COMPARATIVE EXPERIMENT OF BODY-POSITION BASED NAVIGATION IN IMMERSIVE VIRTUAL ENVIRONMENTS

Kikuo Asai
National Institute of Multimedia Education, Mihama-ku, Chiba, Japan

Hideaki Kobayashi
The Graduate University for Advanced Studies, Mihama-ku, Chiba, Japan

Keywords: Human-Computer Interaction, Body-Position Based Navigation, Virtual Environment, Immersive Projection Display.

Abstract: Navigation is one of the basic human-computer interactions in virtual environments, and the technique should be easy to use, cognitively simple, and uncumbersome. However, the interaction in immersive virtual environments requires a factor of the sense of immersion as well as efficiency. We have proposed a body-position based navigation technique that drives a viewpoint with extension and bending of the arms, rotating both arms, and standing on tiptoes and bending the knees. Using various parts of the body may help to enhance the sense of immersion in the virtual environment. Depth images obtained from a polynocular stereo machine are used for tracking the 3D positions of the arms and head of the user in an immersive projection display. We evaluated the body-position system in experiments in which participants performed fly-through tasks in a 3D space, and compared the effectiveness of the body-position system with that of a joystick and a hand-arm gesture interface. The results of the experiment showed that the body-position system was advantageous on moving around at large areas instead of efficiency or accuracy of viewpoint control in virtual environments.

1 INTRODUCTION

Navigation is a basic human-computer interaction (HCI), which is generally a way to move the user into the location where she/he performs the primary tasks. Therefore, the interaction technique should be easy to use, cognitively simple, and uncumbersome (Bowman and Hodges, 1999). However, these factors are not enough for immersive virtual environments (VEs). The navigation in immersive VEs needs essence to enhance the feeling of being there so that the HCI makes fly-through content enjoyable, which takes the user around in the three-dimensional (3D) virtual world.

Most current fly-through systems employ a 3D mouse, joystick, or glove as the input device in VEs (Vince, 1995; Bowman, et al., 2005). Such devices allow for intuitive control of speed and direction of movement because they are easy to use and understand, and are not physically tiring to manipulate. However, these devices are designed focusing on efficiency or accuracy, but not immersion enhancement. We have proposed using body position as a means of navigation. The user's viewpoint moves with movements of both the arms and legs (Figure 1), and hence the user is required to maintain their balance. Using various parts of the body and sensory organs may provide the user with an enhanced sense of immersion, and increase interactivity with the immersive VE.

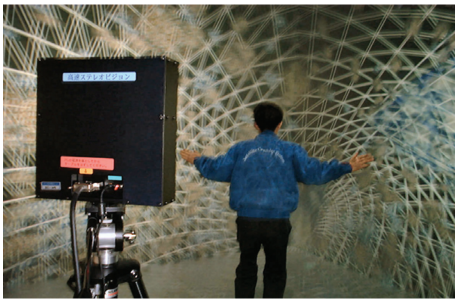


Figure 1: Body-position based interface.

Asai K. and Kobayashi H. (2007).
COMPARATIVE EXPERIMENT OF BODY-POSITION BASED NAVIGATION IN IMMERSIVE VIRTUAL ENVIRONMENTS.
In *Proceedings of the Second International Conference on Computer Graphics Theory and Applications - AS/IE*, pages 106-113
DOI: 10.5220/0002078701060113

The immersive projection display (IPD) (Cruzz-Neira, 1992; Ihren and Frisch, 1999), surrounds users with stable wide-angle images giving them immersive VEs. It uses lightweight stereo glasses, thereby eliminating the need to wear large headgear such as a head-mounted display. IPDs are often used for presenting 3D entertainment contents where the users enjoy navigation itself in the immersive VEs. The navigation of the contents is usually done by a joystick with a magnetic sensor used for pointing at 3D objects. Alternative interfaces should be investigated to gain the user's sense of presence in the virtual world.

It is often difficult for users to decide an appropriate interface device in 3D VEs and development of efficient interaction techniques, because the most effective ways for humans to interact with synthetic 3D environments are still not clear and may depend on the applications (Herndon, et al., 1994). This has given rise to studies on design and evaluation guidelines of the interfaces used in 3D VEs (Gabbard, et al., 1999; Kaur, et al., 1999), but there is little work on the evaluation of 3D interaction by body position in immersive VEs and on the adaptability to their applications in emphasis on the sense of immersion.

We examined performance of the body-position based navigation system by making an experiment of fly-through tasks in 3D VEs in comparison with a joystick and a hand-arm gesture interface often used in IPDs.

2 RELATED WORK

Various techniques have been developed for applications to HCI using body motion. Mapping head movements to navigation has been dedicated to hands-free fly-through applications in VEs, although the sensors have often been implemented in interfaces using a wired approach. A head-directed system (Fuhrmann, et al., 1998) determines speed and direction of navigation. The advantage of such a system is to be simple requiring no additional hardware except a head tracker, but using a head direction leads to limitation of the view direction.

Both head and foot movements have been mapped into viewpoint motion (LaViola, 2001). This system used a floor map as a world in miniature to move to the desired location in the virtual world, and detected the body's lean to enable movement over a small distance. The interaction technique is suited for moving to specific places, but not for moving around in VEs such as fly-through applications.

The physically connected systems constrain natural movements of the user within VEs. The user often has to be aware of any cabling in their immediate vicinity. Holding interface devices leaves the user the feeling of machine manipulation. Studies have also been performed that focus on wireless interaction without attachment of tethered trackers. Body balance was mapped to navigation according to weight shifts detected using weight sensors (Fleischmann, 1999). The user controls speed and direction through the entire body and balance. However, the weight sensors needed a large platform that makes the implementation difficult in the system requiring a floor display such as IPDs.

A vision-based interface is then one of the most suitable candidates for applications in which the user moves the body within VE systems, allowing full freedom of movement. The ALIVE system (Maes, et al., 1997) is a gesture-language system in which the user interacts with virtual creatures, and the movements are controlled by the position of the user's head, hands, and feet, through vision-based tracking. They have been applied to the control of avatars rather than 3D navigation. Magic Carpet (Freeman, et al., 1998) was designed for navigation in a 2D space, but a 3D VE was not considered. The positional interpretation provided by vision-based tracking has been mapped into navigation of 3D game controls or inside a 3D Internet city by movement of the user's body (Wren, et al., 1997; F. Sparacino, et al., 2002).

All these works have been addressed on design issues of HCI in VEs, but not on evaluation issues of user performance, though the prototypes have been developed for demonstrations. HCI by body position has yet to be examined as a possible interface for navigation in 3D VEs, particularly for a vision-based system with an IPD.

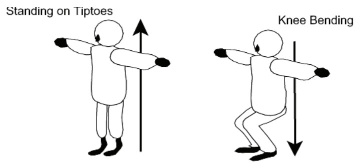
3 SYSTEM OVERVIEW

3.1 Assignment of Body Position

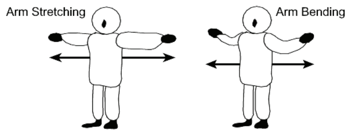
We utilize both arms and legs for navigation, so that the movements contribute to maintaining balance. The user is therefore required to control viewpoint motion and maintain their balance simultaneously. Here, the basic premise of these assignments is an intuitive understanding of the relationship between body position and viewpoint motion. Figure 2 illustrates the assignment of body movements to viewpoint motion. The movements are limited to the following 3 degrees of freedom (DOF) in order to simplify the control parameters:

- Upward and downward movement of the head; standing on tiptoes and bending of knees.
- Horizontal movement of hands; extending and bending arms.
- Left and right rotation of both arms around the axis of the body.

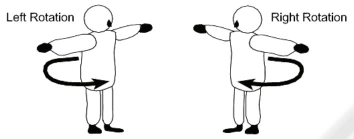
These 3 DOF correspond to vertical motion, forward and backward motion, and counterclockwise and clockwise rotation motion in VEs, respectively.



(a) Upward & Downward Movement



(b) Forward & Backward Movement



(c) Counterclockwise & Clockwise Movement

Figure 2: Assignment of body position.

Flexibility of body movements may generate the various mapping methods between body position and viewpoint motion. We decided to use the above assignment, after the preliminary test obtained the result that the above assignment was more suitable for navigation by body position in VEs, than the other two; 1) the body movements are identical to the above but the viewpoint moves at the same direction to that of extending and bending arms, and 2) upward and downward movements of the arms are assigned to the vertical motion, instead of standing on tiptoes and bending of knees.

3.2 Viewpoint Control

The image sensor must detect the 3D positioning of both edges of the hands and the top of the head. However, we cannot simply apply color or motion segmentation techniques to the extraction of the target, because the background images also change in the screens of the IPD. Depth images are acquired for extracting the body shape by the polynocular stereo machine (Kimura, et al., 1999). Here, we use a primitive structure to simplify the body position, created by generating a straight line between the edges of both hands.

To maintain stability at a point and smooth motion of the viewpoint, the threshold and scale were adjusted based on the properties of the body-position movements using a trial and error process. In forward and backward driving, the width between the both hands when the arms are outstretched is defined as the maximum forward speed, and two thirds of the maximum width is linearly adjusted to represent a speed of zero. When the hands are brought together to less than two thirds of the maximum width, the viewpoint begins to move backwards.

3.3 Implementation

Figure 3 shows the system configuration of the IPD. The VE system TEELeX (Asai, et al., 1999) incorporates a 5.5 multi-screen display designed to generate stereo images with the circular polarization method. The acoustic system is also installed into the display for 3D spatial sounds with 8 speakers. The user's location is taken around the center of the floor screen (3 m x 3 m) along the forward-backward axis diagonal to the screen corners. The multi-camera equipment is placed behind the user at the edge of the screens, such that it does not obstruct the view of the user.

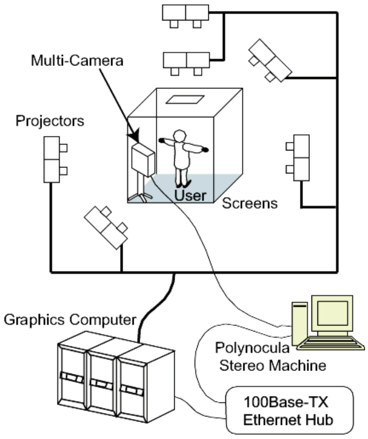


Figure 3: System configuration.

An Onyx2 graphics computer [SGI] is used to generate images of the virtual world. The polynocular stereo machine is a Pentium III 1 GHz PC, connected to the graphics computer. The body-position data from the polynocular stereo machine is multicast with UDP/IP, designating the receiving address. The polynocular stereo machine generates depth images with the size of 280 x 200 pixels at the frame rate of 30 frame/sec, which have the resolution of 4 mm at 1 m, 15 mm at 2 m, and 34 mm at 3 m from the camera, respectively. Figure 4 shows examples of (a) depth image, and (b) extracted skeleton with tracking points.

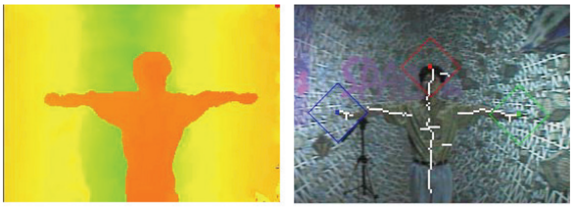

(a) (b)

Figure 4: Examples of (a) depth image, and (b) extracted skeleton with tracking points.

4 EXPERIMENT

We performed an experiment comparing the body-position system with a joystick and a hand-arm gesture interface. The experiment was designed to address evaluation issues on user performance including interaction style and cognitive capability in navigation.

4.1 Method

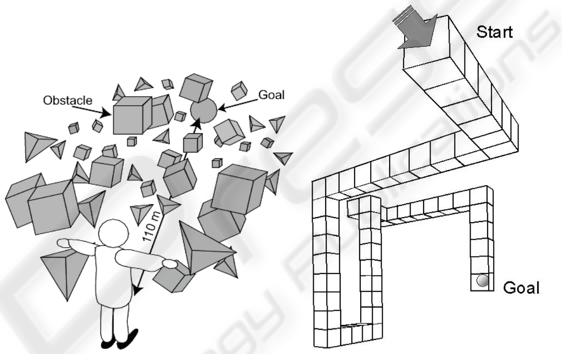
4.1.1 Participants

21 participants took part in the experiment, all of whom were aged from 18 to 28 and had normal vision. Most of the participants had experience of stereoscopic views, but none had experienced navigation using the body-position system in an IPD.

4.1.2 Tasks

We prepared two different tasks for measuring characteristics of the performance. The tasks were designed as follows (Obstacle and Hallway), in which the users were required not only to move directly to a place but also to move dynamically with their surrounding condition and to cope with the cognitive loads in a VE.

In Obstacle, a participant is required to reach a goal as fast as possible while avoiding obstacles in the VE. A sky-blue transparent sphere 5 m in diameter is put as the goal at the opposite end of the VE from the initial point. Blue-white mottled tetrahedrons and brown textured cubes are placed within an area of 60 m from the center of the VE. Figure 5 (a) shows a schematic layout of the obstacles in the VE. The position of the goal is initially fixed, but is modeled so as to elude the user as the user approached. The speed of the goal is set at 0.3 times the viewpoint speed. The 5,000 tetrahedron and cube obstacles are randomly located and revolved around the center of the VE at 0.5 rpm. Therefore, collisions occur even when the viewpoint is motionless. When the viewpoint collides with the obstacles, the viewpoint rebounds elastically resulting in time loss.



(a) Obstacle (b) Hallway

Figure 5: The layout of the VE.

In Hallway, a participant is required to reach a goal through a 3D hallway as fast as possible while memorizing objects such as 3D models, pictures, and Kanji idioms with their loci and directions. This information gathering task was based on the work by D. A. Bowman et al. [1998]. The hallway is composed of 50 blocks of cube with 3 m edges, as shown in Figure 5 (b). 12 objects are placed in the hallway presented with transparent-green walls. We used the popular pictures such as the Mona Lisa and the Scream of Munch, and the familiar idioms with four Kanji characters. When the viewpoint collides with the wall of the hallway, the viewpoint rebounds softly instead of getting out of the hallway.

4.1.3 Apparatus

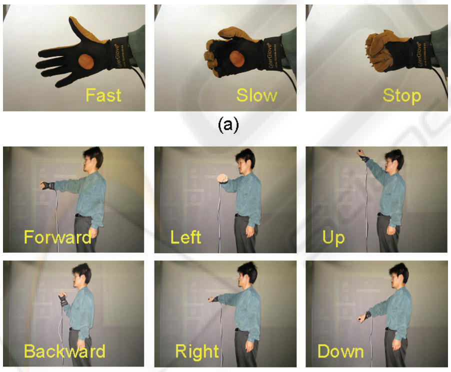
The images were presented via the stereoscopic display at a resolution of 1000 x 1000 dots per screen, and the sounds were output from the 8 speakers. In the tasks, collision sounds notify the participant of the collision besides the background color change to red synchronously. The standing position of the participant was basically fixed around the center in the IPD.

We used a joystick and hand-arm gestures (Figure 6) as interfaces compared with the body-position based navigation system. Wanda $^{\circ,\mathrm{R}}$ [Ascension] was implemented as a joystick, which has been popularly used as a navigation device in IPDs.

(a) (b)

Figure 6: The compared interfaces (a) joystick and (b) hand-arm gestures.

Figure 7 shows the hand-arm gestures. Flathand and fist movements were mapped into adjustment of the speed, and spreading and bending the arm controlled the forward and backward directions, respectively. Shifting the hand left and right rotates the viewpoint left and right, and shifting the hand up and down moves the viewpoint up and down. The hand shapes are detected by resistive bend-sensing of CyberGlove$^{\circ,\mathrm{R}}$ [Immersion] that transforms hand and finger motions into joint-angle data. The positions of the arm are detected by two magnetic sensors of FASTRAK$^{\circ,\mathrm{R}}$ [Polhemus], which are put on the back of the hand and the upper part of the arm.



(a)

(b)

Figure 7: The hand-arm gestures; (a) flathand and fist movements for the speed control, (b) spreading and bending, rotation, and upward and downward movements of the arm.

4.1.4 Measurements

We measured the task completion time for each trial, in which participants complete a series of manipulations for the joystick, the hand-arm gesture interface, and the body-position system. In the Obstacle task, the number of collisions and the loci of the viewpoint were recorded as accuracy measurements for the navigation.

In the Hallway task, the cognitive load is defined as accuracy of the information memorizing task. We counted the number of object/direction/location/surface sets the participant got exactly right, and several variations of partially correct objects, directions, loci, and surfaces. A single response variable V that would encompass all of these values was formulated as

$$V = \sum_{n=1}^{12}(4a_n + 3b_n + 2c_n + d_n) \tag{1}$$

where n expresses the kind of object. a is the number of object/direction/location/surface sets exactly correct, b represents responses that have three of four aspects (object/direction/location/surface) correct, and c is responses where two of four aspects correct, and d is responses where only one of the aspects are correct. The location over one block was regarded as a correct answer.

4.1.5 Procedure

The trials were performed from the Obstacle task for each participant. The Obstacle and Hallway tasks had the trial seven times and twice, respectively. The participants proceeded to the trials using a second interface after the trials using the first interface.

In the body-position system, each task began with capture of the initial position. The participants were told to fix their bodies in a straight standing position with both arms outstretched at shoulder height, being advised to turn their palms upwards so as to reduce fatigue. In the hand-arm gesture interface, each task began with calibration of the hand shapes that slightly depend on a user. The recursive movements of flathand and fist are needed for capturing the hand shapes. The joystick did not need any calibration for the manipulation.

Before the actual experiment began, each participant was allowed to practice controlling the viewpoint until they felt controllable with the interface. During the practice, participants were advised to master staying at one point in the VE, and were allowed to ask any questions about the task at that time. After the word "Start" appeared on the screen, the participants were instructed to reach the goal as quickly as possible and the word "Goal" was displayed to notify the participant of the end of the trial.

4.2 Results

One out of 21 Participants was retired because she did not feel well during the experiment, and the data were excluded from the analysis. The rest of participants experienced all the tasks, but had some trials that were not completed and some extras were done in the tasks.

4.2.1 Obstacle Task

Table 1 lists the experimental results of the task performance for each interface in the Obstacle task. Each value indicates the average among the trials experienced by the participants, and the value inside the parenthesis is the standard deviation. The Obstacle task had 139 trials for the joystick, 140 for the hand-arm gestures, and 139 for the body-position system, respectively. The joystick and body-position system included one participant who did not complete the trial once.

Table 1: Results of the Obstacle task.

	Trials	Time [sec]	Collision number	Speed [km/h]
(a)	139	27.9 (12.1)	0.88 (1.7)	19.6 (15.9)
(b)	140	24.0 (14.5)	0.84 (1.9)	19.5 (6.9)
(c)	139	19.3 (9.3)	1.21 (2.1)	24.3 (15.1)

(a) Joystick, (b) Hand-arm, and (c) Body-position

A one-way ANOVA was performed for the effect of interface on the completion time, collision number, and speed in the VE. The effect of completion time on the interface was statistically significant at F(2,415) = 15.29, $p < 0.01$. Post hoc analyses were conducted in order to compare all possible pairs of the interfaces. The analyses show that the trials with the body-position system were completed significantly faster than those with either joystick or the hand-arm gestures. There was no significant difference between the joystick and the hand-arm gestures in the completion time.

The effects of collision number and speed on the interface had no significant difference at F(2,415) = 1.54 ($p > 0.01$) and F(2,415) = 6.13 ($p > 0.01$), respectively. However, the speed effect had a trend toward significance on the interface. Post hoc analyses comparison of the pairs of the interfaces showed that the navigation with the body-position system had significantly higher speed than that with the hand-arm gestures, and a trend to be faster than that with joystick.

4.2.2 Hallway Task

The table 2 lists the results of the task performance for each interface in the Hallway task. Each value indicates the average among the trials experienced by the participants, and the value inside the parenthesis is the standard deviation. The Hallway task had 34 trials for the joystick, 35 for the hand-arm gestures, and 31 for the body-position system, respectively. Although we designed that each participant had the trial twice for each interface, 6, 5, and 9 participants did complete the trial only once for the joystick, the hand-arm gestures, and the body-position system, respectively, because they felt uncomfortable during the trials.

Table 2: Results of the Hallway task.

	Trials	Time [sec]	Memorizing value	Speed [km/h]
(a)	34	120.3 (35.0)	29.8 (8.8)	7.4 (3.2)
(b)	35	147.0 (32.6)	26.7 (8.6)	5.4 (1.2)
(c)	31	164.3 (29.8)	29.9 (9.9)	5.0 (1.3)

(a) Joystick, (b) Hand-arm, and (c) Body-position

The results of ANOVA showed significant effects of completion time and motion speed on the interface at F(2,97) = 17.48 and F(2,97) = 16.93, $p < 0.01$, respectively, but no significant effect of the interface on the information memorizing at F(2,97) = 1.44, $p > 0.01$. Post hoc analysis comparison showed that trials with the joystick had a significantly shorter time of completion comparing with trials with the body-position system and the hand-arm gestures. The completion time had a trend toward a significant difference between the body-position system and the hand-arm gestures. Post hoc analyses of the motion speed showed the similar result to the completion time. Besides, the trials with joystick had the larger standard deviations in the completion time and the motion speed, compared to those with the hand-arm gestures and body-position system.

4.3 Discussion

We examined the user performance of the navigation using the body-position, compared with the joystick and the hand-arm gestures. We expected that different techniques of navigation would produce different levels of user performance. We found that the task worked as a significant factor that dominated the results as well as the interface was the significant variable.

In the Obstacle task, the body-position had a significant effect on the task performance, and the body-position system was superior to the joystick and the hand-arm interface in the completion time and the motion speed. One of our concerns is that the body-position technique works to drive the motion speed into being kept, and did not simply result in improving performance for the navigation in VEs, because more collisions were observed at higher speeds when using the body-position interface, while there was not clear correlation between the collision frequency and speed for individuals.

One possible scenario is that the body-position interface promoted the participants willingness to travel at higher speeds. This caused situations in which it was easy to collide with the obstacles. The participants who could evade collisions well, reached the goal in a shorter period of time. Otherwise, unavoidable collisions due to high-speed movement resulted in longer times. Another scenario is that the participants tended to dodge obstacles by controlling direction, but not speed when using the body-position interface. Stopping the motion was often observed until the obstacles passed when using the joystick. Keeping the velocity zero was a simple process at the joystick because the participants could just let go off the lever. Conversely, when using the body-position interface, the participants had to search for zero motion position for forward and backward directions.

In the Hallway task, the joystick had an overwhelming effect in the completion time and the motion speed, though the excursion level was quite high between the participants. The significant effect of the joystick could be interpreted as the properties where the joystick makes a user possible to navigate VEs subtly. The travel with the joystick was quite stable, accurately controlling the 3D position, direction, and speed for the viewpoint. The large excursions of the completion time and the motion speed between the participants was possibly caused by differences of the experiences using the similar devices. The excursions between the participants were not so large in the body-position system and the hand-arm gestures, where the participants had no previous experiences, as in the joystick.

Surprisingly, the navigation techniques did not affect information memorizing scores, though we expected that different control techniques would produce different levels of cognitive loads. To interpret the results, we checked the participants' behavior from the recorded video. We found that most of the participants took strategies for performing the Hallway task with careful information memorizing. The participants focused on memorizing the objects, and passed slowly through the hallway, sometimes stopping at the objects for ordering their memory. Thus, the lack of significant difference implies that the information memorizing took priority over the completion time, rather than simply that the level of cognitive loads did not differ between the interfaces.

Overall, the results of the experiment indicated that the body-position system was advantageous on moving around at large areas instead of efficiency or accuracy of navigation in VEs, while the joystick appears to be advantageous on accurate viewpoint motion control. The hand-arm gestures had the middle characteristics between the body-position system and the joystick. It is suggested that the body-position based navigation is suitable for an application getting users enjoying fly-through itself with some entertaining elements.

5 CONCLUSIONS

We have developed a body-position based navigation system as a vision-based interface in an immersive VE. In our implementation, the body-position enables us to navigate VEs via arm and head movements without the need for devices attached to the body. Stable position tracking is achieved using depth images in the IPD. We conducted an empirical evaluation by comparing the body-position system with the joystick and the hand-arm gesture interface. The results of the experiment showed that the performance for the interfaces depended on the task, and the body-position system was advantageous on moving around at large areas instead of efficiency or accuracy of navigation in VEs. This suggests that the body-position interface tends to suit applications in which amusement and enjoyment are important, and conversely may not be suitable for applications that require a high efficiency.

ACKNOWLEDGEMENTS

This study was supported in part by Grants-in-Aid for Scientific Research (18500754).

REFERENCES

K. Asai, N. Osawa, and Y. Y. Sugimoto, Virtual environment system on distance education, Proc. EUROMEDIA'99, pp.242-246 (1999)

D. Bowman, and L. Hodges, Formalizing the design, evaluation, and application of interaction techniques for virtual environments, Journal of Visual Languages and Computing, vol.10, pp.37-53 (1999)

D. Bowman, D. Koller, and L. Hodges, A methodology for the evaluation of travel techniques for immersive virtual environments, Virtual Reality: Research, Development, and Applications, vol.3, pp.120-131 (1998)

D. Bowman, E. Kruijff, J. J. LaViola, Jr., and I. Poupyrev, 3D User Interfaces; Theory and Practice, Addison Wesley (2005)

C. Cruzz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, The cave automatic virtual environment, Communications of the ACM, vol.35, pp.64-72 (1992)

M. Fleischmann, T. Sikora, W. Heiden, W. Strauss, K. Sikora, and J. Speier, The virtual balance: an input device for VR environments, GMD Report, vol.82, pp.1-11 (1999)

W. Freeman, D. B. Anderson, P. A. Beardsley, C. N. Dodges, M. Roth, C. D. Weissman, W. S. Yerazunis, H. Kage, K. Kyuma, Y. Miyake, and K. Tanaka, Computer vision for interactive computer graphics, IEEE Computer Graphics and Applications, vol.18, pp.42-53 (1998)

A. Fuhrmann, D. Schmalstieg, M. Gervautz, Strolling through cyberspace with your hands in your pockets: head directed navigation in virtual environments, Proc. 4th EUROGRAPHICS Workshop on Virtual Environments, pp.216-227 (1998)

J. L. Gabbard, D. Hix, and E. J. Swan, User centered design and evaluation of virtual environments, IEEE Computer Graphics and Applications, vol.19, pp.51-59 (1999)

K. Herndon, A. van Dam, and M. Gleicher, The challenges of 3D interaction, SIGCHI Bull, vol.26, pp.36-43 (1994)

J. Ihren, and K. J. Frisch, The fully immersive CAVE, Proc. 3rd International Projection Technique Workshop, pp.58-63 (1999)

K. Kaur, N. Maiden, and A. Sutcliffe, Interacting with virtual environments: an evaluation of a model of interaction, Interacting with Computers, vol.11, pp.403-426 (1999)

S. Kimura, T. Shinbo, H. Yamaguchi, E. Kawamura, and K. Nakano, A convolver-based real-time stereo machine (SAZAN), Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.457-463 (1999)

J. J. LaViola, D. F. Keefe, and D. Acevedo, Hands-free multi-scale navigation in virtual environment, Proc. 2001 Symposium on Interactive 3D Graphics, pp.9-15 (2001)

P. Maes, B. Blumberg, T. Darrel, and A. Pentland, The ALIVE system: wireless full-body interaction with autonomous agents, ACM Multimedia Systems, vol.5, pp.105-112 (1997)

F Sparacino, C. Wren, A. Azarbayejani, A. Pentland, Browsing 3-D spaces with 3-D vision: body-driven navigation through the Internet city, Proc. International Symposium on 3D Data Processing Visualization and Transmission, pp.214-223 (2005)

J. Vince, Virtual reality systems, ACM Press (1995)

C. R. Wren, F. Sparacino, A. J. Azarbayejani, T. J. Darrell, J. W. Davis, T. E. Starner, A. Kotani, C. M. Chao, M. Hlavac, K. B. Russell, A. Bobick, and A. P. Pentland, Perceptive spaces for performance and entertainment: untethered interaction using computer vision and audition, Applied Artificial Intelligence, vol.11, pp.267-284 (1997)