3D Spatial Interaction with the Wii Remote for Head-Mounted Display Virtual Reality

Yang-Wai Chow

Abstract—This research investigates the design of a low-cost 3D spatial interaction approach using the Wii Remote for immersive Head-Mounted Display (HMD) virtual reality. Current virtual reality applications that incorporate the Wii Remote are either desktop virtual reality applications or systems that use large screen displays. However, the requirements for an HMD virtual reality system differ from such systems. This is mainly because in HMD virtual reality, the display screen does not remain at a fixed location. The user views the virtual environment through display screens that are in front of the user’s eyes and when the user moves his/her head, these screens move as well. This means that the display has to be updated in real-time based on where the user is currently looking. Normal usage of the Wii Remote requires the controller to be pointed in a certain direction, typically towards the display. This is too restrictive for HMD virtual reality systems that ideally require the user to be able to turn around in the virtual environment. Previous work proposed a design to achieve this, however it suffered from a number of drawbacks. The aim of this study is to look into a suitable method of using the Wii Remote for 3D interaction in a space around the user for HMD virtual reality. This paper presents an overview of issues that had to be considered, the system design as well as experimental results.

Keywords—3D interaction, head-mounted display, virtual reality, Wii remote.

I. INTRODUCTION

Virtual reality and 3D virtual environments are becoming increasingly popular and its applications span a variety of different areas. These range from applications like scientific and medical visualization, simulation and training to video games. In fact, video games have been one of the main driving forces behind the many advancements made toward improving virtual environment technology. Much of this progress has focused on areas such as the development of increasingly powerful Graphics Processing Units (GPUs) for the generation of real-time 3D computer graphics. However, the basic interface between humans and gaming systems has received relatively little attention [1].

Interacting with the 3D content present in games and virtual environments generally involves some form of 3D interaction. 3D interaction has been defined in [2] as human-computer interaction in which the user’s tasks are performed directly in a 3D spatial context. This however does not necessarily involve 3D input devices. Conventional 2D input devices like the keyboard and mouse, or the standard gamepads that have numerous buttons and analog controllers are probably not ideal, and certainly not intuitive, for interaction in a 3D spatial context [1]. Furthermore, the unnatural mapping between these 2D devices with 3D content to some extent reduces the user’s immersive experience [3]. Thus, this highlights the impact that the development of 3D spatial interaction devices and techniques can offer to 3D virtual environment interaction, via more natural and intuitive human expression.

While research in virtual reality and 3D User Interfaces (UIs) – user interfaces that involve 3D interaction [2] – has been around for many years, it has met with limited success [4]. However, with the recent advent of the Nintendo Wii there has been a strong push in the direction of 3D interaction devices for games and virtual environments. The Wii’s video game controller, the Wii Remote (informally known as the ‘Wiimote’), presents players with an innovative way of interacting with game content. The somewhat simple yet effective optical tracking and motion sensing technology provided by the Wii Remote has given rise to many interesting interaction possibilities. This has in turn revolutionized the way in which certain video games are being developed and played.

While to date Nintendo has not released any official technical specifications about the technology contained within the Wii Remote, there have been many interested parties among the general public that have tried to reveal and share information about how this game controller operates. In particular, the global hacking community has managed to reverse-engineered many aspects of the Wii Remote [5]. As a result, much of the technical information about the inner workings of this game controller appear in a number of websites and online wikis [6]–[8].

This readily available information shows how easily the Wii Remote can be connected to a computer via a Bluetooth connection. Furthermore, various software libraries have also been written that allow easy access to the features offered by this video game controller [9]. Consequently, many people have applied the Wii Remote to a myriad of applications which make use of a variety of interaction techniques, without making use of the Wii game console.

A number of researchers have also adopted the Wii Remote for various purposes, such as for gesture recognition based applications [3], [10], [11], robot control [12], motion capture [13], and many others [14]–[22]. While a variety of optical

Yang-Wai Chow is with the School of Computer Science and Software Engineering, University of Wollongong, NSW, Australia (e-mail: caseyc@uow.edu.au).
tracking and motion sensing devices have surfaced over the years, the advantage of the Wii Remote is that it is a low-cost wireless device that combines an infrared sensor with accelerometers, vibration feedback, a speaker and a variety of buttons all within a single device. In addition, the infrared sensor specifications in the Wii Remote generally outperform comparably priced webcams in terms of refresh rate and resolution [5]. Furthermore, the game controller can also be connected to a number of other low-cost extensions, like the ‘Nunchuk’ extension, which in addition to 2 buttons and a control stick also has similar motion-sensing technology [23]. This makes the device extremely flexible, as the data it outputs can be interpreted in a variety of ways depending on the intended context [1].

Normal 3D input devices like 3D data gloves, 6 degrees-of-freedom (DOF) sensors and trackers, etc. come with a heavy price tag. As such, this makes the Wii Remote attractive as a low-cost 3D spatial input device. In fact, its low-cost has been the main driving factor behind much research effort [18]–[20]. The cost of high-end virtual reality systems has long been a factor that has hindered its use in mainstream society [24]. As noted in [4], having input devices which can track 6-DOF that are in the price range of the Wii Remote will go a long way to improving 3D spatial interaction and providing much more realistic and immersive virtual environment experiences for the general public [4]. In addition, given its widespread popularity, it is a device that is familiar to many people.

The aim of this study is to design a system using the Wii Remote as a low-cost 3D input device that is suitable for use in an immersive Head-Mounted Display (HMD) virtual reality system. Previous work investigated a number of approaches to achieving this, and also identified various design limitations that this study attempts to address [25].

II. BACKGROUND

This section gives some background to the purpose and motivation behind this research. It also presents a brief overview of 3D user interfaces and interaction techniques, and outlines previous work as well as some of the issues that had to be considered when attempting to use the Wii Remote as a 3D input device in a space around the user.

A. HMD Virtual Reality

There are a number of existing virtual reality applications that make use of the Wii Remote. However, the current virtual reality applications that make use of this game controller are either desktop virtual reality applications [5] or virtual reality systems that involved the use of large screen displays [14], [17], [21], [22]. The requirements for Head-Mounted Display (HMD) virtual reality differ significantly from such systems, primarily because the location of the display screen is not fixed. The user views the virtual environment through display screens that are in front of the user’s eyes and when the user moves his/her head, these screens move as well, and the display has to be updated in real-time based on the user’s head position and orientation. This means that unlike fixed location displays, the user can interact with the virtual environment by physically turning around.

While the Wii Remote can be used for head-tracking in desktop virtual reality [5], head-tracking accuracy is vital in immersive HMD virtual reality systems. In addition, it is necessary for the system to have very low latency, otherwise the user may suffer from a variety of adverse side effects. These adverse side effects have been well documented [26]. In that respect, the Wii Remote is not really suitable for head-tracking in immersive HMD virtual reality. This is because the data outputs from the Wii Remote are not particularly stable even when the device remains stationary. Filtering the raw readings from the controller can reduce the amount of jitter, but at the same time introduces a lag.

Nevertheless, it is possible that the game controller can be used as a hand-held 3D interaction device for applications where perfect accuracy is not essential. Moreover a human user cannot really hold a hand-held device perfectly stationary. In that respect, it is possible that slight inaccuracies or lag in the data outputs may not impact user satisfaction. For example, from the user’s perspective in an application like a virtual reality game, slight inaccuracies and delays might be tolerable as long as it does not impede user task performance in the virtual environment.

There are a number of issues that this study attempts to address. Firstly, in order to adequately interact with the surrounding virtual environment in HMD virtual reality, the user should ideally be able to use the 3D input device in a 360 degree space in the horizontal plane around the user. The conventional Wii Remote setup does not allow for this, because in order to use the controller as a pointing device, normal usage requires it to be pointed in the direction of the ‘sensor bar’. The sensor bar is typically placed at a fixed location either above or below the display device (TV or monitor). This severely restricts the device’s interaction scope. Another issue that needs to be addressed is that a 3D spatial input device should ideally be able to detect 6-DOF, however the Wii Remote cannot reliably detect this when used in the conventional manner.

B. 3D User Interfaces

3D User Interfaces (UIs) are seen as a class of technology that can bridge the gap between 3D interaction and natural human expression [1], [27]. 3D UIs are defined as involving input devices and interaction techniques for effectively controlling 3D computer generated content [4]. Three basic approaches to interaction through input devices are described in [4], and while the described approaches were in regard to video game interaction, they also apply to the broader sense of interaction in virtual environments.

The first approach involves mapping 2D input and button devices, e.g. keyboard and mouse, to elements in the 3D virtual environment. The second method attempts to mimic the real world via 3D input devices that are replicas or physical props of real world devices like steering wheels and musical instruments. Whereas the third approach is true spatial 3D
tracking of the user’s motion and gestures, where users interact in and control elements in the 3D virtual environment with their bodies. LaViola [4] also argues that the second and third approaches hold the most promise in the next level of innovation. As such, these are the approaches that are adopted for the purpose of this study.

As for 3D interaction techniques, these typically consist of the so-called ‘universal 3D tasks’ which involve selection, manipulation, navigation and control system tasks in the virtual environment. These techniques are the fundamental building blocks of 3D user interfaces [2]. This study attempts to address the common Wii game interactions, namely selection and navigation [5] in the design of the system.

C. Design Considerations and Previous Work

The motion sensing technology contained within the Wii Remote consists of 3 linear accelerometers which are oriented along 3 orthogonal axes to sense acceleration along the three axes. Unlike fully self-contained inertial sensing devices which require 3 accelerometers and 3 gyroscopes to determine position and orientation [28], the Wii Remote does not contain any gyroscopes. As such, the game controller can only handle coarse motion sensing and tilt-sensing, i.e. estimating pitch and roll orientation of the controller with respect to gravity. Tilt-sensing can only be performed when the controller is held relatively stationary. Furthermore, the reported positions of the infrared LEDs will fluctuate constantly giving rise to inconsistent pitch and roll estimates even when the device is held relatively stationary.

The Wii Remote also incorporates optical sensing in the form of an infrared camera, mounted in front of the device, which can detect up to 4 infrared light sources. This is usually used in conjunction with the ‘sensor bar’, which typically consists of two clusters of infrared LEDs located at either end of the bar. These infrared light sources allow the controller to be used as a pointer, based on the reported positions of what the infrared camera sees. Relative distance from the infrared light sources can also be estimated using the separation between the reported positions. Optical sensing however will only work when the infrared light sources are within the camera’s limited field-of-view. Various sources have reported different field-of-view measurements [5], [13], [21].

There are two design alternatives that can be used for optical sensing, these are outlined in [28]. The first is the outside-looking-in approach, in which an optical sensor(s) is placed at a fixed location and landmarks (e.g. the infrared LEDs) are mounted on the user. This was the approach adopted thus far in Johnny Chung Lee’s popular Wii Remote projects [5]. The other alternative is the inside-looking-out approach where the sensor is moving whereas the landmarks are placed at fixed locations in the interaction space. Normal usage of the Wii Remote uses this method, where the sensor bar is placed at a fixed position, either above or below the TV, and the user moves the controller.

While both these approaches were considered in previous work, the outside-looking-in alternative was determined to be the most practical for the purposes of HMD virtual reality as it is rather impractical to surround the user with infrared light sources. Fig. 1 gives a depiction of the system design that was examined in [25]. It can be seen that this designed allows a user the freedom of interacting in an area of space surrounding the user. The design employed the use of 2 Wii Remotes. The user held one controller which was used to obtain pitch and roll using tilt-sensing. Two infrared light sources were attached to the first controller and these were used by the overhead Wii Remote to estimate 3D position and yaw. While this approach allowed for limited 6-DOF; 3D positioning, 360 degrees yaw and approximately +/- 45 degrees pitch and roll, there were a number of drawbacks with the system.

For one thing, tilt-sensing could not be done when the device was accelerating due to user hand movement. In other words, this restriction meant that the user was only allowed to move the device rather slowly. Also, the accelerometer readings fluctuated constantly giving rise to inconsistent pitch and roll estimates even when the device was held relatively stationary. Furthermore, the reported positions of the infrared light sources also jittered, thereby affecting other positional estimations. These factors combined made the 6-DOF tracking rather inaccurate and therefore inadequate for any meaningful 3D interaction.

Previous work also examined ray-casting selection and occlusion selection approaches to selection tasks. Selection tasks are what the user does when singling out a specific object or point in a virtual environment [30]. In the ray-casting approach, a ray is projected from a ‘virtual’ 3D interaction entity into the virtual environment. When the ray intersects an object, the user can usually select this object through a button press on the input device. The occlusion approach is similar to the ray casting method in that a ray is projected into the environment; however in this case the ray emanates from the user’s eye, through a point (typically the tip of a virtual ‘wand’) or virtual hand is used as the 3D cursor, then into the
environment. So in this case, the user does not actually see the ray. The object that the user selects is the object that is occluded by the 3D cursor. It was noted that the ray-casting selection approach was more suitable for this particular system design. Hence, this was the approach that was adopted in this study.

III. SYSTEM FRAMEWORK

In an attempt to improve the system design, it was determined that tilt-sensing should be avoided. Instead the 6-DOF estimations were obtained using the reported positions of four infrared light sources that were arranged in a non-planar configuration. These infrared light sources were attached to a Nintendo Wii Zapper gun mount. The gun mount provided an elegant solution that was well suited for ray-casting selection and also reduced the likelihood of the user obstructing line-of-sight between the infrared light sources and the optical sensor. This is because when properly holding the gun mount, the user’s hands would always be below the infrared light sources. In addition, if line-of-sight was lost, pitch and roll estimations could still be obtained through tilt-sensing. Fig. 2 illustrates this system setup. A Polhemus Patriot 6-DOF magnetic tracker was used to obtain user position and orientation, and an eMagin Z800 HMD was used to display the virtual environment.

A Kalman filter was introduced in order to improve and smoothen the readings from the overhead optical sensor. This recursive predict-correct filter is commonly used in a variety of systems, including tracking systems and virtual reality applications [31]. Others have also proposed the Kalman filter as a means of improving the Wii Remote’s accelerometer data outputs [32].

The model of the virtual object that the user handled to interact in the virtual environment was chosen to roughly correspond to the physical input device that the user was holding. This was done to help improve the user’s sense of immersion, through more intuitive interaction. The user could manipulate the virtual object by physically moving and/or rotating the input device. This provided spatial 3D tracking of the input device as long as the infrared light sources were not blocked and remained within the view of the overhead optical sensor.

In addition to manipulating the virtual device, the user could also interact with the virtual environment in a number of other ways. The user could perform selection tasks in the virtual environment via ray-casting selection. A ray originating from the virtual object was projected into the virtual environment, and in order to carry out a selection task the user could point the ray in a certain direction and press the gun mount’s trigger to select.

The user could navigate through the virtual environment using the analog control stick on the Wii Remote’s Nunchuk extension that was also attached to the gun mount. User translation through the virtual environment was in the direction that the user was looking in.

Another form of interaction was also implemented whereby users could swipe the virtual gun at other objects in the virtual environment. If the virtual gun collided with the other objects, it would knock them over. The amount of force that was exerted on that object was determined by using the physical input device’s speed and direction of movement, thus portraying a more realistic interaction response. The Wii Remote’s vibration motor or ‘rumble’ was used to provide tactile feedback in response to the collision.

IV. EVALUATION

Tests were conducted in order to ascertain the stability and accuracy of the system. The experimental setup was similar to that used in the previous study [25]. A number of targets were placed at various locations in the virtual environment. The user’s task was to attempt to accurately aim the ray emanating...
from the tip of the virtual gun at the centre of the target, press the trigger button and to try to hold it stable in that position over a few seconds.

A total of 16 targets were placed in the virtual environment. These were positioned at yaw angles of 0, 30, 60 and 90 degrees and at pitch angles of 0, 15, 30 and 45 degrees. Angles in other quadrants were not used as they were deemed to simply mirror these. Also, for human factors reasons, normal users generally will not often look at very high or low elevation angles in HMD virtual reality. Fig. 3 shows a screenshot of the virtual environment that was used in the experiment.

Accuracy measurements were taken by determining how much the intersection point between the ray and the target object missed the target’s bulls-eye by. 1000 readings were taken for each target at 60Hz. This meant that the user had to attempt to hold the input device steady at the target for around 16 seconds. The tests were repeated for targets at distances of approximately 5 and 10 metres away from the user, as well as when the device was used at distances of approximately 1, 1.5 and 2 metres away from the overhead Wii Remote. This was done because it was anticipated that accuracy would decrease the further away the infrared light sources were from optical sensor.

![Fig. 4 Average accuracy for targets at 5 metres](image1)

![Fig. 5 Standard deviation for targets at 5 metres](image2)

![Fig. 6 Average accuracy for targets at 10 metres](image3)
Fig. 4 shows the average of how much the ray missed the target’s centre, and the standard deviation of the accuracy measurements depicted in fig. 5. These results collectively indicate that the accuracy decreased when the device was tilted at steeper angles (i.e. the higher the pitch the less accurate it became). Furthermore, as anticipated accuracy also decreased the further away the device was used from the overhead optical sensor. The reason for this lies in the fact that when the input device is used closer to the optical sensor, the reported positions of the projected image points from the infrared camera would be closer together, thus making it harder to estimate the device’s physical orientation. This same explanation also applies for the decrease in accuracy at high elevation angles.

Similar conclusions can be drawn from the average accuracy measurements and standard deviation when the targets were at distances of 10 metres away from the user, these are shown in Fig. 6 and fig. 7 respectively. Fig. 4 and fig. 6 also show the drop in accuracy with distance away from the user.

From these results, it was concluded that the input device should ideally be used at a distance of 1 to 1.5 metres away from the overhead Wii Remote, and should not exceed 2 metres, otherwise the accuracy of the estimations will rapidly degrade. This however gives rise to a problem in terms of the total interaction area covered by the optical sensor. Because of the Wii Remote’s limited field-of-view, the closer the infrared light sources are used with respect to the overhead optical sensor, the smaller the area covered below the sensor. At the very least this means that the user cannot extend his/her hand outside the field-of-view. One solution to this is to use multiple overhead optical sensors. Two/three overhead Wii Remotes placed a certain distance apart will probably be enough to create an interaction space which is adequate for a HMD virtual reality application where the user has to sit on a swivel chair which is placed at a fixed location.

The use of the Kalman filter significantly smoothen the position and orientation estimates. So even though it introduced a slight delay in terms of the response time when the user moved the physical input device and its virtual representation moved correspondingly in the virtual environment, it was essential for the usability of the system.

V. Conclusion

This paper has presented a design for low-cost system for 3D spatial interaction using Wii Remotes for immersive HMD virtual reality. While the system is not perfectly accurate, it can be used in virtual reality applications where slight inaccuracies will not impede user task performance in the virtual environment. An example of such an application is for a virtual reality game. Future work will involve increasing the spatial interaction area by using multiple overhead optical sensors and using the system to set up a virtual reality game. Usability testing will also be performed in order to ascertain how well a human user can use the system, as well as to obtain feedback on their user experience and level of satisfaction.

Acknowledgment

The author would like to acknowledge the support of the UOW URC Small Grant used for this research.

References


