Natural Full Body Interaction for Navigation in Dismounted Soldier Training

Brian M. Williamson, Dr. Chadwick Wingrave, Dr. Joseph J. LaViola Jr.
University of Central Florida
Orlando, FL
bwilliam@eecs.ucf.edu, cwingrav@eecs.ucf.edu, jjl@eecs.ucf.edu

Tim Roberts and Pat Garrity
U.S. Army Simulation & Training Technology Center
Orlando, FL
tim.e.roberts@us.army.mil
pat.garrity@us.army.mil

ABSTRACT

Realistic interaction in 3D Virtual Environments (VEs) gives Soldiers a greater task focus (e.g. train as you fight), instead of focusing on the interface or learning training behaviors unrelated to the task. Full-body interfaces include all training interactions, such as navigation, team communication, firing and manipulating weapons, and other relevant domain tasks. For VEs that deal with training and simulation, appropriate interaction mechanisms are critical to providing realistic and immersive experiences that closely mimic real world activities being taught as the computer interface becomes invisible to the user. In particular, dismounted Soldier training requires a VE experience that closely resembles the real world; not only in terms of visual fidelity, but also in terms of the actions Soldiers need to perform. In this paper, we present a prototype system for dismounted Soldier training, where users can navigate through a virtual environment using natural full body motions. Soldiers can run, jump, crouch, and turn in the VE while still holding their weapon by simply using their bodies, mimicking the motions they would normally do in the real world. Our system makes use of video game console motion controllers including the Microsoft Kinect, Playstation Move, and Nintendo Wiimote, combined with the Unity 3D game engine to support untethered interaction. Our system makes use of a set of heuristic recognition rules that recognize various actions taken from the Kinect's depth image 3D skeleton representation. These rules support seamless transitions between realistic physical interactions (e.g., actually walking and running) and proxied physical interactions (e.g. walking and running in place) that support locomotion in the larger VE. We discuss the details of our prototype and provide an informal evaluation using subject matter experts.

KEYWORDS

Dismounted Soldier, Virtual Environments, Full Body Interfaces, Immersive, Full Body Motions

ABOUT THE AUTHORS

Brian Williamson has a master's in computer science from the University of Central Florida. His interests include computer graphics, video game design, user interfaces, military applications, and GPS/INS (Inertial Navigation System) navigation. His current project involves navigation techniques using the Wii Remote and MotionPlus, Microsoft Kinect, and Playstation Move technologies.

Chadwick A. Wingrave is a postdoc in the University of Central Florida's School of Electrical Engineering and Computer Science and a member of the university's Interactive Systems and User Experience Lab. His primary research interests involve 3D user interaction, human-computer interaction, and 3D user-interaction infrastructure for virtual, augmented, mobile, desktop, and entertainment environments. Chadwick has a Ph.D. in Computer Science from Virginia Polytechnic Institute and State University.

Joseph J. LaViola Jr. is the SAIC Faculty Fellow and assistant professor in the Department of Electrical Engineering and Computer Science and directs the Interactive Systems and User Experience Lab at the University of Central Florida. He is also an adjunct assistant research professor in the Computer Science Department at Brown University. His primary research interests include pen-based interactive computing, 3D spatial interfaces for video games, predictive motion tracking, multimodal interaction in virtual environments, and user interface evaluation. His work has appeared in journals such as ACM TOCHI, IEEE PAMI, Presence, and IEEE Computer Graphics & Applications, and he has presented research at conferences including ACM SIGGRAPH, ACM CHI, the ACM Symposium on Interactive 3D Graphics, IEEE Virtual Reality, and Eurographics Virtual Environments. He has also co-authored "3D User Interfaces: Theory and Practice," the first comprehensive book on 3D user interfaces. In 2009, he won an NSF Career Award to conduct research on mathematical sketching. Joseph received a Sc.M. in Computer Science in 2000, a Sc.M. in Applied Mathematics in 2001, and a Ph.D. in Computer Science in 2005 from Brown University.

Timothy Roberts is a Science and Technology Manager of Dismounted Soldier Simulation and Training Technologies at the U.S. Army Research Laboratory (ARL) Simulation and Training Technology Center (STTC). He currently works in the Dismounted Soldier Simulation Technologies conducting R&D in the area of dismounted soldier training and simulation. His research interests include dismounted soldier training environments, games for training, virtual locomotion, multi-modal interfaces, path planning, control systems and robotics. He received a BS in Electrical Engineering in 2002 and a MSEE in Control Systems and Robotics in 2007 from the University of Central Florida.

Pat Garrity is a Chief Engineer at the Army Research Laboratory's (ARL) Simulation and Training Technology Center (STTC). He currently works in Dismounted Soldier Simulation Technologies conducting research and development in the area of dismounted Soldier training & simulation where he was the Army's Science & Technology Manager for the Embedded Training for Dismounted Soldiers program. His current interests include Human-In-The-Loop (HITL) networked simulators, virtual and augmented reality, and immersive dismounted training applications. He earned his B.S. in Computer Engineering from the University of South Florida in 1985 and his M.S. in Simulation Systems from the University of Central Florida in 1994.

Natural Full Body Interaction for Navigation in Dismounted Soldier Training

Brian M. Williamson, Dr. Chadwick Wingrave, Dr. Joseph J. LaViola Jr.
University of Central Florida
Orlando, FL
bwilliam@eecs.ucf.edu, cwingrav@eecs.ucf.edu, jjl@eecs.ucf.edu

Tim Roberts and Pat Garrity
U.S. Army Simulation & Training Technology Center
Orlando, FL
tim.e.roberts@us.army.mil
pat.garrity@us.army.mil

INTRODUCTION

Existing dismounted Soldier training simulators can lead to improperly trained responses if systems do not train Soldiers in realistic reactions to battlefield situations. As such, natural movements to control a virtual avatar train appropriate responses, not simulator responses, and keep with the mantra of "train as we fight". High-end simulators enable more natural training but remain costly and encumber the Soldier with obtrusive sensors. This makes them impractical for large squad training at the scale and frequency necessary to serve current training needs.

In this paper, we present software and algorithmic advances using current off the shelf hardware to make natural full-body interaction training, at the scale and frequency needed, a possibility. A prototype, dubbed RealEdge (shown in Figure 1), interprets natural movements for navigation and locomotion control.

Figure 1 – RealEdge prototype utilizing the Sony Move and Microsoft Kinect.

RealEdge addresses the following navigation tasks:

- Real-time orientation mapped to the user
- Near-field locomotion in a constrained space
- Long-range locomotion inside the simulation
- Tactical gesture recognition (jumping, crouching, grenade throwing)
- Weapon tracking, including aim and reload.

We focus on recent gaming technologies created by Microsoft, Sony, and Nintendo known as the Kinect, Move and Wii Remote (Wiimote). In addition to being inexpensive, these devices are often used by Soldiers, making their availability ubiquitous which is important for being repurposed as training devices.

Microsoft's Kinect is capable of producing a colored RGB image along with a correlated depth image. This creates a real-time image with both color and distance associated with each pixel. We used the Primsense drivers, developed by OpenNI, and the NITE body skeleton tracking system as seen in Figure 2 (Primesense, 2010). This software takes in the color and depth image and produces a skeleton output of up to four tracked users. This skeleton data is composed of twenty-four joints all with three-dimensional position data and rotation matrices for orientation of the joint[1]. Using this data we are capable of understanding how the user's body is moving and translate this to solve several navigation challenges.

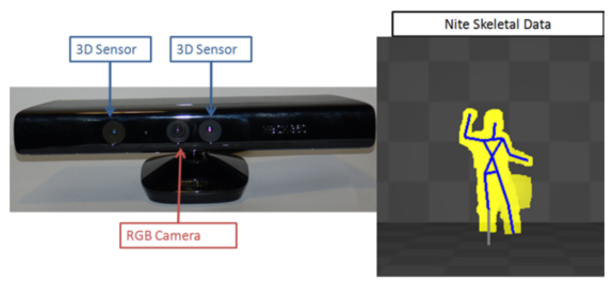

Figure 2 - Kinect sensor with example output from NITE skeleton given the Kinect depth image.

The Playstation Move is a controller and camera that is accessed with the Move.Me software application (Sony, 2011). The controller combines accelerometers, gyroscopes, and magnetometers to know the device's position and orientation. Along with this, the Playstation Eye camera picks up the glowing sphere on top of the controller to know exact position in real-time

[1] The Primsense NITE Algorithms documentation in the NITE package mentions orientation as being noisier than position data and being unable to recognize some "twist" rotations (Primesense, 2010).

while visible. This data is processed on the Sony Playstation 3 and streamed to a connected computer as position, acceleration, angular velocity, orientation, and other ancillary data. This is useful for determining the exact position and orientation of an object the user is holding attached to the controller, such as the Playstation Move Sharp Shooter shown in Figure 3.

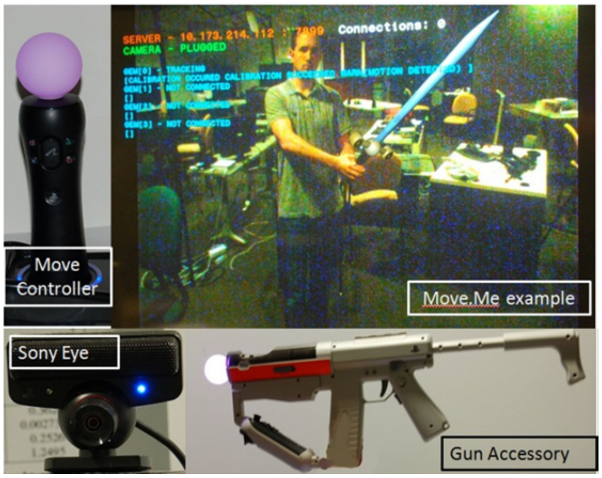

Figure 3 - The Playstation Move Controller and PlaystationEye system. Also shown is a gun accessory (bottom right) used in our prototype and an example of the Move.Me software using just the controller.

The Nintendo Wiimote (seen in Figure 4) was used as an alternative design for the weapon used in RealEdge. Similar to the Playstation Move, it combines accelerometers and gyroscopes along with infrared data to determine the position and orientation of the controller. The software drivers, however, are only capable of providing the raw data (Peek, 2008), leaving it to the programmer to compute, as best as possible, position and orientation information (Williamson, et al., 2010).

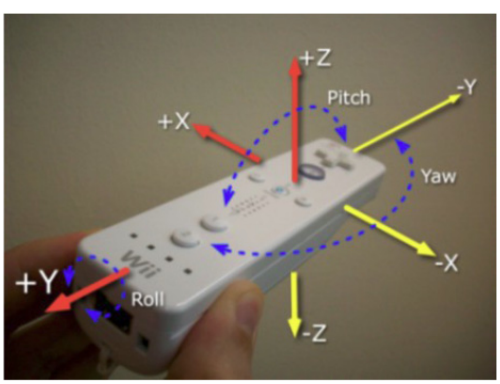

Figure 4. Wiimote with axes illustrated.

In section two we review other literature in the field of simulation training and 3D user interface techniques. Section three discusses our prototype simulation design and software iterations that have been developed. Section four presents the final prototype and evaluation of the system by subject matter experts. Section five concludes the paper while laying-out future work.

RELATED WORK

The idea of using virtual environments for training has been researched for many years. As in one case, virtual training was shown to successfully train Soldiers to follow routes in a way that applied to real world settings (Witmer, Bailey, Knerr, 1995). This research was expanded upon in later years to show that as technology has improved, virtual environments could be used to train Soldiers in several unique areas at a time (Knerr, et al., 2003). Though these systems are high fidelity, their costs can vary from eighty thousand dollars to over four hundred thousand dollars (Knerr, 2006) and at times be obtrusive to the user, which may affect their sense of presence negatively as compared to an unobtrusive system (Childs, 2010). Our proposition is that similar simulation challenges can be resolved with lower cost systems without sacrificing presence or accuracy.

One major challenge we face is locomotion, which is defined as the motor component of travel (Bowman, et al., 2005). We considered the ability to maneuver in real time essential, as it provides the user higher levels of cognitive understandings of the space around them (Zanbaka, et al., 2004). Walking or running in place, however, is shown to also maintain presence, while providing an alternative of using minimum area (Slater, Usoh, Steed, 1995). While research has been made to provide natural walking over long ranges (Interrante, Ries, Anderson, 2007), this typically involves very expensive equipment which counters our low cost objective.

The use of console systems has also become prevalent in research to resolving virtual locomotion problems. The Wiimote broke ground as a device used for interaction techniques in several research papers (Lee, 2008) (Shirai, Geslin, Richir, 2007) (Wingrave, et al., 2010). In our previous research (Williamson, Wingrave, LaViola, 2010) we looked at the Wiimote as a means to resolve many of the locomotion problems examined in this work. However, the Wiimote had limited data when compared to what is offered by the Kinect and Move. As such, research is emerging with these new technologies, such as a depth camera being used as a touch sensor (Wilson, 2010).

What previous research shows us is that simulators can be used to train Soldiers, as long as they maintain presence and higher leveling understanding of the situation when resolving issues, such as locomotion and orientation. We also see that commercial systems are emerging as low cost solutions to previously expensive problems, giving us an opportunity to combine both, creating a system of high fidelity, without large expenses.

REALEDGE DESIGN

RealEdge Setup

We began with the Microsoft Kinect, to recognize full body movements, and combined it with the Playstation Move and the Nintendo Wiimote, as an alternate design choice for the weapon proxy. The hardware was positioned as closely together as possible to better map coordinates between systems. We placed the Kinect centered and on top of the screens. The Wiimote, when used, had its infrared transmitter placed directly below the Kinect. The Move's Playstation Eye was placed on top of it. The prototype recognition software was written in C# and used to control movement within the Unity 3D Engine, specifically, Unity's existing bootcamp demonstration available with the 3D engine (Unity, 2011)

In Table 1 we go through the price of the system to illustrate its cost, as compared with other systems. All the prices listed include other hardware needed, such as the Wiimote requiring an infrared bar and the Move requiring a Playstation Eye.

	Variances in Equipment Cost	
	Low Estimate	**High Estimate**
PC	$1000	$3500
Screen	$600	$5500
Wiimote	$35	$55
Kinect	$135	$150
Playstation 3	$300	$350
Playstation Move	$80	$95
Accessories	$30	$60
Total (Wii)	$1800	$9265
Total (Move)	$2145	$9655

Table 1. Cost Breakdown of the prototype system for a low to high-end system.

Orientation and Locomotion Techniques

Our first goal was proving the Kinect could be used to support realistic locomotion and navigation techniques.

We first mapped the gaze-based orientation of the simulator's camera to the user's viewpoint (Norton, et al., 2010). This was done by receiving the orientation of the head joint and position of the torso joint from the Kinect software and tying the main camera's rotation quaternion and player controller movement to these values. This feature gave the user the ability to quickly look one direction, and move to quick reactions in the relatively small area of the Kinect's field of view.

The next goal was to handle long range movement within the simulation, while allowing the user to move small amounts in the real world. For this we chose running and walking in place based on previous research (Whitton, et al., 2005). With this technique, the Soldier can stand in one spot jogging, while navigating large distances inside the simulation. To recognize the movement, we used

$$W = \begin{cases} D_{lhk} - \overline{D}_{lhk} < C \quad \cap \; D_{rhk} = \overline{D}_{rhk} \pm C \\ D_{rhk} - \overline{D}_{rhk} < C \quad \cap \; D_{lhk} = \overline{D}_{lhk} \pm C \end{cases} \quad (1)$$

whereD_{lhk} is the vertical distance between the left hip and knee, D_{rhk} is the vertical distance between the right hip and knee, $\overline{D}$ is the average vertical distance,C is a constant threshold and W is the Boolean value that determines if the gesture has occurred this frame or not. This equation determines if the vertical distance between the knee and hip are decreasing for one leg and not for the other to some threshold away from the mean. The recognition of the gesture occurs before the user has placed their foot down, and thus has minimal latency on the action to reaction in the simulation.

There was also a feature added to perform a detached form of turning. With orientation tied to the torso joint, the user had realistic turning, ideal for a system with a head mounted display, but the prototype made use of a single television screen. As such, the Soldier needed an ability to turn without completely turning away. We chose to implement a bent knee pose, where the user bends either knee to the right or the left to indicate which direction they want the simulation to turn. The amount their knee was bent outward mapped directly to the angular velocity of their turn in the system, giving the user control over the speed of their turn. We also implemented a design with rotation of the hips affecting the angular velocity, but were inhibited by the NITE system's representation of the hip joint's orientation. A final design for evaluation relied on the upper body, which is more visible in the Kinect's field of view than the knees. This involved the user rotating with their shoulders as in the real time turning, and sticking their hand out to signal to RealEdge to hold that turn angle as they moved back. After evaluating each method with subject matter experts, the bent knee pose was preferred as it combined reliability of recognition with a natural gesture.

Tactical Gestures

Tactical gestures were used to recognize maneuvering techniques the Soldier may need, such as jumping and crouching. These gestures were combined with the locomotion techniques stated above so the user can jump forward as well as jump in place.

The jump gesture was determined by equation

$$J = H_y - \overline{H}_y > C \tag{2}$$

where H_y is the Y position of the head, $\overline{H}_y$ is the average position, C is a constant threshold and J is the Boolean value that determines whether the jump has occurred this frame. This equation (2) allows the system to determine if the head has passed some threshold over the average height of the user.

Similarly, the crouch gesture was determined with the heuristic

$$Cr = H_y - \overline{H}_y < C \tag{3}$$

where the variables have the same representation as in (2), with, C being a possibly different constant value and Cr being the Boolean value that determines if a crouch occurred.

We extended our gesture recognition to add a grenade toss gesture. This involves the user reaching to their hip with their right hand and throwing the hand overhead as though a grenade was grabbed and tossed. Implementation achieved by heuristics looking at the distance between the right hand and right hip, followed by how quickly the right hand rose above the head for completion.

Weapon Tracking

Two iterations were performed to achieve accurate weapon orientation and firing. We began by using the Wiimote's infrared position to adjust the position of the simulated gun. The mapping was a scaling value, taking the infrared position to be between [-0.5,0.5] and adjusting that to world coordinates. The trigger on the Wiimote caused the gun to fire in the simulation. Issues immediately surfaced with the range of the infrared transmitter, causing the data to disappear if the Wiimote was moved too far. Furthermore, rotation of the controller was also reflected as position movement in the infrared data, though this can be factored out by using the Wii Motion Plus attachment, which provides angular velocity data, to determine when rotation has occurred.

The second iteration used the Sony Playstation Move.Me system. This software turns the Playstation 3 into a server for the Move recognition and controller data, transmitting it to the computer via the TCP protocol. This data provides the simulator with both spatial information and a quaternion representing the orientation of the Motion Controller. Though not used, angular velocity, controller acceleration, and other data are available. The controller was then attached to the Playstation Sharpshooter accessory to appear more like a rifle and the accessory's trigger caused the simulator's gun to fire. This implementation had much better range and accuracy than the Wiimote, as well as offloading the computation of the position to the Playstation hardware.

EVALUATION

The RealEdge system has been used several times for demonstration purposes at the U.S. Army Research Laboratory Simulation and Training Technology Center (ARL STTC) in the Dismounted Soldier laboratory at STTC. The original system used only the Kinect as a sensor. During the demonstrations and evaluations, the following issues were noticed. First the participant must bring his/her knees up very high when walking for the system to recognize the walking gesture. Also, the Kinect sensor from Microsoft does not track a participant's feet very well. Another main issue is that the system did allow a participant to turn his/her body, but only by using their shoulders. Because of how the NITE software (Primesense, 2011) predicts orientation you cannot use your waist to steer the locomotion. This issue makes it very hard to turn and walk. During the evaluations, a "drive" gesture was also found. If a participant simply raises only one leg the sensor recognizes that the participant is always walking and performs a "drive" mode. Using this "drive" mode enforces negative training as the system is designed to enforce walking which also induces physiological symptoms such as tiredness, fatigue and sweat. If a participant uses the "drive" mode, no physiological symptoms occur.

The second update of the RealEdge system allows a participant to perform the same gestures as before, but also allows running, jumping, jumping over objects and crouching. A "throw grenade" gesture was also added, as well as a gun animation to shoot in the virtual environment. One issue noticed during evaluations and demonstrations is that the gesture used to "throw grenade" is recognized as a jump because the participant's arms raise up. This caused a grenade to be thrown on accident. Otherwise, the gestures work well, allowing for jumping, running (as long as they keep their knees up), walking and crouching with no issues.

The demonstration system has been receiving positive comments from evaluators and on-lookers. Much of this is attributed to the system being natural and not needing to encumber the user with sensors. This system has much promise for the future as the Army is currently researching natural, low-cost means of performing these gestures for future Army programs of record.

CONCLUSION AND FUTURE WORK

We have successfully developed RealEdge, a prototype system capable of tracking a Soldier's natural movements for navigation and locomotion. Furthermore, this prototype utilized commercial off the shelf equipment to achieve a viable combat simulator control without driving up costs or adding extra weight and encumbrances to the user.

The evaluations from the demo installation show us there is still much work to be done in addressing the needs of the simulation. Many of the issues stated are the result of using heuristics for gesture recognition, which is prone to false positives or people being able to perform unintended actions such as the above state "drive" gesture. However, this form of recognition was put into the prototype as a placeholder for a more complex system being developed concurrently. This new system will utilize Hidden Markov Models to recognize several gestures in real time, with continuously streaming data and in parallel on GPUs.

Future work will replace the plastic gun accessory with a realistic modeled gun, with realistic feel and weight distribution. We will be adding nonverbal communication such as hand gestures and support for multiple users at once. Lastly, RealEdge will undergo formative lab testing and summative multi-subject evaluations to prove the training simulation efficacy.

While there remains much work, this prototype presents a system for natural avatar control in training that does not require the cost and intrusion of other virtual environment solutions. This can provide inexpensive realistic dismounted Soldier training at the scale and frequency required for Army needs.

REFERENCES

Bowman, D., Kruijff, E., LaViola, J., &Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Addison-Wesley.

Childs, M. (2010). *Learner's Experience of Presence in Virtual Worlds*. University of Warwick, Institute of Education.

Interrante, V., Ries, B., & Anderson, L. (2007). Seven League Boots: A new Metaphor for Augmented Locomotion through Moderately Large Scale Immersive Virtual Environments. *Symposium of 3D User Interfaces*.

Knerr, B., Lampton, D., Thomas M., Comer, B., Grosse, J., Centric, J., et al. (2003). *Virtual Environments for Dismounted Soldier Simulation, Training, and Mission Rehearsal: Result of the FY 2002 Culminating Event*. U.S. Army Research Institute for the Behavioral and Social Sciences.

Knerr, B (2006). *Current Issues in the Use of Virtual Simulations for Dismounted Soldier Training.* U.S. Army Research Institute for the Behavioral and Social Sciences.

Lee, J. (2008). Hacking the Nintendo Wii Remote. *IEEE Pervasive Computing* , pages 39-45.

Norton, J., Wingrave C., & LaViola J.J. (2010). Exploring Strategies and Guidelines for Developing Full Body Video Game Interfaces.*Proceedings of the Fifth International Conference on the Foundations of Digital Games 2010*, p 155-162.

Peek, B. (2008). *WiimoteLib -.Net Managed Library for Nintendo Wii Remote*. Retrieved September 20, 2009 from http://www.brianpeek.com.

Primesense (2010). *Primsense NITE 1.1 Middleware Datasheet*. Retrieved May 3rd, 2011 from http://www.primesense.com/?p=515

Shirai, A., Geslin, E., &Richir, S. (2007). WiiMedia: motion analysis methods and applications using a consumer video game controller. *SIGGRAPH Sandbox*.

Slater, M., Usoh, M., & Steed, A. (1995). Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, pages 201-219.

Sony (2011). *Move.Me | Playstation Move*. Retrieved May 15th, 2011 from http://us.playstation.com/ps3/playstation-move/move-me/

Unity (2011). *UNITY: Unity 3 Engine Features*. Retrieved May 3rd, 2011 from http://unity3d.com/unity/engine/

Whitton, M., Cohn, J., Feasel, J., Zimmons, P., Razzaque, S., Poulton, S., et al. (2005). Comparing VE Locomotion Interfaces. *IEEE Virtual Reality*, 123-130.

Williamson, B.M., Wingrave C., & LaViola J.J. (2010). RealNav: Exploring Natural Interfaces for Locomotion in Video Games. *Proceedings of 3D User Interfaces 2010*, pages 3-10.

Wilson, A. (2010). Using a Depth Camera as a Table Sensor. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 69-72.

Wingrave, C., Williamson, B., Varcholik, P., Rose, J., Miller, A., Charbonneau, E., et al. (2010). Wii Remote and Beyond: Using Spatially Convenient Devices for 3DUIs.*IEEE Computer Graphics and Applications*, pages (71-85).

Witmer, B., Bailey, J., Knerr, B. (1995). *Training Dismounted Soldiers in Virtual Environments: Route Learning and Transfer*. U.S. Army Research Institute for the Behavioral and Social Sciences.

Zanbaka, C., Lok, B., Babu, S., Xiao, D., Ulinksi, A., & Hodges, L. (2004). Effect of Travel Technique on Cognition in Virtual Environments. *IEEE Virtual Reality*, pages 149-156.