Optimization of Computer Generated Three Dimensional Models for Decreased Latency in Virtual Environments

Varun Aggarwal

Soar Technology, Inc.
Orlando, FL

varunneilaggarwal@gmail.com

Denise Nicholson

Soar Technology, Inc.
Orlando, FL

denise.nicholson@soartech.com

ABSTRACT

Due to demand for realism and high-fidelity experiences by today's users, rendering virtual environments proves to be a computational challenge for lightweight computing platforms (e.g. mobile devices). Traditional simulated environments typically use as much processing power as available to render the entire scene in high detail, limiting simulations to higher end computers. One approach to optimize processing power for three dimensional models is to use varying, decreased, level of detail (LOD) for distant representations (Sik & Pattanaik, 2011). This research attempts to future optimize resources by expanding the adaptive LOD approach based on the object's location in the field of view (FOV) in addition to the object's distance. Such FOV adaptation would take advantage of state-of-the-art head and gaze tracking capabilities. This paper presents results from an initial investigation focused on identifying the minimal LOD that objects can be reduced to before the they become unrecognizable. A simulation was designed that presented randomized sets of objects of various LOD. Subjects were asked to choose an object from the group based on an on-screen prompt. The speed and accuracy of each subject's response was recorded to determine the LOD at which there was no difference in recognition from the full-detail objects. The researchers concluded that the minimum required LOD for recognition without sacrificing speed or accuracy lies between 20% and 80% depending on the shape and distinct features of each object. Specific levels of detail were determined for six objects of different feature complexity to be used in further research studies.

ABOUT THE AUTHORS

Varun Aggarwal is a student at Lake Highland Preparatory School (LHPS) and an intern at Soar Technology. As part of LHPS' Advanced STEM Program emphasizing Internship and Research Experiences (ASPIRE), Varun conducts research under the supervision of Dr. Denise Nicholson. He is interested in computer science techniques to improve low cost Virtual Reality and Augmented Reality displays for a variety of military and commercial human performance applications. Mr. Aggarwal is a member of the Debate Team and HazMat Robotics Team at LHPS, and member of the award-winning Exploding Bacon FIRST Robotics Competition (FRC) Team.

Denise Nicholson, Ph.D., CMSP, is the VP of Intelligent Training at Soar Technology, Inc. She is also the Director of X, leading an effort to eXplore, identify and pursue innovative applications of intelligent systems for critical and challenging problems. Dr. Nicholson has over three decades of experience conducting R&D for military applications in government, academia and industry positions. Denise has a Ph.D. and M.S. in Optical Sciences from the University of Arizona, and a B.S. in Electrical Computer Engineering from Clarkson University. She has contributed to IITSEC as an author and sub-committee member, has over 100 publications, and was co-editor of the PSI Handbook of Virtual Environments for Training and Education [Three Volumes]: Developments for the Military and Beyond.

Optimization of Computer Generated Three Dimensional Models for Decreased Latency in Virtual Environments

Varun Aggarwal

Soar Technology, Inc.
Orlando, FL

varunneilaggarwal@gmail.com

Denise Nicholson

Soar Technology, Inc.
Orlando, FL

denise.nicholson@soartech.com

BACKGROUND

History of Simulated Reality

The idea of creating simulated environments with the goal of mimicking real ones has been around for hundreds, if not thousands, of years. The earliest examples of simulated reality date back to the early 19th century, when artists began to paint 360-degree murals on the walls of circular rooms. The goal of these panoramic paintings was to make a viewer standing inside the room feel as though they were actually inside the painting. These paintings were a poor substitute for reality, but they paved the way for future advancements in simulated reality.

In 1838, researcher Charles Wheatstone demonstrated a simple but incredibly important concept. He proved that showing the right and left eyes a slightly different picture, adjusted for their relative position to the head, created an image that looked three-dimensional. He went on to create the stereoscope, a device that used this principle to allow people to view pictures as three-dimensional scenes in front of them (“History of Virtual Reality”, 2015).

Over 100 years later, the stagnating field of simulated reality was brought back from the brink of collapse by a few major developments. In 1960, the very first simulated reality head-mounted display (HMD) was created, named the Telesphere Mask by its inventor, Morton Heilig. In 1961, the Philco Corporation developed the first HMD with head tracking, which allowed operators to manipulate cameras from afar, giving a sense of presence in areas that were too dangerous to send humans. In 1968, Ivan Sutherland’s Sword of Damocles connected an HMD to a computer for the very first time, providing a basic virtual environment that could show the user any scene that the computer was capable of processing. In 1987, the term virtual reality was coined by Jaron Lanier, founder of the Visual Programming Lab, which sold the very first consumer VR HMD, the EyePhone 1. The EyePhone 1 cost over $9,000 and required a further $9,000 in peripherals, dooming it to failure before it ever hit the market. However, the system succeeded in generating public interest in simulated reality. In 1995, Nintendo released the Virtual Boy, the first VR video game console. A year later, the Virtual Boy was discontinued due to poor sales, likely a result of the uncomfortable headset design and lack of color graphics (“History of Virtual Reality”, 2015).

After the failure of the virtual boy, there was very little major progress in consumer VR, until 2012, when Oculus VR began developing the Oculus Rift. In 2013, Oculus released their first developer kit, the DK1, selling over 50,000 headsets despite being an unfinished prototype. In 2014, Oculus released the DK2, which improved the DK1’s resolution and positional tracking technology. The DK2 was a huge success, with over 100,000 headsets sold (Hayden, 2015). On November 27, 2015, the Gear VR mobile VR headset was released by Oculus, becoming the first successful consumer VR headset. In 2016, the Oculus Rift, HTC Vive, Playstation VR, and Daydream View were all released th the market. In 2016, sales for these headsets reached a combined total of over 6 million units, proving the consumer viability of VR (Durbin, 2016).

Current State of Simulated Reality

In the field of simulated reality, there are two major markets driving the industry. The first is to provide education and entertainment to the public. With VR, people can freely travel around the world and even back in time, experiencing the world in a way that would not be possible otherwise. Students can attend classes on other continents, or even bring the top professors in the world right into their own homes or classrooms. Another growing market is in the area of providing effective and realistic training programs for doctors, soldiers, and other professionals. In a virtual environment, a doctor can practice performing difficult operations without risking the safety of a patient, while a soldier can gain the combat experience necessary to stay alive in a fight (Nicholson, 2009).

In order to accomplish both of these goals, simulated environments need to be as realistic as possible. However, as these environments, and the digital models they are composed of, become more realistic, they also become increasingly difficult to render. The increased level of detail requires larger files with more data to be processed. This, in turn, requires more powerful hardware to run the necessary calculations. The level of graphical processing capability required operate a VR headset that meets the current consumer standard is far beyond that of any computer that is readily available to the public. The monetary resources required to build or purchase such a machine makes widespread, consumer-ready VR impossible at the moment. In time, this problem will become even more pronounced, as advancements in the technological capability of VR headsets will require even more detailed simulated environments (Cohn, 2009).

Due to demand for realism and high-fidelity experiences by today's users, rendering virtual environments within acceptable latency limits proves to be a computational challenge for lightweight computing platforms (e.g. mobile devices). Traditional simulated environments typically use as much processing power as available to render the entire scene in high detail, limiting simulations to higher end computers. One approach to optimize processing power and reducing latency for three dimensional models is to use varying, decreased, level of detail (LOD) for distant representations (Sik & Pattanaik, 2011).

This research attempts to future optimize resources by expanding the adaptive LOD approach based on the object's location in the field of view (FOV) in addition to the object's distance. Such FOV adaptation would take advantage of state-of-the-art head and gaze tracking capabilities. With the development of new VR headsets that are capable of tracking eye movements, it is now feasible for rendering engines to use a process called foveated rendering, where objects outside of the fovea are not rendered at full detail ("Perceptually-Based Foveated Virtual Reality", 2017).

PURPOSE

Long Term Goal

The goal of this research is to reduce the amount of processing power required to render a virtual environment. This project is the first phase of a three-year plan to design a method of rendering virtual environments with increased efficiency, by reducing the LOD of objects outside the user's central gaze area. There has been research conducted on decreasing the resolution of the parts of the environment seen by peripheral vision, but a software-only method incorporated within the rendering engine does not exist. The overarching goal of this three-year project is to determine how the presence of less detailed objects outside of the user's central gaze effects the quality and perceived realism of a simulation. The research presented in this paper is the first phase of the project, with the goal of identifying the point at which specific objects have the minimum possible detail and are still recognizable. If the decreased detail has no effect on the quality of a simulation, then the new method of rendering can be applied to any virtual environment, in order to increase efficiency and allow simulations to be run using less-powerful hardware. The primary impact of this development would be a reduction of financial barriers to consumer VR technology, expanding the user base of VR beyond enthusiasts to average consumers.

Short Term Engineering Goal of Phase 1

The engineering goal of this project is to create three dimensional models of six specific objects at the minimum level of detail required for them to be recognized.

- The experimental models should have a maximum of 80% of the detail of the initial models.
- There should be no statistically significant difference between the time and accuracy of recognition of the experimental models and the initial models.

METHODS

A simulation was designed that presented randomized sets of objects of various LOD. Subjects were asked to choose an object from the group based on an on-screen prompt. The speed and accuracy of each subject's response was recorded to determine the LOD at which there was no difference in recognition from the full-detail objects.

Object Collection

All of the objects used in this research were taken from the Unity 3D Asset Store (Figure 1). Using the "Free" and "Most Popular" filters, the most commonly used objects that were available for public use were identified, and the top fifty distinct types of objects were recorded and downloaded. In order to prevent duplicates that would confuse the participants and interfere with the results, objects of the same type, such as two different variations of a "car" object, were removed, and replaced with objects further down the list. After the identification of the top fifty objects, six were chosen as experimental models, with the remaining forty-four serving as background objects. These six objects, which included an apple, desk, street light, tiger, tree, and vase, were chosen based on three factors. They each represented a unique shape with varying feature complexity, a unique category of object, and were the most used objects in their categories.

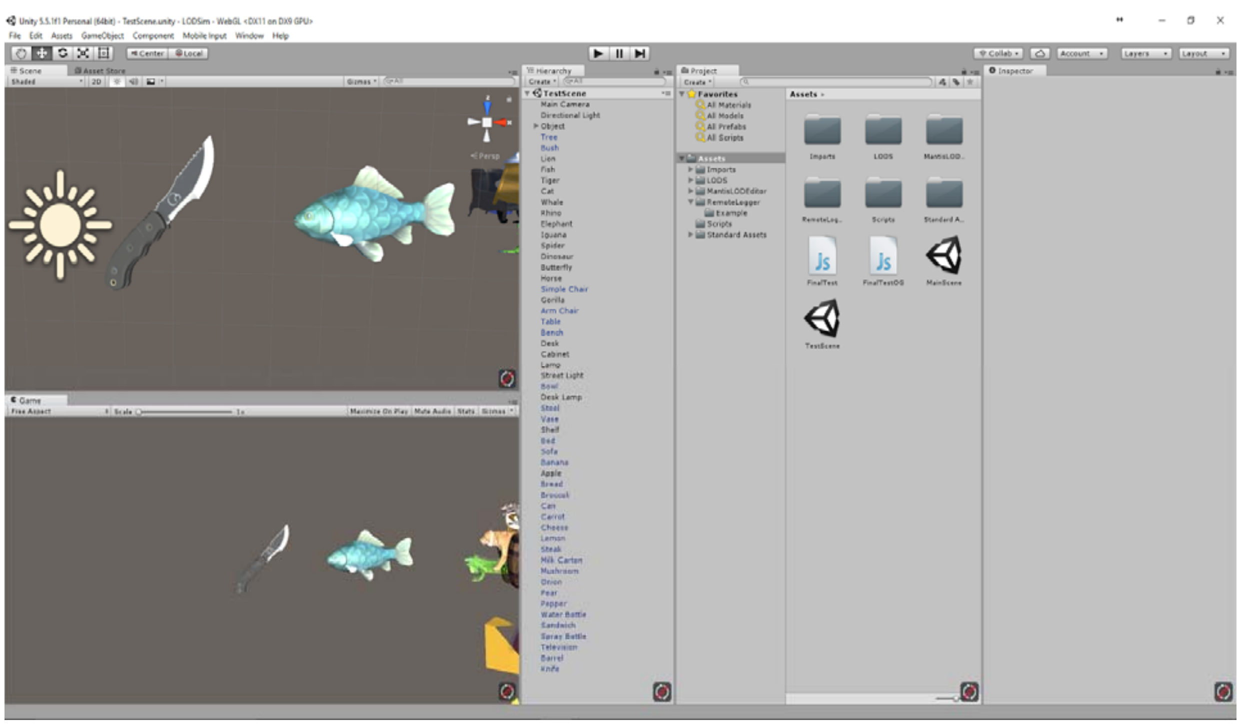

Figure 1: Creation of Simulation in Unity

After all of the objects had been chosen and sorted into experimental and background categories, the researchers took all of the objects composed of multiple meshes and replaced them with similar objects composed of a single mesh, in order to reduce the complexity of the objects and ensure accurate results. The final fifty objects were then loaded into the Unity 3D Engine, for use in simulations. Each object was run through the Mantis Integrated Unity Level of Detail (LOD) tool five times, a software tool that reduces detail by collapsing polygons in an approximately linear scale, to create six different copies of each object. As shown below in Figure 2 for a Fish, the copies each had a different level of detail, from zero percent detail (LOD Group 0) to one hundred percent detail (LOD Group 5), in increments of twenty percent. These three hundred objects were sorted into six different groups based on their level of detail, so that group zero had fifty objects with zero percent detail, and group five, the control group, had fifty objects with one hundred percent detail.

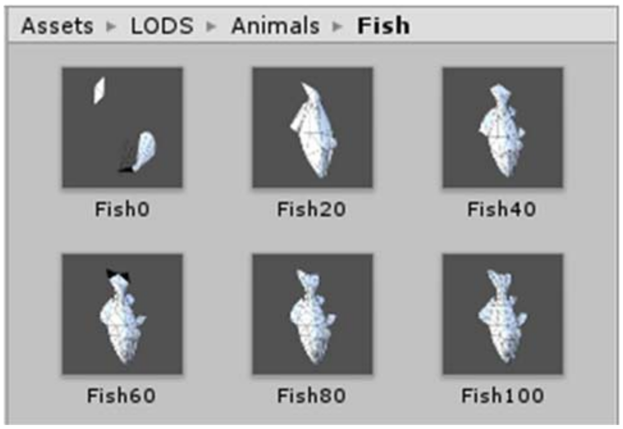

Figure 2: Levels of Detail for "Fish" Object

Simulation Design

Using the Unity 3D Editor, a simulation that displayed random sets of objects was written in JavaScript. The simulation was designed to display a set of nine randomly selected background objects and one experimental object in a 5 x 2 grid. At any given time, all ten objects had the same level of detail, with each level of detail being used six times, in a random order. Each experimental object would also be used six times, in an order determined by the level of detail, so that each level of detail for each object would be displayed exactly once, for a total of thirty-six sets of objects. The simulation included a prompt at the top of the screen, identifying the experimental model by name and asking users to click on it.

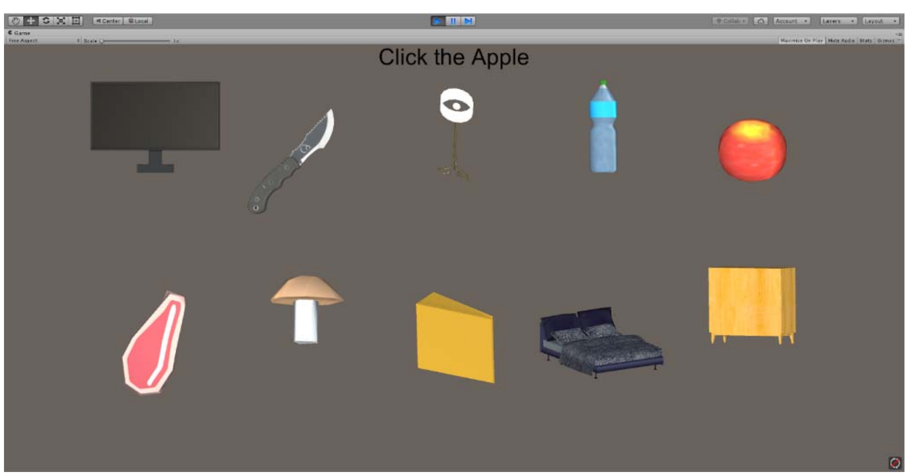

Figure 3: Final Simulation

The amount of time elapsed before the participant clicked and the accuracy of the selection were written to the debug log, along with the name and level of detail of the experimental object. Using preexisting open source code for c# and Google Scripts, a logging system that took all entries from the debug log and wrote them to a google sheets file was created. After the completion of the simulation, it was compiled as a web application using Unity's built in compiler and uploaded to GitHub, where it was hosted on GitHub Pages as a web applet accessible to anyone with the link.

Data Collection and Analysis

In order to collect the large amounts of data required for statistically significant results, a Human Intelligence Task(HIT) was posted on Amazon Mechanical Turk, a platform for crowdsourcing data, asking workers to fill out a demographics questionnaire and give consent before running the simulation on a personal computer. In addition to these workers, emails containing the same set of instructions were sent out to some of the researcher's contacts, who had expressed interest in taking part. After data was collected from one hundred different participants, a single spreadsheet containing each data point was compiled. Due to the low upload speeds experienced by some participants, a portion of the data was not uploaded correctly and had to be deleted. After ensuring that all remaining data was valid, the raw data from the spreadsheet was compiled and entered into the Minitab statistical analysis software. A binary logistic regression comparing each data set to the control was run in order to determine the minimum level of detail required for recognition accuracy on par with that of the control, for each of the six experimental models. Additionally, an ANOVA test with a Tukey pairwise comparison was run to determine the minimum level of detail required for recognition speed on par with that of the control, for each of the six models.

RESULTS

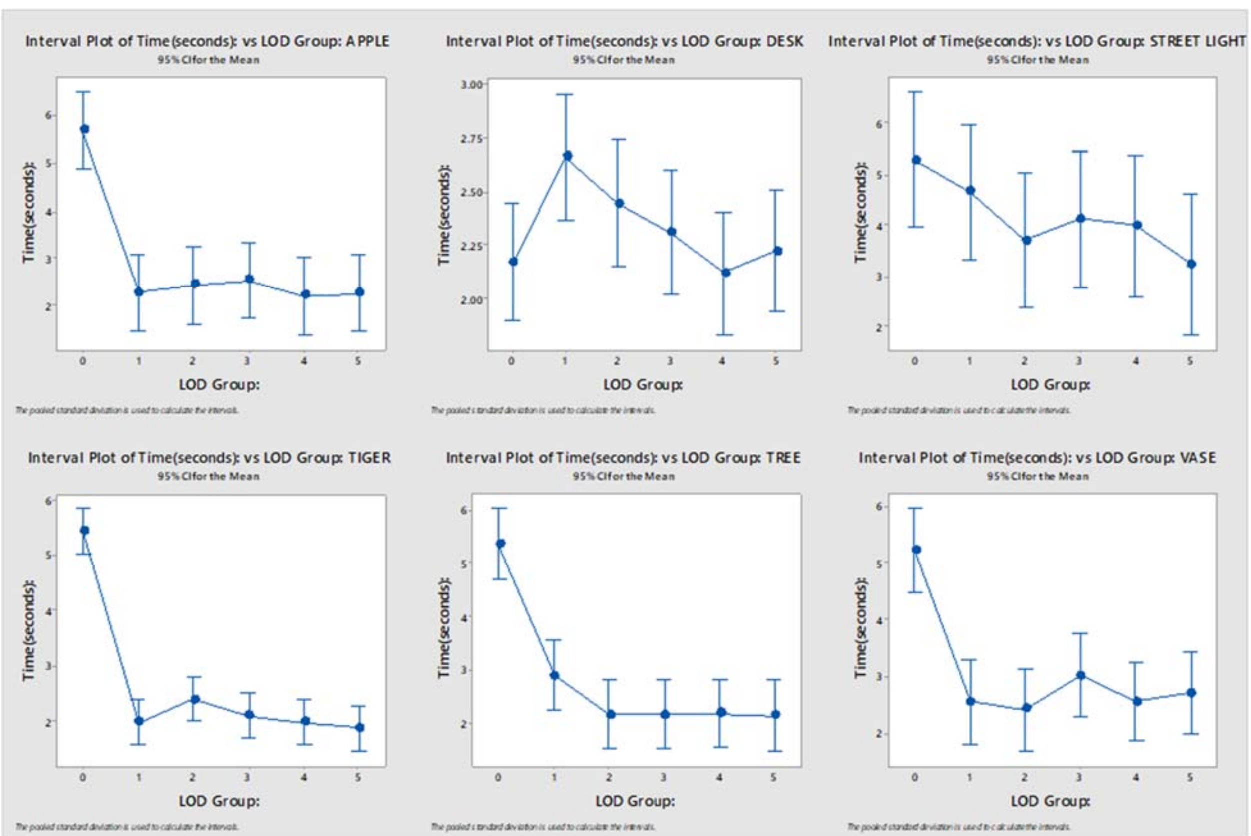

Figure 4: Response Time by Level of Detail (LOD)

As shown below in Table 1, the speeds of response of LOD Groups 1 and higher all had P-value >/=0.05 to relation to the speed of response of the control, LOD Group 5. In addition, the Desk and Street Light objects response times were similar even when reduced to minimum detail in Group 0.

Table 1: Tukey Test P-value Results for Speed of Response Compared to Control

Object	All	LOD 0 vs. LOD 5	LOD 1 vs. LOD 5	LOD 2 vs. LOD 5	LOD 3 vs. LOD 5	LOD 4 vs. LOD 5
Apple	0.000	0.000	1.000	1.000	0.998	1.000
Desk	0.098	1.000	0.279	0.891	0.998	0.997
Street Light	0.341	0.271	0.679	0.996	0.943	0.972
Tiger	0.000	0.000	0.999	0.449	0.969	0.999
Tree	0.000	0.000	0.618	1.000	1.000	1.000
Vase	0.000	0.000	1.000	0.993	0.992	1.000

Table 2 shows the accuracy of recognition of each group, and compares each to the accuracy with which participants recognized the objects at 100% detail in LOD Group 5. The cells highlighted in yellow represent an accuracy of above 90% of the accuracy of Group 5, while the cells highlighted in green represents accuracies above 95%.

Table 2 Accuracy of Recognition by LOD Regression Prediction

Object	LOD 0	LOD 1	LOD 2	LOD 3	LOD 4	LOD 5
Apple	0.428	0.670	0.847	0.938	0.976	0.991
Desk	0.806	0.855	0.893	0.923	0.944	0.960
Street Light	0.409	0.544	0.673	0.780	0.859	0.913
Tiger	0.339	0.762	0.952	0.992	0.998	0.999
Tree	0.624	0.778	0.881	0.940	0.971	0.986
Vase	0.606	0.751	0.855	0.920	0.958	0.978
(yellow)	= Predicted accuracy is 90% of Predicted accuracy at LOD 5			(green)	= Predicted accuracy is 95% of Predicted accuracy at LOD 5	

The results of this study show that the minimum level of detail (LOD) required for speed and accuracy to be identical to that of the control, assuming a required match rate of 90%, is:

- Apple: 60% (LOD Group 3)
- Desk: 40% (LOD Group 2)
- Street Light: 80% (LOD Group 4)
- Tiger: 40% (LOD Group 2)
- Tree: 60% (LOD Group 3)
- Vase: 60% (LOD Group 3)

If the required match rate is increased to 95%, the following levels of detail are required:

- Apple: 80% (LOD Group 4)
- Desk: 60% (LOD Group 3)
- Street Light: 100% (LOD Group 5)
- Tiger: 40% (LOD Group 2)
- Tree: 60% (LOD Group 3)
- Vase: 80% (LOD Group 4)

CONCLUSION

The results of this study demonstrate that it is possible to significantly reduce the level of detail required to recognize virtual objects, with polygon reductions of up to 60% in some cases. This means that object-based foveated reality has the potential to reduce the processing cost of virtual environments by a similar margin, if future studies verify similar recognition results when the object is outside of the central fovea gaze. Using the data gathered in this research, the overall hypothesis of this project can be tested, showing the potential for object-based foveated reality to reduce the hardware requirements of virtual environments and lower the cost barrier to consumer virtual reality.

Future Application

The next step in this research involves using these models in a simulation that places them outside of the central fovea gaze, and surveying participants about simulation comfort, realism and quality of immersion. The expected simulated experience will focus the subject on conducting a primary task which is occupying the central field of view, with the lower LOD objects used as background items. If comfort and immersion are not impacted, this method of foveated rendering can be implemented in other VR simulations. The models that correspond to each of these objects and LOD have been stored in a public GitHub repository for future use in LOD research and practical application. A match rate of 90% is acceptable for most consumer applications, but 95% should be used for industrial or scientific purposes.

REFERENCES

Cohn, J., Nicholson, D., & Schmorrow, D. (2009). *Virtual Environments for Training and Education: Integrated Systems, Training Evaluations, and Future Directions*. Westport, CT: Praeger Security International.

Conditt, Jessica. "Use Your Fingers to Play in Vive's World with the Manus VR Glove." *Engadget*. 14 July 2016. Web. 06 Feb. 2017.

Durbin, Joe. "VR Sales Analysis: Rift and Vive to Sell Under 500,000 in 2016, PS VR to Top 2 Million." *UploadVR*. UploadVR, 14 Nov. 2016. Web. 15 Feb. 2017.

Hayden, Scott. "Oculus Reveals More than 175,000 Rift Development Kits Sold" *Road to VR*. 12 June 2015. Web. 15 Feb. 2017.

"History Of Virtual Reality." *Virtual Reality Society*. 10 Jan. 2015. Web. 15 Feb. 2017.

"Perceptually-Based Foveated Virtual Reality" *Nvidia Corporation*. Web. 06 Feb. 2017.

"FOVE Shows Glimpse of Foveated Rendering Progress." *Road to VR*. 13 Feb. 2016. Web. 06 Feb. 2017.

Nicholson, D., Schmorrow, D., & Cohn, J. (2009). *Virtual Environments for Training and Education: VE Components and Training Technologies*. Westport, CT: Praeger Security International.

Sik, L.L., Pattanaik, S. (2011). Rendering Large Scale Synthetic Urban Terrain. In *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL*.