

A Study of Navigation and Selection Techniques in Virtual Environments Using Microsoft Kinect®

Peter Dam¹, Priscilla Braz², and Alberto Raposo^{1,2}

¹Tecgraf/PUC-Rio, Rio de Janeiro, Brazil

peter@tecgraf.puc-rio.br

²Dept. of Informatics/PUC-Rio, Rio de Janeiro, Brazil

{pbraz, abraposo}@inf.puc-rio.br

Abstract. This work proposes and studies several navigation and selection techniques in virtual environments using Microsoft Kinect®. This device was chosen because it allows the user to interact with the system without need of hand-held devices or having a device attached to the body. This way we intend to increase the degree of virtual presence and, possibly, reduce the distance between the virtual world and the real world. Through these techniques we strive to allow the user to move and interact with objects in the virtual world in a way similar to how s/he would do so in the real physical world. For this work three navigation and three selection techniques were implemented. A series of tests were undertaken to evaluate aspects such as ease of use, mental effort, time spent to complete tasks, fluidity of navigation, amongst other factors for each proposed technique and the combination of them.

Keywords: 3D Interaction, Virtual Reality, Gesture Recognition, HCI.

1 Introduction

Virtual Environments, due to enabling realistic and immersive experiences, have seen an increase in importance. Its use in areas such as games, simulation and training, medicine and architectural visualization has pushed the visualization technologies to rapid evolution. However, the way we interact with these environments hasn't evolved as fast, leaving a noticeable gap and hindering the interaction capabilities, since many inherently tri-dimensional tasks have been performed using technologies developed primarily to solve bi-dimensional tasks.

The objective of this work is to propose and study techniques that allow the user to interact in a complete manner using only corporal movements to perform tasks in a virtual environment, especially training and simulation, where the user normally needs to navigate through a scene and interact with equipment. For this, three selection and three navigation techniques have been proposed using Microsoft Kinect® as an input device. These techniques use corporal gestures, most of which aim to keep a certain fidelity to the respective actions in the real world in attempt to increase the naturalness of tri-dimensional interaction.

This paper is organized the following way: section 2 speaks of related work, section 3 presents the proposed techniques, section 4 presents results and analysis of user tests, and section 5 brings the conclusion.

2 Related Work

There are several researches in the virtual environment interaction area, but very few of those, up to the current date, make use of Microsoft Kinect®, due to it being a relatively new technology. For this reason, the study of related work focused on work about interaction in virtual environments.

According to Bowman and Hodges [1], interaction in virtual environments is divided into three types: locomotion (navigation), selection and manipulation, where, in many cases, the last two are combined, but can be dissociated. Since in this work both locomotion and selection have been considered, researches about either case have been considered in related work.

Selection. Sibert and Jacob [2] present a selection based on gaze direction. It is based upon a directional ray controlled by the direction of the eyes' gaze, eliminating the need of hand-held devices or devices attached to the user. The selection is triggered when the gaze rests upon an object for a certain amount of time. The idea of relating time to selection intention is contemplated in the *Hover* technique, presented in this paper. Rodrigues et al. [3] studied the advantages of applying multi-touch interface concepts in virtual reality environments by mapping 3D space into a virtual touch screen. To enable this, they proposed a wireless glove which is worn by the user and tracked by a specific configuration of Nintendo WiiMote® controllers. The index finger's position is tracked, mapping the axes into system coordinates. The X and Y axes are used to control the mouse cursor on the screen, while the Z axis is used to determine selection intent by establishing a threshold in the real world as if it were a screen. If the finger passes beyond this threshold the selection is activated and a command is triggered, sending haptic feedback, present in the glove. Even though the glove was designed for and tested in 2D interfaces, it inspired the *Push* technique, specifically the gesture of passing an imaginary plane in front of the user to confirm selection (or generating a "click"); and, consequently, also inspired the *Hold* technique.

Navigation. One technique that consists in putting the foot in a certain position to navigate is the *Dance Pad Travel Interface*, proposed by Beckhaus, Blom and Haringer [4]. This technique consists of a physical platform (created for the game *Dance-Dance Revolution*), which has directional buttons. The user steps on these buttons and a displacement is created in the direction represented by these buttons. To control the viewing direction the user steps on the directional arrows. One of the navigation techniques proposed in this work (*Virtual Foot Dpad*) was inspired by the *Dance Pad Travel Interface*. During the development of this technique a very similar technique was found in the game *Rise of Nightmares* for the XBOX/Kinect console.

Bouguila, Ishii and Sato [5] created a physical device, similar to a platform, which detects the user's feet and, when moved a certain distance away from the center, activate movement in that direction. To control the viewing direction the user turns his whole body in the desired direction. Because of this, a portion of the user's field of view might not be occupied by the viewing screen, so the device slowly rotates to align the user to the screen again. This work inspired the idea of allowing the user to completely leave a virtual circle, creating a movement vector with origin in the circle's center in the direction of the user's position. This led to the creation of the *Virtual Circle* technique.

3 Proposed Techniques

The proposed techniques use information obtained from Microsoft Kinect® as the only data input device. OpenNI [6] was used for communication between the device and the system.

3.1 Selection Techniques

First a virtual hand was developed to follow the user's hand movements in the real world. Moving this virtual hand over objects in the scene enables selection of this object, however the gesture required to select the object depends on which technique is being used. Unlike in Bowman and Hodges [7], due to our work focusing on selection and not exactly manipulation, we did not find the "lever" problem, where the object is attached to the extreme of a selection ray, making it difficult to properly manipulate the object.

Hover. This technique is based on the idea that the user will focus her/his attention on an object when s/he wishes to select it [2]. When the user wishes to select an object s/he needs to hover with the virtual hand over that object. A timer will appear and, once emptied, the object will be selected (Fig. 1). When the virtual hand intercepts a selectable object a "pre-counter" is started, introduced to avoid the "Midas Touch" effect, described by Jacob et al. [8]. This allows the user to freely move the virtual hand without actually triggering many visual timers all the time.

There are two ways to de-select an object with this technique. The first requires the user to move the virtual hand away from the selected object and, after a short time, it will be de-selected. This may not be possible if the object is attached to the virtual hand on all 3 axes, so a second de-selection method was created. The second method requires the user to overlap both hands, which will start a timer to confirm the intention of de-selection and, consequently, de-select the object once the timer runs out.

Push. The idea for this technique came from having a virtual plane in front of the user, described by Rodrigues et al. [3]. The user stretches her/his arm and, once it passes a certain threshold, the selection is triggered. The user must then withdraw her/his arm and may interact with the object. To release the object s/he repeats the gesture.

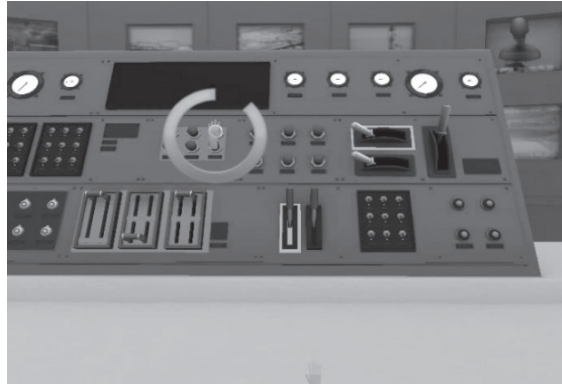


Fig. 1. Hover technique timer

The gesture of stretching the arm is detected through the arm's angle, more specifically the angle between the vectors formed by the elbow to the wrist and the elbow to the shoulder, as seen in Fig. 2. Once the angle reaches a pre-established limit, the system activates the selection (or de-selection). One problem present in this technique, described by Rodrigues et al. [3], is the involuntary movement along the X and/or Y axes while the user performs gesture of stretching her/his arm. This problem is more noticeable in cases where interaction requires a higher precision or when the object to be selected is very small on the screen, but for larger objects this problem rarely is an issue.



Fig. 2. Arm openness angle

Hold. This technique is based on the previous one, as an alternative. Selection is activated in this technique when the user stretches her/his arm, but, unlike the previous one, s/he must maintain the arm stretched during the interaction. De-selection is done by withdrawing the arm.

3.2 Navigation Methods

For a complete interaction experience the user must be allowed to select and to navigate through the scene. To enable this, three navigation techniques were created. Two of the proposed techniques use *Body Turn* to control the view point orientation. *Body*

Turn is a sub-part of these techniques and consists of the user turning her/his shoulders in the direction in which s/he wishes to rotate the view point, while maintaining the central direction of the body facing the screen. This allows the user to control the view and movement direction without the screen exiting her/his field of view.

Virtual Foot DPad. This technique was inspired by the work of Beckhaus, Blom and Haringer [4], where they created a physical platform on which the user steps on directional arrows to move in the corresponding direction. The idea was to make a virtual version of this platform. Three joints were used to achieve this: torso, left foot and right foot. The distance of each foot to the torso is calculated and, once one of the feet reaches a certain distance a movement is generated in that direction. This technique uses the previously described *Body Turn* to allow the user to control the view point orientation.

Dial DPads. Based on first person games for touch screen devices, such as *iPhone* and *iPad*, this technique uses dials that the user interacts with using virtual hands (Fig. 3). The idea is that it works in a fashion similar to a touch screen, but in larger scale and, instead of using fingers on a screen, the user uses hands. Two dials are displayed on the screen, one in each inferior corner. To the left is the movement control dial and to the right is the view point orientation dial. The user places her/his hand over the dial and stretches the arm to activate it.

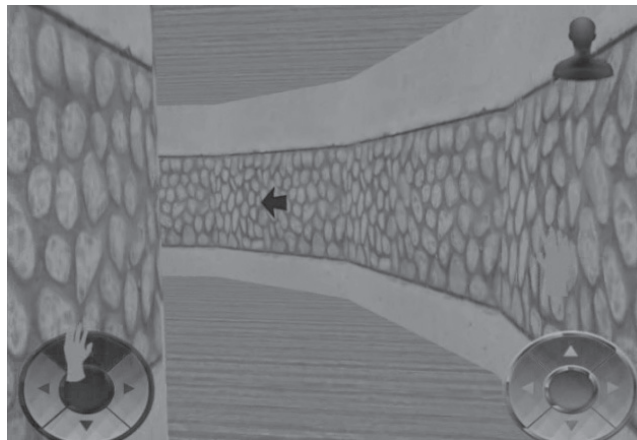


Fig. 3. Dial DPads controls

Virtual Circle. In this technique the system needs to store the position from which the user started the interaction and generates a virtual circle at this spot. The circle is fixed and the user can be compared to an analog joystick. To move in any direction the user simply moves in that direction enough to leave the virtual circle. A vector is then created from the center to the circle to the user's current position, defining direction and speed of the movement (Fig. 4). To stop the movement the user steps back into the circle. For view point orientation the technique uses *Body Turn*.



Fig. 4. Virtual Circle movement vector

4 Evaluation and Analysis of Test Results

4.1 Evaluation

Selection and navigation tasks were identified for the tests in a 3D virtual environment to exercise the interaction techniques being evaluated.

Three use scenarios were defined for execution of the tasks and evaluation of the interaction techniques, described below.

Scenario 1. In the first scenario only navigation was contemplated, alternating between the three navigation techniques proposed in this work. This scenario was a corridor, with two 90° curves and a section with a U-turn. The user needed to reach the end of this course, where there would be a red light. Once close enough to this light it would turn off and the user needed to turn around and go back to the initial point.

Scenario 2. In this scenario only selection was tested, alternating between the three selection techniques proposed in this work. In this scenario the user had a control panel placed in front of him/her containing a series of levers and buttons (Fig. 1). The user needed to first press several buttons following a specific order, according to which one was lit. After that a series of three red levers needed to be dragged up or down a track to a specific point and released once the indicator showed an acceptable position. At last, two green levers needed to be manipulated simultaneously until the end of their respective tracks.

Scenario 3. In this scenario navigation and selection were evaluated, alternating between the navigation and selection techniques. For this test we discarded *Dial Dpads*, because this technique makes use of hands, potentially creating conflict with the three selection techniques. The other two navigation techniques were used in combination with the three selection techniques, creating a total of six combinations. Each of these combinations were tested. This scenario tested the proficiency with buttons and levers, besides a new task: carry a ball while navigating and interacting with other objects at the same time (Fig. 5).

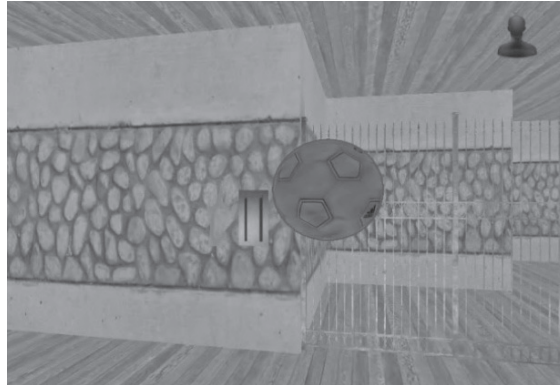


Fig. 5. User carrying a ball while navigating in Scenario 3

The order of the tests was changed for each user to avoid that learning had any influence in the general result of the test. In total 9 users were evaluated during the tests using the same physical set up: a room with enough space for free movement with a single large screen.

4.2 Analysis of the Results

Navigation. Mental effort reflects the degree of interaction fidelity of each technique. *Virtual Circle* had the greatest degree of interaction fidelity and, consequently, demanded less mental effort from the users. Similarly, *Virtual Foot*, which had the second greatest degree of interaction fidelity, demanded greater mental effort.

Comparing one leg of the path amongst the three techniques (Fig. 6) it is possible to observe that the users had a considerably better performance during the U-turn when using *Virtual Circle*. However, to walk in a straight line they performed better with *Virtual Foot*. The reason behind this is that *Virtual Circle* is completely analogical, so if the user moves slightly to any side the movement vector will not be 100% parallel to the walls, creating a slight deviation to one of the sides. This is visible in the initial part (from starting point until the first curve).

Selection. The repetition of the gesture for selection and de-selection, present in the *Push* technique, did not please the users, who had trouble with that. *Hover*, on the other hand, was criticized for introducing a delay to be able to select an object, being the least immediate of the three techniques. Despite this, *Hover* was the preferred technique in all tasks. Oppositely, *Push* was the worst in the opinion of the users.

It was made clear that for tasks that require high precision, such as the case of the red levers, the involuntary movement along the X and Y axes highly hinders the interaction, consequently affecting the users' preference of the technique.

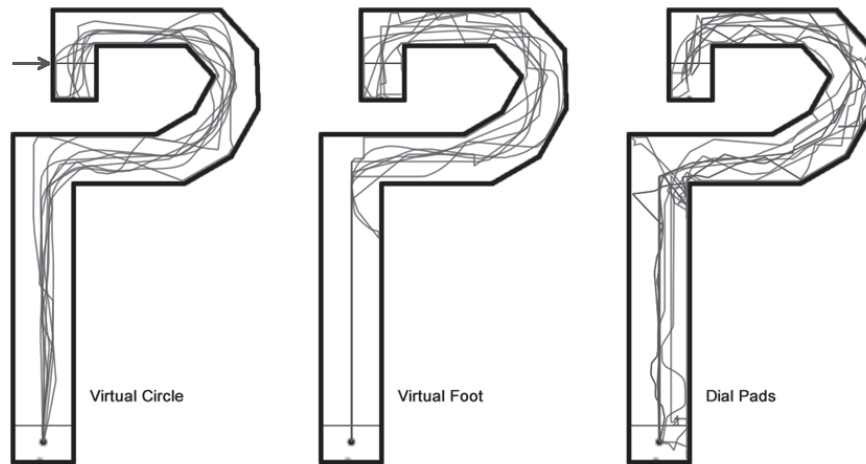


Fig. 6. Path outline for the first leg of the course

Curiously in selection, contrary to navigation, the technique with least interaction fidelity was the one the users preferred. Bowman et al. [6] speak about interaction fidelity, questioning if a technique with higher interaction fidelity means it is necessarily better.

Combination of Navigation and Selection. When comparing directly the navigation techniques, we observed that the *Virtual Circle* technique was, in fact, considered slightly better in pair with selection, while the mental effort was very similar, showing that the change in navigation techniques did not have great impact on selection. However, it is possible to observe that strictly comparing navigation tasks, the users preferred *Virtual Circle*.

The technique that had most user technical faults (executing actions by mistake) was *Hold*, with large difference to the second placed technique *Push*. *Hover* did not have any mistakes of this type. These errors were caused by the user withdrawing her/his arm when s/he shouldn't have.

Fig. 7 shows the average execution time for the tasks, considering the order in which they were performed, not sorted by technique. The average time was considered for each 1st task of all users, then for each 2nd task, and so on. The completion and collision timings show that, no matter which technique combination used, there is a learning curve, indicated by the decreasing lines for task completion. The 4th task causes an increase in completion time compared to the 3rd task. This is due to changing the navigation technique: the first three tests were applied using one of the navigation techniques, then the last three were applied using a different technique.

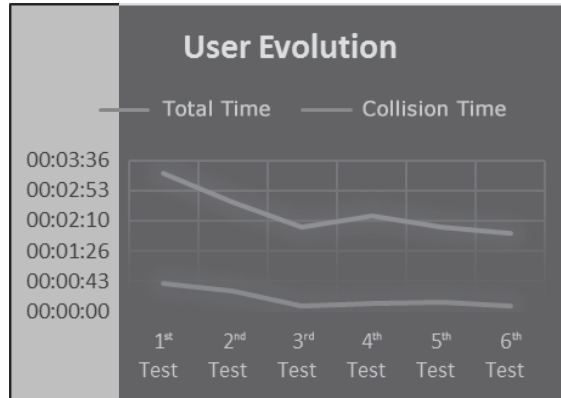


Fig. 7. User evolution based on average test execution times

The combination of navigation and selection consolidated *Hover* as the overall preferred technique by the users. This preference happened because the users felt more secure to carry objects while navigating, since they didn't have the risk of accidentally dropping the ball.

Besides that, *Virtual Circle* continued to be the preferred navigation technique, but not as evidently as in the first scenario. This was because, in the third scenario, the user was less prone to collision since the environment was more ample than the first, which had narrow corridors.

5 Conclusion and Future Work

One of the advantages initially predicted with these techniques was the possibility of interacting with both hands at the same time, a possibility not currently easily supported by current devices. To evaluate this advantage, amongst others, as well as limitations imposed by the techniques, we had to develop user tests. Through these tests we identified which techniques allow a satisfactory interaction, enabling the user to perform tasks in a virtual environment, such as exploring and interacting with objects (despite not being able to rotate and scale them, the users could select and move them).

It was possible to observe that there is clearly a learning curve and, after several tasks, the users would discover ways to use the techniques in which they felt more comfortable. Even though no techniques, in general, had a poor performance, each user, in the end, felt more comfortable with a certain navigation and selection technique. Despite this, it was not possible to compare these techniques with techniques the users were already familiar with, due to the possibility of using both hands simultaneously.

At last, it was possible to verify that Microsoft Kinect® enables the creation of techniques with high degree of interaction fidelity that allow several user actions in a virtual environment in a comfortable manner, besides increasing the user's virtual presence. After some improvements, especially in the implementation of the

techniques, we believe that they can be used in virtual reality applications to control a character and, possibly, to perform more complex tasks than currently possible, mainly due to the possibility of using both hands simultaneously.

Acknowledgements. Tecgraf is an institute mainly supported by Petrobras. Alberto Raposo thanks CNPq for the individual grant (process 470009/2011-0).

References

1. Bowman, D., Hodges, L.: Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *Journal of Visual Languages and Computing* 10(1), 37–53 (1999)
2. Sibert, L., Jacob, R.: Evaluation of Eye Gaze Interaction. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2000)*, pp. 281–288. ACM, New York (2000)
3. Rodrigues, P., Raposo, A., Soares, L.: A Virtual Touch Interaction Device for Immersive Applications. *The Int. J. Virtual Reality* 10(4), 1–10 (2011)
4. Beckhaus, S., Blom, K., Haringer, M.: Intuitive, Hands-free Travel Interfaces for Virtual Environments. In: *New Directions in 3D User Interfaces Workshop of IEEE VR 2005*, pp. 57–60 (2005)
5. Bouguila, L., Ishii, M., Sato, M.: Virtual Locomotion System for Human-Scale Virtual Environments. In: *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI 2002)*, pp. 227–230. ACM, New York (2002)
6. OpenNI, <http://openni.org/>
7. Bowman, D., Hodges, L.: An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In: *Proceedings of the 1997 Symposium on Interactive 3D Graphics (I3D 1997)*, p. 35. ACM, New York (1997)
8. Jacob, R., Leggett, J., Myers, B., Pausch, R.: An Agenda for Human-Computer Interaction Research: Interaction Styles and Input/Output Devices. *Behaviour & Information Technology* 12(2), 69–79 (1993)