**Limitations of commodity tracking devices in virtual environments**

Eliot Winer, PhD1     Timothy Morgan1      Justin Gosselin2       Maya Hughes3       Allison Smith4

1Iowa State University, Ames, IA

2Florida Gulf Coast University, Fort Myers, FL

3Jackson State University, Jackson, MS

4Wilkes University, Wilkes-Barre, PA

**Abstract**

Virtual reality has seen increasing interest in the past twenty years with applications ranging from entertainment to medical imaging. While most virtual reality applications require expensive, task-specific equipment, there has been a movement toward the use of commodity devices for a variety of applications. Devices such as the Oculus Rift and the Microsoft Kinect have gained popularity in not just the entertainment industry, but also scientific and academic communities. These devices are inexpensive, easy to use, and suitable for many applications. However, these devices often are marred by high latency, poor accuracy, and other limitations. In an effort to determine the feasibility of using such devices for controlling virtual environments, this study examines the limiting factors of commodity devices using the Microsoft Kinect. The latency and accuracy are measured and analyzed using a simple method based on elementary kinematics.

**Introduction**

The use of commodity tracking devices has increased in recent years due to their affordability and ease of use. Although commodity devices can be used for entertainment, there are other demands that the commodity devices can accommodate such as virtual caves and medical preparations. These devices often include a variety of interaction modules, ranging from voice recognition to skeletal tracking [1]. This would introduce a new way of creating a virtual environment at a cheaper cost with the open source domains and other easy-to-replace hardware. With enough research from testing and experimenting, it is possible that commodity tracking systems could possibly surpass the more expensive devices [2]. The overall scope of this research is to help build on a research conducted by Morgan, et al. in the use of networked Kinects to increase the coverage area in virtual environments [13].In the future, this will be used to enhance scientific visualization. Specifically, it will be used in constructing a 3D virtual environment to visualize x-ray systems rather than degrading the x-rays to 2D images. Using these body-based, commodity tracking devices would require little to no prior knowledge about the virtual environment and the Microsoft Kinect. Because the Kinect does not require any restrictive gear, it will allow individuals, who are not experts in virtual reality, to have a significant immersive experience. This leads to the research question: What are the limitations and advantages of using body-based, commodity tracking devices for controlling virtual environments? During the literature review, we found that latency, accuracy, and crosstalk [when using multiple devices] are some of these limitations within the Microsoft Kinect System.

A major limitation that arises with commodity devices is dynamic tracking errors. Azuma [5] defines dynamic tracking errors in head-mounted displays as errors that “have no effect [on the user] until either the viewpoint or the objects begin moving.”  End-to-end latency is caused by a variety of factors, including transmission delay and the update of double-buffered displays [6,10]. Examples include end-to-end latency, which is the time between the user’s input and the related output. In a virtual reality setting, this latency becomes problematic when there is a noticeable delay between a user’s action and the corresponding visual change [6]. Also, it impacts effectiveness of virtual environments, hampering the quality of commodity devices [1,7]. Research shows that errors in virtual environments can break a user’s immersion and cause disorientation [6]. This effect of dynamic error,latency, is known to cause simulator sickness, oscillopsia, nausea, and issues in perceptual stability [6-8]. Studies conducted by Moss, et al. [9] found that the threshold mean for latency detection in head-mounted displays is 147.64 ms (SD = 84.91).Other studies have shown that this threshold value is as low as 14.3 ms (SD = 2.7 ms) [10]. This study aims to explore the limitations in body-based tracking devices due to threshold values of latency. While previous research has focused on the latency of head-mounted displays [6-10], this study evaluates the latency of body-based, commodity device, e.g., Microsoft Kinects.

Although dynamic tracking errors pose difficulties in virtual environments, they are not the sole problem with commodity tracking devices. Static tracking errors, defined by Azuma [5] as mistakes that “cause registration errors even when the user’s viewpoint and the objects in the environment remain completely still”, are also a problem. These include issues such as crosstalk and positional accuracy. The introduction of crosstalk caused by multiple Kinects is considered when the eyes detect identity switches and mismatching from of images occurs within the devices [1]. Crosstalk is problematic when it affects image quality, depth quality, and causes image discomfort [11]. In the methods section of this paper, the positional accuracy of the Microsoft Kinect v1 and Kinect v2 sensors is examined.

It has been shown through other studies, that the Kinect is accurate when detecting large movements, but not very accurate for small movements [Galna, et al., 2014]. If clasping your hands is considered to be a small movement, how could the low accuracy affect the way people have to interact with a virtual environment? Would it be enough to clench a fist in front of you in order to grab a virtual object? Would the user have to perform a bigger action instead? These questions are important to determine because in order to control a virtual environment, motions will have to be made by the user, be interpreted by the program, and seen accurately enough by the Kinect. In the methods section of this paper, the positional accuracy of the Microsoft Kinect v1 and Kinect v2 sensors is examined.

Overall, this study aims to analyze the feasibility of using commodity body-based tracking devices to control virtual environments. By investigating the limitations of these devices, the usefulness of these devices for virtual reality applications will be ascertained. Examining the advantages of using these affordable devices instead of expensive, high-powered devices will also help discern the feasibility of using these devices. This is achieved by analyzing important issues in commodity devices, including latency, accuracy, coverage area, and other aspects. Specifically, this study incorporates the use of the Microsoft Kinect v1 and v2, as these are very common commodity, body-based tracking devices.

**Methods**

In order to determine the limitations of the Microsoft Kinect v1 and Kinect v2, it is necessary to decide what factors are important to study. Since latency has been shown to have an effect when dealing with Kinects [4], a test will be performed to determine those values for both versions of the Kinect. Initial experimentation will be done with a single Kinect. This will allow us to examine the feasibility of using Kinects for virtual environments. When performing this test for a single Kinect, an almost-frictionless ramp will be used and compared against basic physics. A metal ball will be rolled down a ramp (2.4 meters), and the motion will be captured using the Kinect. A high speed camera will be used to record the ball rolling down the ramp, as well as the screen on which the Kinect is outputting its data to. Physics will be used to determine how much time it should take for the ball to reach a certain point on the ramp, and the high speed camera will be used to determine the time difference between when the ball should reach that point, and when the Kinect registers that the ball is at that point. The difference between these times will be the latency.

Two tests will be conducted, for each Kinect version, in order to determine the positional accuracy for each. The first test will be measuring the accuracy by using a color image from the Kinects to find the position of a static, 2D mark. The second test will measure the accuracy of the depth camera to determine the position of a static, 3D object. These tests will use the same setup as the latency experiments, however, only the Kinect camera will be used, since they will be static experiments. The 2D mark and 3D object will be placed on the ramp, and the actual distance from the mark/object to the camera on the Kinect will be measured and used for comparison to the position found by the Kinect. A still frame will be captured with the Kinect camera, and then the position from the Kinect will be found using the resolution (pixels) and scope/field of view of the cameras. The Kinect v1 has a resolution of 320 x 240 for the depth camera and 640 x 480 for the color camera, and it has a scope of 57° x 43°. The Kinect v2 has a resolution of 512 x 424 for the depth camera, and 1920 x 1080 from the color camera, and it has a scope of 70° x 60°. By knowing the scope of each, we can determine how much area each pixel from the Kinect image takes up. This will then allow us to determine the position of that mark/object based on some reference point.

Expanding on research previously conducted by Morgan, et al. [13], a demo application will be created that allows users to interact in a virtual environment controlled by multiple Kinects, ideally in a cave scenario. This application will be created in a manner that will allow the user to complete a task. The app may be edited to include a variable latency that will allow for testing user effect due to latency. Similar to the study conducted by Meehan, et al. [7], a demanding virtual environment will be created. Users will be tasked with using gestures to throw a virtual ball down a virtual pit. Before user testing the demo application, crosstalk of networked Kinects will have to be addressed and analyzed.

Although multiple Kinects can be linked together as a network, there are other possible limitations such as crosstalk. In a previous research there has been a study where a gentle shake of the Kinect unit can reduce the crosstalk [14]. This formed an experiment in which a team developed a vibrating motor that ranged from 15Hz to 120Hz to cause a motion blur that reduces much of the cross talk. There were rubber bands being used to allow mounting and free vibrations from the motors. There were different ways of testing where the team used a vibrating motor hot glued on top of the Kinect. These methods can be used for or more Kinects, if there is a decision to network them, and Velcro to stabilize the Kinect.

**Acknowledgements**

This material funded in part by National Science Foundation Grant CNS-1461160. We would like to give thanks to Stacy MacAllister, Paul Easker, and Glen Galvin for their help with providing/finding equipment for our testing.

**References**

[1] K. Ponto, J. Kohlmann and R. Tredinnick, "DSCVR: Designing a Commodity

Hybrid Virtual Reality System," *Springer-Verlag,* vol. 19, no. 1, pp. 57-70, 2014.

[2] T. Schou and H. J. Gardner, “A Wii remote, a game engine, five sensor bars and a virtual reality theatre,” Proc. OZCHI’07, CHISIG Annu. Conf. Human-Computer Interact., pp. 231–234, 2007.

[3] B. Lange, S. Koenig, C.-Y. Chang, E. McConnell, E. Suma, M. Bolas, and A. Rizzo, “Designing informed game-based rehabilitation tasks leveraging advances in virtual reality,” Disabil. Rehabil., vol. 34, no. 22, pp. 1863–1870, 2012.

[4] M. A. Livingston, J. Sebastian, Z. Ali, and J. W. Decker, “Performance measurements for the Microsoft Kinect skeleton,” *2012 IEEE Virtual Real.*, pp. 119-120, 2012.

[5] R. T. Azuma, “A Survey of Augmented Reality,” Presence Teleoperators Virtual Environ., vol. 6, no. 4, p. 355, 1997.

[6] R. S. Allison, L. R. Harris, M. Jenkin, U. Jasiobedzka, and J. E. Zacher, “Tolerance of temporal delay in virtual environments,” Proc. IEEE Virtual Real. 2001, pp. 1–8, 2001.

[7] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks, “Effect of latency on presence in stressful virtual environments,” Proc. - Virtual Real. Annu. Int. Symp., vol. 2003, pp. 141–148, 2003.

[8] A. Ash, S. Palmisano, D. G. Govan, and J. Kim, “Display lag and gain effects on vection experienced by active observers,” Aviat. Sp. Environ. Med., vol. 82, no. 8, pp. 763–769, 2011.

[9] J. D. Moss, E. R. Muth, R. a. Tyrrell, and B. R. Stephens, “Perceptual thresholds for display lag in a real visual environment are not affected by field of view or psychophysical technique,” Displays, vol. 31, no. 3, pp. 143–149, 2010.

[10] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill, “Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity,” Proc. 1st Symp. Appl. Percept. Graph. Vis. - APGV ’04, no. study 1, p. 39, 2004.

[11] H. Sohn, Y. J. Jung and Y. M. Ro, "Crosstalk Reduction in Stereoscopic 3D Displays: Disparity Adjustment Using Crosstalk Visibility," *Optics Express,* vol. 22, no. 3, pp. 3375-3392, 2014.

[12] C. Ellis, S. Z. Masood, M. F. Tappen, J. J. Laviola, and R. Sukthankar, “Exploring the trade-off between accuracy and observational latency in action recognition,” Int. J. Comput. Vis., vol. 101, no. 3, pp. 420–436, 2013.

[13] T. B. Morgan, D. Jarrell, and J. M. Vance, “Poster: Rapid development of natural user interaction using kinect sensors and VRPN,” 2014 IEEE Symp. 3D User Interfaces, pp. 163–164, 2014.