, 2, 2013, from http://spectrum.ieee.org/automaton/robotics/industrial-robots/fukushima-robot-operator-diaries

Helton, W. S., Head, J., & Blaschke, B. A. (2013). Cornering Law The Difficulty of Negotiating Corners With an Unmanned Ground Vehicle. Human Factors: The Journal of the Human Factors and Ergonomics Society, 0018720813490952. doi:10.1177/0018720813490952

Hollister, S. (2014) *BlueShark: where the US Navy dreams up the battleship interfaces of tomorrow.* The Verge. Retrieved January 26, 2014, from http://www.theverge.com/2014/1/26/5346772/blueshark-us-navy-oculus-rift-virtual-interface

Lakoff, G., Johnson, M (1980). *The Metaphors We Live By [Kindle version].* Retrieved from Amazon.com

Lapointe, J.F., & Massicotte, P. (2003). Using VR to improve the performance of low-earth orbit space robot operations. Cyberpsychology & Behavior, 6(5), 545–8. doi:10.1089/109493103769710578

Mantovani, G., & Riva, G. (1999). "*Real*" *Presence: how different ontologies generate different criteria for presence, telepresence, and virtual presence*. Presence: Teleoperators and Virtual Environments, 540–550. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/105474699566459

Nielsen, J. (2000). *Why You Only Need to Test with 5 Users. Nielsen Norman Group.* retrieved June 25, 2013, from http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Nguyen, L. A., Edwards, L. J., Flueckiger, L., Neveu, C., Schwehr, K., & Wagner, M. D. (2001). *Virtual Reality Interfaces for Visualization and Control of Remote Vehicles*, 59–68.

Norman, D. (2002). *The Design of Everyday Things [Kindle version]*. Retrieved from Amazon.com (Original work published 1988)

Norman, D. A. (2007). *The Design of Future Things [Kindle version]*. Retrieved from Amazon.com

O'Hara, K., Gonzalez, G., Sellen, A., Penney, G., Varnavas, A., Mentis, H., Carrell, T. (2014). Touchless Interaction in Surgery. Commun. ACM, 57(1), 70–77. doi: 10.1145/2541883.2541899

Oviatt, S. (1999). Ten Myths of Multimodal Interaction. Commun. ACM, 42(11), 74–81. doi: 10.1145/319382.319398

Saffer, D. (2009). *Designing for Interaction: Creating Innovative Applications and Devices (2nd Edition)(Voices That Matter) [Kindle version]*. Retrieved from Amazon.com

Saffer, D. (2011). *Designing Devices [Kindle version].* Retrieved from Amazon.com

Sanguino, T. J. M., Márquez, J. M. A., Carlson, T., & Millán, J. R. (2012). Interaction and Evaluation of an Augmented Virtuality Assistance System for Teleoperated Robots.

Shedroff, N., & Noessel, C., (2012). Make it so: interaction design lessons from science fiction *[Kindle version].* Retrieved from Amazon.com

Six degrees of freedom. (n.d.). In Wikipedia. Retrieved March 23, 2014, from http://en.wikipedia.org/wiki/Six_degrees_of_freedom

Stoker, C. R., Zbinden, E., Blackmon, T. T., Kanefsky, B., Hagen, J., Neveu, C., … Sims, M. (1999). Analyzing Pathfinder data using virtual reality and superresolved imaging. Journal of Geophysical Research: Planets, 104(E4), 8889–8906. doi:10.1029/1998JE900019

Syslo, M. M., & Kwiatkowska, A. B. (2005). Informatics versus information technology: how much informatics is needed to use information technology – a school perspective, *From computer literacy to informatics fundamentals*. Presented at the ISSEP'05: International Conference on Informatics in Secondary Schools, Evolution and Perspectives, Klagenfurt, Austria (pp. 178–188) Springer-Verlag.

Ware, C. and Franck, G. *Evaluating stereo and motion cues for visualizing information nets in three dimensions*. ACM Transactions on Graphic 15, 2 (1996), 121–140.

Whelan, T., Kaess, M., Fallon, M., Johannsson, H., Leonard, J., & McDonald, J. (2012). Kintinuous: Spatially Extended KinectFusion. Retrieved from http://dspace.mit.edu/handle/1721.1/71756

Yates, F. A., (2011). The Art of Memory *[Kindle version].* Retrieved from Amazon.com

Chapter 5

Appendix

Research Ethics Board

August 14, 2013

Dear Hudson Pridham,

RE: OCADU 118, “A VR simulation for robot navigation and control.”

The OCAD University Research Ethics Board has reviewed the above-named submission. The protocol and the consent forms dated August 14, 2013 are approved for use for the next 12 months. If the study is expected to continue beyond the expiry date (August 13, 2014) you are responsible for ensuring the study receives re-approval. Your final approval number is **2013-30.**

There are two changes that are required to your Consent Form. Please note that the proper name of the Faculty is the Faculty of Design not the Faculty of Inclusive Design. In the section on “Voluntary Participation”, you must also provide an exact date by which participants can withdraw from the student. Please submit a copy of your revised form.

Before proceeding with your project, compliance with other required University approvals/certifications, institutional requirements, or governmental authorizations may be required. It is your responsibility to ensure that the ethical guidelines and approvals of those facilities or institutions are obtained and filed with the OCAD U REB prior to the initiation of any research.

If, during the course of the research, there are any serious adverse events, changes in the approved protocol or consent form or any new information that must be considered with respect to the study, these should be brought to the immediate attention of the Board.

Research Ethics Board

The REB must also be notified of the completion or termination of this study and a final report provided. The template is attached.

Best wishes for the successful completion of your project.

Yours sincerely,

Tony Kerr, Chair, OCAD U Research Ethics Board