Navigating and Selecting in the Virtual Supermarket: Review and Update of Classic Interaction Techniques

Patrick Renner, Timo Dankert, Dorothe Schneider,
Nikita Mattar, Thies Pfeiffer

Artificial Intelligence Group, Faculty of Technology, Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
email: {prenner, tdankert, dschneid, nmattar, tpfeiffe}@techfak.uni-bielefeld.de

Abstract: Classic techniques for navigation and selection such as Image-Plane and World in Miniature have been around for more than 20 years. In the course of a seminar on interaction in virtual reality we reconsidered five methods for navigation and two for selection. These methods were significantly extended by the use of up-to-date hardware such as finger-tracking devices and the Nintendo Wii Balance Board and evaluated in a virtual supermarket scenario. Two user studies, one on experts and one on novices, revealed information on usability and efficiency. As an outcome, the combination of Ray-Casting and Walking in Place turned out to be the fastest.

Keywords: Interaction, Navigation, Selection, HCI, Immersive Virtual Reality

1 Introduction

Navigation and selection in virtual reality have been well-known challenges for more than 20 years. What appears to be easy in real-life requires the right idea and often a complex interplay of hard- and software to work in virtual reality. Making travel and selection in virtual environments similar to the real world is a goal that is still strived after today.

In the course of the IEEE Symposium on 3D User Interfaces 2010, the Video Submission Contest 3D UI Grand Prize was announced to find innovative solutions to these classic 3D UI problems in a specific supermarket scenario (see figure 1). Participation in the contest required sending a video and performance data such as time on task from a usability study.

Supermarket scenario The organisers provided the scenario coded in VRML. The supermarket is furnished mainly like in real-life, with shelves holding e.g. bottles of milk, bread, and washing agent, or freezers with frozen pizzas. Instead of a cash register, there is a purple table on top of which are a red frozen pizza, a green cup, and a red milk carton. The task is to start at the table, fetch the objects corresponding to the examples from the supermarket, carry them back to the table, and place them next to the originals.

Our contribution At the same time we, among 16 other students, participated in a course on "Interaction in Virtual Reality" given by Nikita Mattar and Thies Pfeiffer. We imple-
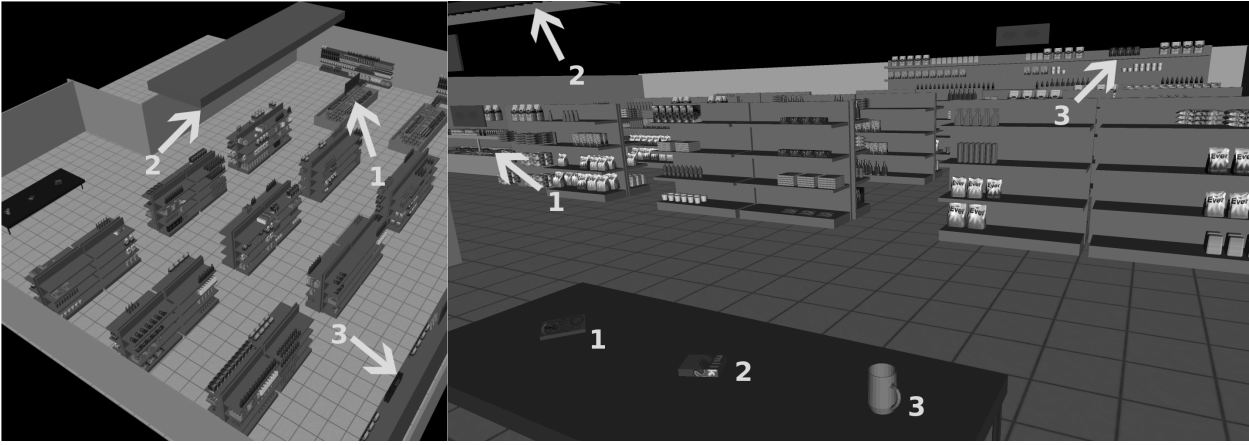
Figure 1: Two views of the supermarket. The locations of the 3 target objects are marked.

mented well-known interaction techniques which had already been evaluated and deemed usable. For participation in the 3D UI Grand Price we chose Path Drawing, Lean-Based Velocity, Walking in Place, World in Miniature, and Grabbing in the Air as navigation methods, as well as Ray-Casting and Image Plane for selection and manipulation. We then combined pairs of one navigation and one selection method each to allow for a full interaction with the scenario. As these techniques were mostly described in the early 1990s, the pre-conditions for them had changed significantly, taking the technical progress into consideration. Especially a 3-sided CAVE and input devices like the Nintendo Wii Remote and the Nintendo Wii Balance Board are newer technical alternatives which could be profited from. Hence, our implementations and evaluations of the techniques differ distinctly from the originals.

2   Equipment

The techniques were all implemented for use in a 3-sided CAVE called TRI-SPACE with a polarised light stereo projection and an ART tracking system. The viewpoint is relative to the position and orientation of the user's head.

Input devices used for the techniques are: The cordless **Nintendo Wii Balance Board** for navigation, which is divided into 4 touch-sensitive areas and measures the pressure applied by the user's feet. The cordless **Nintendo Wii Remote** with eleven buttons, an accelerometer and 2D-tracking by IR. We equipped it with additional markers in order to track its position and orientation by the ART system. We call this modified version "WiiMod". For "Grabbing in the Air" and "Image Plane", we used **ART fingertracking devices** to track the position of three fingers on each hand. On this basis, we defined some hand gestures to function as triggers.

Below the titles of each method are symbols which describe some of their characteristics. Table 1 explains those symbols.

[two hands symbol]	Method needs two hands	[feet symbol]	Method needs feet
[3] .. [7]	Degrees of freedom	[one hand symbol]	Method needs one hand
[head symbol]	Method needs head		

Table 1: Explanation of symbols

3 Navigation methods

This section presents our five methods for navigation through a virtual scene. The structure of each method description is as follows: In the beginning, we depict the design as it was described in former papers. Afterwards, our approach is outlined, followed by the relevant technical details.

3.1 Path Drawing

[one hand symbol]	3

Original Design The navigation method "Path drawing for 3D walkthrough" [IKMT98] is an extension of the flying technique (cf. [MCR90a]) for non-immersive set-ups. The user can navigate an avatar through the virtual world by drawing a path onto the ground and the drawn stroke is projected onto the walking surface as visual feedback. Afterwards, the viewpoint is dynamically adjusted by being moved along the tangents of the path.
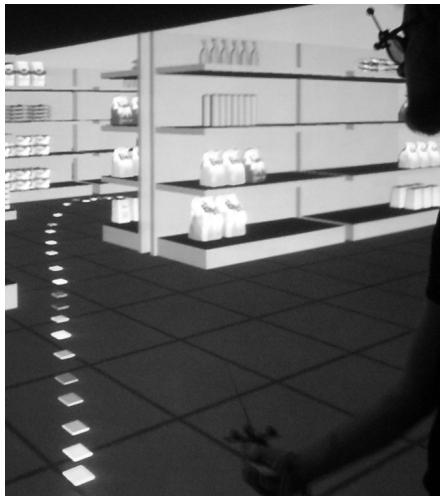

Figure 2: Pathdrawing

Using this method, the user gets the feeling that the avatar which represents him on-screen is moving along the drawn path.

The system can be operated in different ways. The user can, for example, define his desired path in detail by drawing a long stroke from his current position to his target. If the exact path does not matter to him, he can also draw a short stroke near the target, which means that he will reach it in a straight line. Finally, if he wants to change the viewpoint's orientation, he can do so by drawing a short stroke near his feet. Drawing mistakes can also be corrected simply by drawing a new path.

A user study ([IKMT98]) revealed that the Path Drawing technique was more intuitive and more widely accepted than both the flying and the driving method. While the driving method suffers from the need to continuously press a button during movement, the flying method offers only a limited control over where to navigate and is prone to causing disorientation. In the spectrum defined by those two methods, Path Drawing lies in the middle and thus, combines both of their advantages while avoiding their disadvantages. The desired path can be controlled easily, yet the user does not need to hold down the button the entire time, which means that he has his equipment free for other things.

Our Realisation In our immersive set-up there is no avatar and the user himself moves through the supermarket. Furthermore, we adapted the original method to the supermarket's planar 2D walking surface and eliminated the need to fit the drawn stroke onto a 3D surface.

Technical Details The means of drawing is realised with the WiiMod which functions as a "virtual pen". Drawing is triggered by a specific button. The path is drawn onto the floor of the virtual supermarket, at the point where a ray projected from the WiiMod and the walking surface intersect. The drawn path is then represented as colourful squares on the ground which give off a small glow so that they are easily distinguishable from the ground even in darker scenes.

3.2 Lean-Based Velocity

👣 3

Original Design [FHvdVN09] and [HBTF09] describe an empirical study of different modes of travel using the Nintendo Wii Balance Board. They explored two different input methods of navigating the 3D environment - one velocity-based, the other one acceleration-based.

Depending on which method is used, a user's forwards and backwards leaning motion controls either his velocity or acceleration. Standing still has no effect, which is somewhat counter-intuitive and requires a longer adaption phase. Both methods let the user change his moving direction by leaning sideways, thus rotating his view around his own up-axis. A "user vector" is calculated which represents the gist of the user's intended motion. The sigmoid functions used to calculate the user vector were chosen so that small movements would not alter it drastically while still preserving the swiftness of the user's perceived motion. So moving at higher speeds, rapidly changing directions, or moving for a long period of time would still be possible and not strenuous.
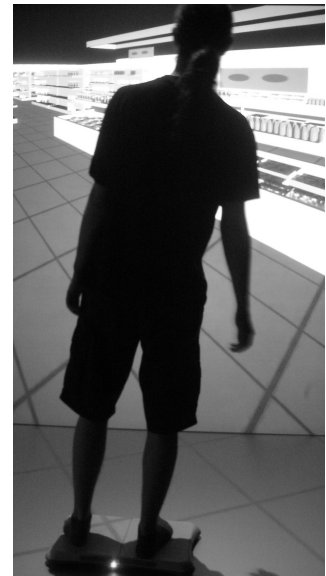

Figure 3: Lean-Based velocity

Our Realisation Just like in the original paper's velocity-based implementation, we let the user steer by shifting his weight to either side (figure 3) and move with leaning motions. In a departure from the original, we found that the supplied supermarket scene is small enough for velocity-based movement to be practical. The specified task contains movement periods followed by selection and manipulation periods. So in our implementation, the user does not need to pay attention to actively not moving while selecting and vice versa.

Technical Details We work with the supplied user vector as per [HBTF09] and use the methods described in [MCR90b] to determine whether the motions would still behave intuitively while the user was busy using other devices for selection and manipulation purposes.

To let selection tasks be perfomed without inadvertent movements, we restrained our model accordingly.

We also restrained the user vector, which itself is not weight-dependent, and project it onto the xz plane to determine the user's wish to rotate or move. Rotation is directly calculated from the degrees on the xz plane, while movement is calculated as a function of the angle between normal vector and user vector. We set minimum and maximum angles of leaning to either side, so that the avatar behaves reasonably while standing still, getting on and off the board, and moving. We compensate for the fact that users are usually able to lean forwards much easier than backwards with different angles.

The user vector incorporates an acceleration threshold, so there is still non-linear behaviour. This was expected by most users; i.e. leaning farther results in faster movement.

3.3 Walking in Place

| 👣 | ✊ | 3 |

Original Design Walking in Place is an easily understandable method for navigation based on [RSS+02]. The user's head position is tracked and its up-and-down motion during walking is translated into steps by a neural network. If a user wants to change his walking direction, he needs to physically rotate his own body; the scene will change around him accordingly. The rotation is determined by tracking the user's torso and then setting it in relation to the position of the front wall. After a change of direction, the viewpoint is slowly oriented back to the front wall in as subtle a way as possible, so that the user is not irritated by a sudden change which he did not cause.
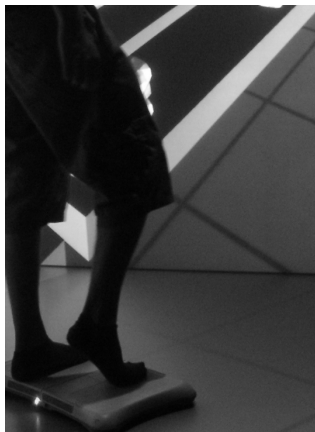
Our Realisation The main difference to the original method is that we do not measure the user's head movements and thus, do not need a neural network in order to determine if the user is currently walking or not. Instead, we use a simplified system which is operated directly by the user's feet. Thus, a walking motion in small steps is achieved by continuously lifting your feet and placing them down on the sensors again (figure 4). Note that the length of the lifting period determines the movement's velocity.

For rotation, we make use of the user's head orientation. Sudden viewpoint changes are interpreted as changes of walking direction. Similarly to the original paper, the viewpoint is then dynamically adjusted until the user faces the front wall again. The greater the change of direction, the faster the readjustment speed. This way, no physical rotation of the user's body is necessary, which means the method is less tiring than the original.

Figure 4: Walking in Place

Technical Details To recognise the steps we use the Nintendo Wii Balance Board. The orientation is determined with the tracked glasses.

3.4 World in Miniature

| 🖐🖐 | 3 |

Original Design The World in Miniature (WIM), based on [SCP95], is an immersive, intuitive way of navigating through a computer-generated environment by using a miniature model of the virtual world as a map. Especially in big and complex worlds, the WIM gives the user an additional, clear, aerial view of the scene. An advantage of this method is that the user has a great amount of control over the virtual world.

In the original paper, the user's non-dominant hand is equipped with a clipboard attached to a position tracker, so that the user can control the position and orientation of the WIM and, for example, lift it into or out of his visual field. The dominant hand holds a cue ball, a device equipped with two buttons and another sensor which is used for selection and manipulation.

Our Realisation The WIM is used like a map on which the user's position is indicated. In contrast to the original paper, the user selects his target position instead of moving an avatar through the map and is then instantly moved to his new destination (flying technique). The WIM can also be turned on or off as the user desires, which provides interference-free interaction with the environment. The user can also resize the WIM, rotate the world by its Y-axis and do small steps into the current viewpoint's direction. It is even possible to take an object in the WIM and transfer it to the original world.
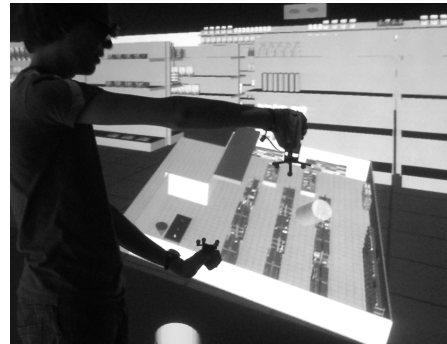

Figure 5: World in Miniature

Technical Details In his dominant hand, the user holds a WiiMod with which he controls each of the various actions with a specific button. A ray originates from this device, which is used for navigation, selection, and manipulation purposes. The non-dominant hand is vested with a simple tracked device which provides the system with a three-dimensional position for the WIM (figure 5).

3.5 Grabbing in the Air (Original name: Scene in Hand)

| 🖐🖐 | 7 |

Original Design Grabbing in the Air is an extended form of the "Scene in Hand" metaphor as it was described by [WO90]. The original idea is to establish a direct linkage between the scene and the translation and rotation of the user's hand. The user can move by grabbing the scene wherever his hand happens to be and then dragging it around.

Originally, a 6 DOF input device called "bat" was used for this method. However, the handling of this instrument was the test users' major complaint, as using it was tiring. (If,

for example, one wanted to do large rotations of the scene, one had to rotate the bat all the way as well.)

The original implementation had only a single, pre-defined centre of rotation. This made the technique feel natural for hand-sized scenes, but difficult to operate for larger ones, as small changes in rotation could have large effects on the user's sense of orientation if the focus of the current viewpoint was too far from the centre of rotation.

Figure 6: Grabbing in the Air

Our Realisation   The main difference between the original implementation and our realisation is that our approach allows the placement of "handles" as suggested by [WO90]. The user can decide which points in the world he would like to apply his manipulation to. This makes rotation easier and enhances the user's control over the scene.

The user can grab points in the virtual world and manipulate them as if he had a firm handle on those points, even if they lie in thin air. This, unlike most other navigation approaches, gives the user the impression that he himself is standing still and the world moves around him. As already mentioned, it also simplifies rotation because the physical constraints of the original bat are no longer a concern; now, rotations can be realised through a series of consecutive movements.

The possible navigation techniques are as follows: For translation the user grabs a point in the scene which then follows the hand movement (figure 6). He achieves rotation by grabbing two points in the scene and moving his hands; the scene will rotate around their centre point. Lastly, if the user wants to scale the scene, he uses both hands to grab points in the scene and changes their distance to each other, which activates the zoom in and out functions.

Technical Details   Our implementation can be operated in two ways. The first uses two WiiMods and realises a grabbing gesture via a specific button. The second makes use of two ART fingertracking devices with which a grabbing gesture equals a grasping gesture between thumb and index finger.

Both approaches have advantages and disadvantages. While the WiiMods are easy and intuitive to use and do not require calibration, the fingertracking devices provide better immersion because the user gets the feeling of manipulating the scene "with his own hands" without the need for any hand-held equipment.

4   Selection Methods

This section presents our two methods for selecting objects in a virtual scene. The structure of the method descriptions is the same as in the navigation part.

4.1 Ray-Casting

✋	6

Original Design The Ray-Casting method uses a virtual ray of light to select objects as described in [BH97]. The user can point the ray in his desired direction; it will then highlight the first object it hits. When highlighted objects are selected for manipulation, they will be attached to the ray and follow its movements. A problem is that the object's distance to the user stays constant during the interaction, which

Figure 7: Ray-Casting

is why complex manipulations such as rotations require extra effort.

Our Realisation We implemented a version of Ray-Casting in which the ray's length is finite and can be changed by the user. The interaction metaphor is not a ray of light, but a virtual pointing staff which can be arbitrarily extended and shortened. This holds two advantages: The first is that objects can be manipulated in an easier way because the user can, for example, pull objects closer to himself. Furthermore, the selection of covered objects is possible because the ray can pass through the objects in front.

When the end of the staff hits an object, its bounding box becomes visible to the user and the ray changes its colour. The user can then press a specific button to attach the highlighted object to the staff's end (figure 7). By pressing another specific button and moving the hand forwards or backwards, he can scale the staff's length. A slow movement causes only a small adaption of the length, whereas a fast one allows rapid scaling proportional to the speed of the movement. The adjustment lasts until the button used for scaling is released. Scaling the staff to its minimal size allows for rotation of the selected object around all axes, which gives the impression of a hand-centered manipulation.

Technical Details Our realisation of Ray-Casting uses two buttons on our WiiMod, one of which is used to attach objects to the end of the pointing staff while the other one activates the scaling mode for the staff's length.

4.2 Image Plane: Head Crusher

✋✋	7

Original Design The original design was proposed by [PFC+97]. They defined four different pointing gestures: Head Crusher, Sticky Finger, Lifting Palm, and Framing Hands. A ray from the user's eyes to the gestures' position was used to determine which object was in focus. The gesture we focus on is Head Crusher, for which the user forms a window with thumb and index finger of the same hand so that the enclosed area shows the projection of the object to be selected. The object is then selected with a gesture performed by the other hand.

Figure 8: Image Plane

Our Realisation With our approach, selection is realised in a way similar to the original design (figure 8). Once the right object is selected, the user can switch into manipulation mode by performing a grasping gesture with the other hand. He can then manipulate the object by grabbing it with both hands.

The manipulation methods are: translation, for which the user grabs the object with both hands, which causes the object to follow the user's hand movement; rotation, which the user executes by specifying the centre of rotation with a grasping gesture and moving his other hand around it; and scaling, which happens when the user grabs the object with both hands and varies the distance between them.

Technical Details The ART tracking system is used to track hand and head movements to recognise gestures and determine the starting point of the ray from the user's eyes.

5 Evaluation

The presented methods for navigation and selection have been combined to create a non-exhaustive selection of five reasonable interaction methods (see left column of Table 2) for the supermarket task (see section 1). We chose to combine them in order to give the user full control over the virtual scene. Furthermore, had we tested the methods separately, interferences between navigation and selection would not be accounted for.

Our original intention was to measure the time on task for 20 novice users. However, it turned out that interacting and orientating themselves in the immersive 3D environment was too difficult for most of them. Some required over half an hour for a single task and thus, were not able to complete the full set of tasks within reasonable time. This gives rise to a set of interesting questions worthy of being addressed. But in order to gather the time on task data for the contest we had to change our study and rely on seven experts with prior virtual reality experiences.

We did, however, have the novices fill out a questionnaire about issues such as easiness, fatigue, fun, learnability, and realism for the one interaction method they had used. Some results of these interviews are picked up in the discussion.

5.1 Usability Study on Experts

The experts were given some time to train the interaction techniques before time was taken. They were instructed to solve the task as quickly as possible. Their results are presented in Table 2.

The pair of Ray-Casting and Walking in Place was the fastest at the median (310 sec). Considering only the methods combined with Ray-Casting, World in Miniature was the

Condition	Time on Task (sec)		
	median	mean	sd
Image Plane/Grabbing in the Air	532	540	161
Image Plane/Lean-Based Velocity	358	370	114
Ray-Casting/Walking in Place	310	349	98
Ray-Casting/World in Miniature	359	376	123
Ray-Casting/Path Drawing	404	391	112

Table 2: Results of the survey

second best (but had the fastest single run with 180 sec), followed by Path Drawing, but the differences are quite small.

Regarding the combinations with Image Plane, Lean-Based Velocity was the fastest method (358 sec for median) and the second fastest overall. Image Plane and Grabbing in the Air is the only pair which differs significantly from the others. It is the slowest method with a median of 532 seconds.
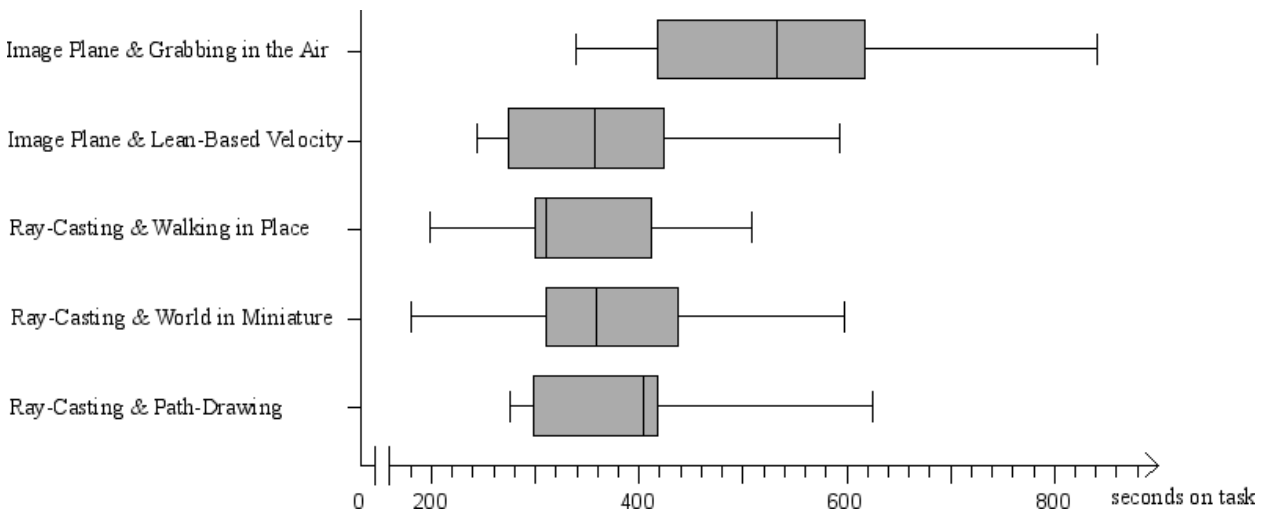

Figure 9: Results of the survey (whiskers represent the min/max times)

6 Discussion

Current technology triggered the opportunity of implementing multiple ways of interaction both in relatively short time and with low-cost input devices. Taking the results of the studies into consideration, we can make some observations:

Navigating with the Wii Balance Board was fastest for both selection methods (Ray-Casting and Image Plane). In these cases, there is a strong dissociation between navigation, done by the feet, and selection, done by the hands, which minimises the interference between the two.

The novice users assessed Path Drawing as the simplest and most usable navigation method. However, when used by experts, it is also the slowest of the three navigation methods combined with selection by Ray-Casting. This reveals a disadvantage: In Path Drawing, the user moves with a constant speed. World in Miniature is slightly faster, as it provides the possibility of "jumping" to a specified point in the scene, while it requires some orientation on the map. The speed of movement in Walking in Place is determined by the actions of the user within the scene, and it is the fastest of the combinations.

Image Plane/Grabbing in the Air clearly was the slowest interaction method. This might result from its complexity. Both are 7DoF methods gaining their input from ART finger-tracking devices. Especially Grabbing in the Air possesses a high functionality. It is the only navigation technique that allows reaching every part of the scene in 3D space because there is no restriction to a specific surface. Furthermore, the novice users reported this combination to be rather difficult and hard to learn. Nevertheless, we assume that extensive training will trigger much better results.

Considering these findings, the designer of a virtual environment has to ponder carefully which technique he adopts for his domain. Distributing navigation and selection to different input devices and modalities seems to increase speed, but it is more expensive and it takes longer to set up. Fingertracking has been reported by the participants to increase immersion, but it has to be carefully calibrated beforehand.

For installations with a higher frequency of short-time novice users, methods which are robust and quick to set up, e.g. those using the Nintendo Wii Remote (Ray-Casting or World in Miniature), are preferable. The overall duration of the interaction session might also be decreased by distributing the interaction to different modalities. Expert users might prefer the higher immersion and the more fine-grained control provided by fingertracking devices. In general, in every case, advantages and disadvantages of the techniques should be weighed.

In order to get a better impression of the different methods, our video for participation in the 3D UI Grand Prize can be found at the following address: http://www.youtube.com/watch?v=KaNE4946LDk.

7   Acknowledgements

References

[BH97]   D. Bowman and L. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings*

of the 1997 symposium on Interactive 3D graphics, pages 35–38, New York, USA, 1997.

[FHvdVN09]   W. Fikkert, N. Hoeijmakers, P. van der Vet, and A. Nijholt. Navigating a Maze with Balance Board and Wiimote. In *Proceedings 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 09), Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (LNICST), Vol. 9*, pages 187–192, Amsterdam, Netherlands, 2009.

[HBTF09]   A. Hilsendeger, S. Brandauer, J. Tolksdorf, and C. Fröhlich. Navigation in Virtual Reality with the Wii Balance Board. In *Virtuelle und Erweiterte Realität - Sechster Workshop der GI-Fachgruppe VR/AR*, pages 269–280, Aachen, Germany, 2009.

[IKMT98]   T. Igarashi, R. Kadobayashi, K. Mase, and H. Tanaka. Path drawing for 3D walkthrough. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 173–174, NY, USA, 1998.

[MCR90a]   J. Mackinlay, S. Card, and G. Robertson. Rapid controlled movement through a virtual 3d workspace. In *SIGGRAPH '90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 171–176, New York, NY, USA, 1990. ACM.

[MCR90b]   J. Mackinlay, S. Card, and G. Robertson. A semantic analysis of the design space of input devices. *Hum.-Comput. Interact.*, 5(2):145–190, 1990.

[PFC$^+$97]   J. Pierce, A. Forsberg, M. Conway, S. Hong, R. Zeleznik, and M. Mine. Image plane interaction techniques in 3d immersive environments. In *I3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 39–ff., New York, NY, USA, 1997. ACM.

[RSS$^+$02]   S. Razzaque, D. Swapp, M. Slater, M. Whitton, and A. Steed. Redirected walking in place. In *ACM International Conference Proceeding Series; Vol. 23*, pages 123–130, Barcelona, Spain, 2002.

[SCP95]   R. Stoakley, M. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *CHI '95: Proceedings of the SIGCHI*, pages 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[WO90]   C. Ware and S. Osborne. Exploration and virtual camera control in virtual three dimensional environments. In *Proceedings of the 1990 symposium on Interactive 3D graphics*, pages 175–183, New York, USA, 1990.