

## Research Article

# Analysis of User Requirements in Interactive 3D Video Systems

Haiyue Yuan, Janko Čalić, and Ahmet Kondož

*I-Lab, Multimedia Communications Research, Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford Gu2 7XH, UK*

Correspondence should be addressed to Janko Čalić, j.calic@surrey.ac.uk

Received 29 March 2012; Revised 1 August 2012; Accepted 16 August 2012

Academic Editor: Kiyoshi Kiyokawa

Copyright © 2012 Haiyue Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The recent development of three dimensional (3D) display technologies has resulted in a proliferation of 3D video production and broadcasting, attracting a lot of research into capture, compression and delivery of stereoscopic content. However, the predominant design practice of interactions with 3D video content has failed to address its differences and possibilities in comparison to the existing 2D video interactions. This paper presents a study of user requirements related to interaction with the stereoscopic 3D video. The study suggests that the change of view, zoom in/out, dynamic video browsing, and textual information are the most relevant interactions with stereoscopic 3D video. In addition, we identified a strong demand for object selection that resulted in a follow-up study of user preferences in 3D selection using virtual-hand and ray-casting metaphors. These results indicate that interaction modality affects users' decision of object selection in terms of chosen location in 3D, while user attitudes do not have significant impact. Furthermore, the ray-casting-based interaction modality using Wiimote can outperform the volume-based interaction modality using mouse and keyboard for object positioning accuracy.

## 1. Introduction

With the recent development of 3D stereoscopic display technology, 3D movies and 3D TV programmes are becoming a commonplace in our everyday lives. The launch of a number of broadcasted 3D channels, such as Sky 3D and BBC HD, TV viewers can immerse into 3D experience in their own living room. There has been a significant amount of ongoing related research into 3D content capture, production, and delivery. However, to the best of our knowledge, there has been very little research towards meaningful user interaction with real 3D video content. In terms of interaction design, there has been no evidence of differentiation between 2D and 3D video content. However, compared to the 2D video content, 3D video provides an additional viewing dimension and thus offers more immersive experience to the audience. Given this crucial characteristic of 3D video medium, surprisingly little attention has been dedicated towards developing an intuitive interactive technique for 3D video.

The aim of our research is to study user practices and propose technical solutions and design guidelines to develop

intuitive interaction for 3D video content. In this paper, we follow the methodology outlined in our previous paper [1] by initially eliciting user requirements of stereoscopic 3D video interaction with an emphasis on potential interactive functionalities and interaction modalities, followed by a user preference study that investigates the impact of user attitudes, interaction modalities, depth profiles, and dominant eye on the selection task in 3D.

## 2. Related Work

There have been a number of studies that introduced advanced interactive 2D video user interfaces, facilitating intuitive interaction with video content. Two interactive video players, DRAGON (DRAGGable object navigation) [2] and DimP (direct manipulation player) [3], offer direct object manipulation of a video scene. Here, the user can browse the video by selecting and dragging an object in the scene instead of using the timeline slide. In addition, other features such as motion trajectories and annotations were

used by Goldman [4], providing more categories for direct interaction with video content.

There has been a large body of research conducted on 3D interaction with computer generated (CG)/animated content. Bowman et al. [5] outline that 3D interaction consists of three common tasks: object manipulation, viewpoint manipulation, and application control. Object manipulation is usually related to tasks such as pointing, selecting, and rotating. Viewpoint manipulation refers to navigation in the virtual reality environment, as well as manipulating the zooming parameters, while the application control integrates the 2D control user interface with 3D environment to enhance the compatibility of 2D user interface. Thanks to the development of stereoscopic display technology, 3D video is able to offer an immersive experience to wide audience. However, compared with the plethora of research for 2D video interaction, there is very little research focusing on interacting with 3D video content. So far, many researchers have looked into the possible benefits of improving 3D interaction using stereoscopic technique especially in virtual reality and 3D user interface communities. Most of the research evaluate the stereo benefits for completing individual tasks such as selection or positioning. Research [6–10] reveals that stereoscopic viewing can help interaction in terms of improving user performance and depth perception. One of the motivations of our study is to see whether any of the benefits of stereoscopic viewing that have been demonstrated in interaction with 3D CG content would be advantageous for interaction with 3D video content.

A lot of research has been dedicated to develop intuitive interaction modalities for 3D stereoscopic CG content in virtual reality and 3D user interface communities. Park et al. [11] present an interactive 3D TV interface with an intelligent remote controller, which enables the user to change the viewpoint from the controller according to visual attention model. Similarly, Tamai et al. [12] introduce view control interface in 3D stereo environment using Wiimote. Ki and Kwon [13] developed a gaze-based interaction application, which is based on the calculation of degree of eye rotation and pupil centre distance to interact with 3D content. Furthermore, Steincke et al. [14] introduced the concept of interscopic interaction which means that the visualisation of 3D data is using stereoscopic techniques whereas the user interaction is performed via 2D graphical user interfaces. In their more recent work [15], they present an interscopic multitouch surfaces (iMUTS) application to support intuitive interaction with either 2D content and 3D content. In the same context of interscopic interaction, Valkov et al. [24] investigated user preferences of haptic interaction with 3D stereoscopic object on a 2D surface.

### 3. User Requirement Study

The aim is to elicit the user requirement and user preference for interacting with 3D stereoscopic TV in terms of interactive functionalities and interaction modalities. Interview is commonly used as a method to explore specific issue [17]

TABLE 1: Psychographic information.

Number	15
Age	24–30
Gender	Male: 12, female: 3
Occupation	Research students: 8, research staff: 4 Students: 2, employee: 1
3D experience	Extensively: 6, regularly: 3, rarely: 6

in user requirement analysis. Semistructured interview was implemented in this study to identify the requirements.

*3.1. Participants.* This study included a total number of 15 participants. 12 participants are male and 3 participants are female. Participants aged from 24 to 30 years old. 10 participants are from the same research centre and studying or working in 3D video-related research areas. Other 5 participants are nonexpert in 3D video technology. Each participant has previous experience of watching 3D video. Table 1 describes the psychographic information of all participants.

*3.2. Procedure.* The literature review and current practice of using 2D TV/video and 3D TV/video were used as the base to form the structure of the interview. It consists of four parts: (1) to gather background information for each participant; (2) to learn about the current usage of interactive service or applications for 2D video content; (3) to identify the user requirements for interactive functionalities; (4) to elicit the requirements for user interface to facilitate intuitive interactions. All the interviews were recorded using either audio recorder or video recorder, and transcribed entirely afterwards. The categorisation scheme was used to analyze the transcripts.

*3.3. Results.* Our results contain two main parts. One is requirement for interactive functionalities, another one is interaction modalities.

*3.3.1. Interactive Functionality Requirements.* During the interviews, we asked participants about what types of interactive functionalities for 2D video interaction can be applied to 3D video interaction. The discussion resulted in the common agreement that the general interactive functionalities for 2D video interaction such as “play,” “pause,” and “fast forward” can be applied for 3D video interaction. The analysis of transcripts focused on the interaction functionalities, which are tailored for 3D video content but not necessary for 2D video content.

*Changing the Angle of View.* One of the expected functionalities for the future 3D interactive video system was changing the angle of view. However, there was a differentiation of opinions between participants regarding the way of achieving this objective. One proposition was that the user can select an object or a region then manipulate it to change the viewpoint of the scene accordingly. Another proposition was to track

viewer's head to change the angle of view. However based on the current technologies of 3D video production, it is more practical to implement this functionality using 3D multiview video rather than 3D stereoscopic video. The production of 3D multi-view video requires multiple cameras to capture the scene; therefore; it has the capability to render different views to the consumers. On the contrary, the production of 3D stereoscopic video content involves only single frontal parallel stereo camera so that there is limited source of captured scene to be rendered to the consumer. Speaking of content requirement, although movie and sports program have been extensively mentioned in this case, there are some interesting comments regarding this issue.

*Participant 1: "Mostly action one, or in a time of goal, or nice shooting in baseball/basketball, I would like to change view in that time."*

*Participant 2: "For example, to watch live concert or live show, you can choose the position you want to watch thereby you have different angle of view."*

*Zoom In/Out.* Being able to zoom in/out the 3D video content was one demanding requirement. It was expected to allow user to firstly select an object and then change the depth of the chosen object to make the illusion like pull the object close to audience, while keeping other objects in the scene at the original depth and original scale. The opposite recommendation was to zoom in the whole scene while all the objects in the scene should be scaled accordingly to keep the relative scale. There was no conclusive agreement of which way is more appropriate, it is a matter of user personalised choice. The possible solution might be providing compatible zoom in/out which can satisfy both requirements. The potential challenge of this issue in future work is to investigate the user preference of depth sensitivity, which can facilitate zoom in/out functionality and also improve user experience. The demanding video contents for this functionality were sports program, national geographical program, and documentary program were most in demand.

*Textual Information.* Textual information-based interaction allows the user to select an object in the scene to obtain corresponding information of the chosen object, which could be displayed in the format of text on the screen. The inspiration of having this interaction metaphor is related to the fact that the particular scene or object or event happening in the scene may not be the subject to what you are actually watching. If this happens, the response from the user is to search on the Internet or anywhere else. The potential challenge for the textual information-based interaction is to define where the text should be displayed and how the text is displayed in 3D without distraction. Participants would like to use this interaction to access information of the of interest object. For instance,

*Participant 3: "Some program may contain some terminologies which I don't know them before,*

*so probably I have difficulties to understand this program, for instance I cannot understand the movie "Matrix" the first time I watched it."*

The implication of this interaction can be used in documentary program, or getting knowledge of the footballer while watching a football game, or to obtain information of an actor/actress in a movie.

*Dynamic Video Browsing.* All the participants found it interesting when they were watching the demo video of direct manipulation video player [3, 4]. As a concept of select and drag an object in the scene to browse the video instead of timeline slide inspired by direct manipulation video player [3, 4], it can be adapted for 3D video content. The most interesting part for this interaction is to allow the user to browse the video in three dimensions. However, the concern was that the applicability mainly depends on the video content. It was not necessary to have this function for most of the programs, but for application like video analysis such as high-speed collision of objects, sports analysis, and surveillance analysis, where the observer or operator can test the exact moment of incidents happening to make a judge. For example, the operator can directly manipulate the football in the video reply to see whether the football crossed the line or not instead of dragging the timeline controller on the video player.

*3.3.2. User Interface Requirements.* The objective for this part is to find out the user preference of interaction modalities that can support 3D video interaction that proposed in previous stage. The dominant candidate was the hand gesture; however, the concern of using hand gesture was critical. It is mainly because (1) the hand gesture might lack of accuracy in the case of selecting an object; (2) deal; with the chaos caused by involuntary movement; (3) design; an effective system for multiple users; (4) implementing privacy control. Consider the above concerns, the alternatives were various such as small device with touch pad, virtual laser pointer, and digitalised glove. Although there was no conclusive result of user interface, the common agreement was that the user interface should merge the reality and virtual environment to offer immersive experience. The graphical representation of the derived user requirements is depicted in Figure 1.

Last but not least, it is not surprising to find that selection was frequently used as the first step for each interaction mentioned above. In the use case discussed during the interview, participants always firstly select the object in the video and then conduct different interaction with the video content. This is consistent with the findings from previous literatures, which indicate that selection is one of the essential building blocks of all interactive virtual environment systems. It is a process of identifying particular objects, which are the targets for the subsequent actions. The most significant characteristic of 3D video is the depth illusion caused by the disparity between left and right images. Unlike the ordinary selection task, to achieve accurate selection for 3D video content needs to acquire information

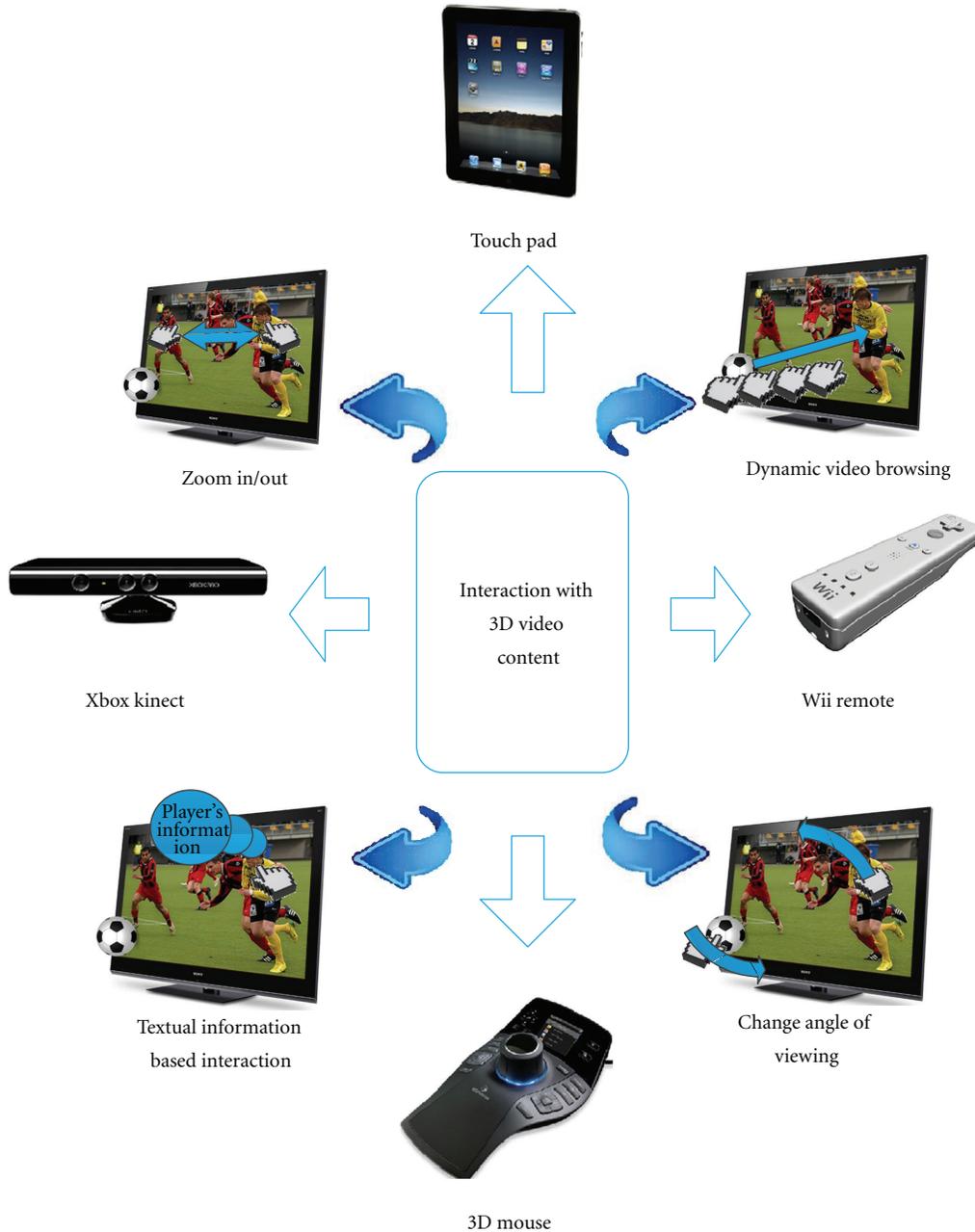


FIGURE 1: Derived user requirements for 3D video interaction.

of disparity, this makes selection in this case more complex and important.

#### 4. User Preference of Object Selection in Stereoscopic 3D Virtual Environment

According to the findings from user requirement analysis, the conclusive agreement among all participants was that selection can be considered as the fundamental requirement for proposed 3D video interaction. Selection has been considered as one of the primary techniques for interactive applications especially in 3D virtual environment [18]. It

is a process of identifying particular objects, which are the targets for the subsequent actions. A large number of research has been looking into various techniques to support accurate and comfort selection. However, few has been focusing on studying user preference and user behavior of selection in virtual environment. In this part, we present a preference study from user's perspective, that investigate the impact of user attitudes, interaction modalities, depth profiles, and dominant eye on object selection in stereoscopic 3D environment.

Selection has been extensively addressed in previous literatures. Most of the selection techniques are variations of

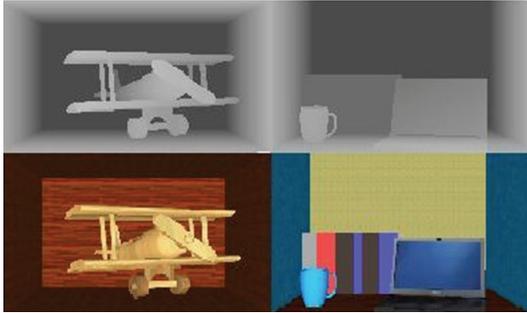


FIGURE 2: Two examples of the 3D scenes with their depth maps used in user preference study.



FIGURE 3: Experimental setup used in the user preference study with Wiimote.

the following two main classes: volume-based selection and ray-based selection [19–21]. Volume-based selection uses the virtual hand/cursor and cone selection to select an object, where requires intersection or collision detection between the virtual hand and the 3D object. As one of the variation of volume based selection, Go-Go interaction technique enables to extend user’s arm length to select object at further distance [16]. Ray-based selection casts a virtual ray into the virtual world to select an object, which is hit by the virtual ray. The way of casting a virtual ray results in two main kinds of variations of ray-based selection. The ray cast from the hand is usually referred as ray-casting technique. The ray is cast from eye and passes through another point in the space that the user can control (e.g., the position of the tip of the finger, or a pointing device). This technique is usually referred to as image plane selection or occlusion selection.

We used both volume-based selection and ray-based selection techniques as the basis to develop two different interaction modalities respectively in this study in order to investigate their impact on the user preference of 3D selective position.

*4.1. Experiment Design.* We conducted series of experiments to ask participants to finish the object selection task in 3D using two interaction modalities and two different user

TABLE 2: Psychographic information.

Number	15
Age	21–28
Gender	Male: 14, female: 1
Occupation	Research students: 13, research staff: 2
3D experience	Extensively: 8, regularly: 5, rarely: 2
Dominate eye	Left: 5, right: 10

attitudes within twenty different depth profiles. The interaction modalities are designed based on the most frequently used selection techniques in 3D interaction. One is the implementation of volume-based selection using mouse and keyboard. Another one is based on ray-casting technique using Wiimote. The user attitudes refer to two different requirements for participants to complete the task: take time to select and select as soon as possible. Depth profile was used to simulate the different 3D scene (see Figure 2).

The reason we created different depth profile was twofold; one was attempting to find out the relationship between user preferred selective position in 3D and associated depth profile, another one was to simulate 3D scene. In order to build a controlled experimental environment, we used 3D stereoscopic CG (Computer Generated)/animated content in this study. Our intention was to learn user behaviour from this experiment and generate results of user preference of object selection in 3D, which can be transferable benefits for the 3D stereoscopic video interaction in our future work.

*4.2. Participants.* There were 15 participants recruited for this experiment. Table 2 describes the psychographic information of all participants. They are all research students in the same research lab. Participants aged from 21 to 28, and contained 1 female and 14 males. All the participants have previous experience of watching 3D stereoscopic video and playing 3D game. Before conducting the experiments, we implemented a Dolman method known as hole-in-the-card to test each participant’s dominant eye, 5 of them are left eye dominant, and 10 of them are right eye dominant. In addition, participants took a Randot stereoacuity test, and all of them had accepted stereo perception.

*4.3. Apparatus.* The experiment was performed on a 46” JVC stereoscopic display with passive polarization glasses (model number GD-463D10). The resolution of the display is  $1920 \times 1080$  and the recommended viewing distance is 2 meters from the screen. The supported format for stereoscopic content is left and right side-by-side representation. We used mouse, keyboard, two Wiimotes with motion plus and a Wii sensor bar in the experiments. We produced and rendered the stereoscopic 3D content using OGRE (open source 3D graphics engine) [22] and use WiiYourself [23] to access Wiimote usage data. Figure 3 presents the setup during the experiment using Wiimote.

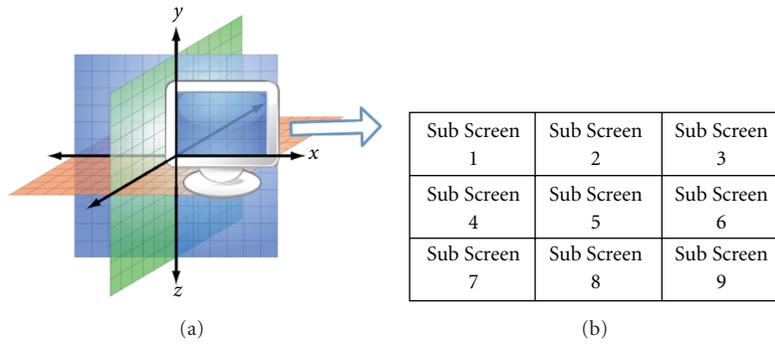


FIGURE 4: (a) Representation of coordinates system (b) Sub-screen distribution.

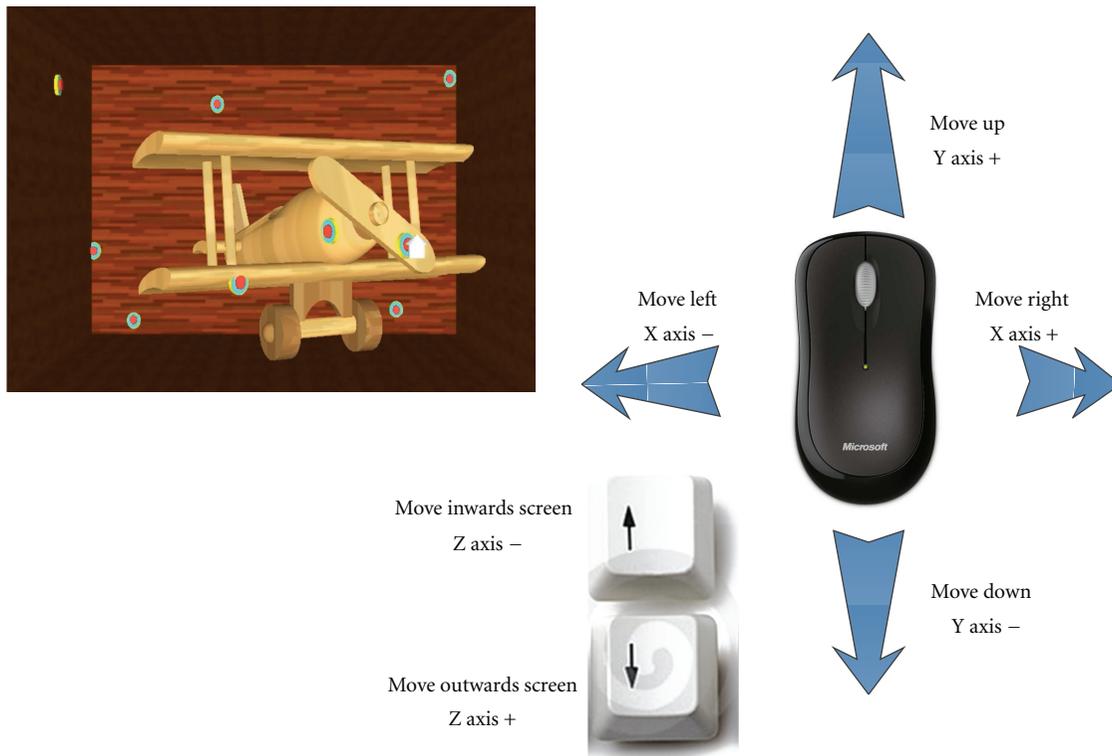


FIGURE 5: Volume-based selection techniques using mouse and keyboard with screenshot.

4.4. Procedure. A within-subjects design was used in which three factors were varied: user attitudes (take your time, as soon as possible), interaction modalities (mouse+keyboard, Wiimote), and depth profiles. As one of the dependent variables, task completion time was calculated from the moment that the object is selected to the moment that object is placed against the destination. Accuracy is another dependent variable, which measured the distance of placed object away from the destination. The smaller the distance is, the higher the accuracy is. The whole experiment is designed based on OGRE coordinates system (please see Figure 4(a)) and consists of two parts. We implemented the volume-based selection technique as a virtual cursor interaction modality in part 1 (please see Figure 5). Mouse is used to

control 2-dimensional movement of the virtual cursor along  $x$ - and  $y$ -axis, and we use arrow key on the keyboard to move the virtual cursor inwards and outwards along  $z$ -axis. The selection is indicated by a mouse click, followed by a collision test activation. If the object is chosen successfully, the bonding box of the chosen object will be visible for the participant. In part 2, we implemented ray-based selection technique (please see Figure 6) to design an interaction modality of virtual laser pointer, which combined Wiimote, Wii motion plus, and Wii sensor bar. The combination of Wiimote and Wii sensor bar is used to locate the position of source of ray. The Wii motion plus is used to detect the degree of pitch and yaw of Wiimote which indicate the orientation of the source of the ray. The selection is executed

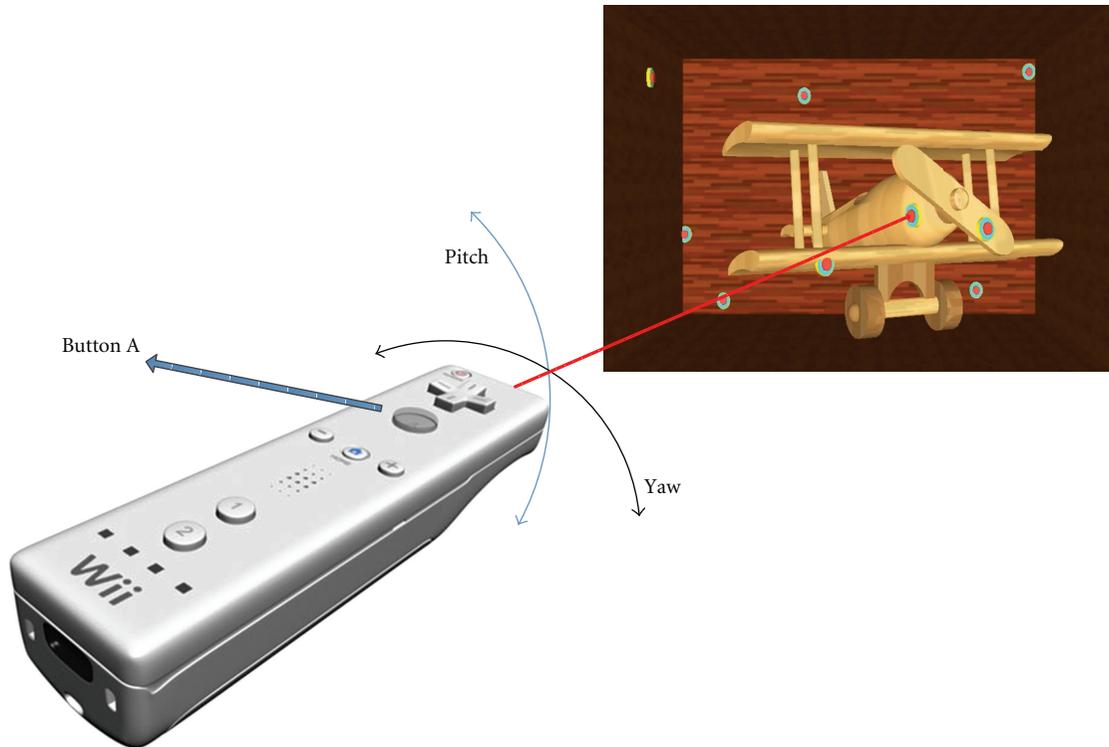


FIGURE 6: Ray-casting selection techniques using Wiimote with screenshot.

by pressing the button A, which emits a ray to the scene. Once the ray hits the object, the appearance of bounding box of the chosen object indicates effective selection.

There were 15 participants in total. We divided them into three groups, 5 participants each group. In order to cancel the learning effect, we apply counterbalancing to assign the order of task to each group. For groups 1 and 3, participants finish experiment part 1 and followed by conducting experiment part 2. For group 2, participants conduct experiment part 1 firstly and then finish experiment part 2.

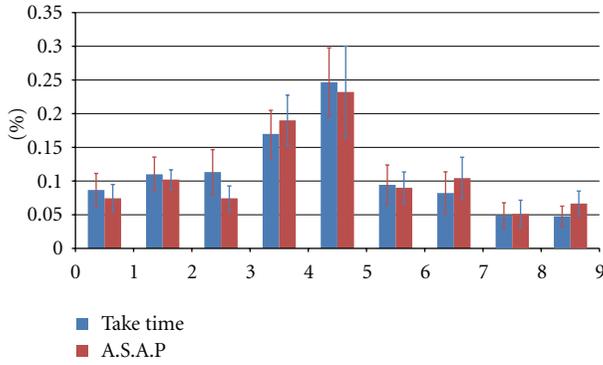
Each part contained 2 sets. For the first set, each participant was asked to take time to choose one object which he/she liked the most and then put the selected object into the destination. For the second set, each participant was required to do the same task as quick as possible. For each set, [each participant needed to finish the selection task with 20 different depth profiles each trial for 3 trials.] The display was divided into 9 subscreens (please see Figure 4(b)). The purpose of introducing subscreens is to find out the popularity of each subregion on the display in terms of selection rate. For each trial, 1 object is allocated to a random position within its corresponding sub screens, so that 9 objects for 9 subscreens in total for participants to choose from. Each participant needs to choose only 1 object for each trial. At end of the experiment, we can obtain the status of how many times the object has been chosen from each sub screen, and thus to get the popularity of each subscreen. Overall each participant completed the task for  $2 \times 2 \times 20 \times 3$

trials for the whole experiment. It took around 30 minutes to complete each part of the experiment and one hour for the whole experiment.

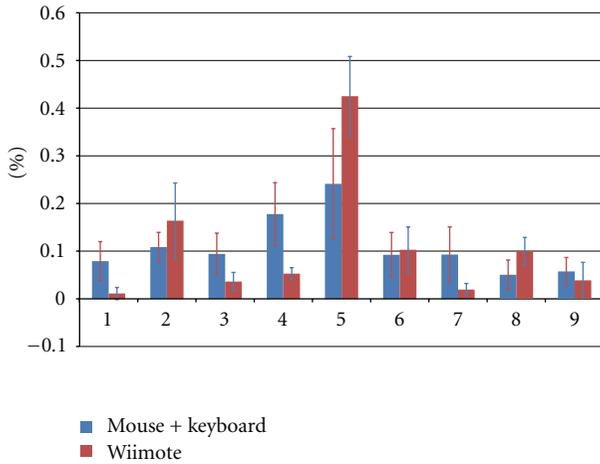
#### 4.5. Experimental Results

**4.5.1. User Attitude Impact.** The participant was asked to choose the object in two different attitudes: one was to *take time* to choose the object which he/she liked the most and then put it into the destination, another attitude was to choose the object A.S.A.P (as soon as possible) and then put it into the destination. ANOVA (analysis of variance) was used to analyze the statistic difference between two attitudes regarding the task completed time and task completed accuracy, respectively. How far the placed object away from the selected destination was used to indicate the accuracy. The smaller the distance is, the higher the accuracy is. ANOVA showed a significant main effect ( $F_{1,1783} = 101.7, P = 0.000$ , see Table 3) of user attitude on the task completed time. It is not surprising that the participant spent about one more second on average to complete the task in Take Time attitude than in A.S.A.P attitude. For the accuracy, there was no significant difference between the two groups ( $F_{1,1783} = 0.99, P = 0.319$ ), which indicated that the user attitude did not have significant impact on the accuracy of completing the task.

In addition, we investigated the impact of user attitude towards the matter of where the participant wants to select



(a) Subscreens



(b) Subscreens

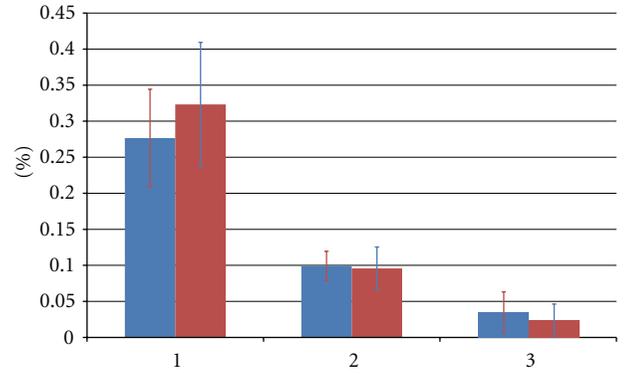
FIGURE 7: Percentage of chosen objects across subscreen at  $X$ - $Y$  plane: (a) user attitudes; (b) interaction modalities.

TABLE 3: Impact of user attitudes on task completion time and accuracy.

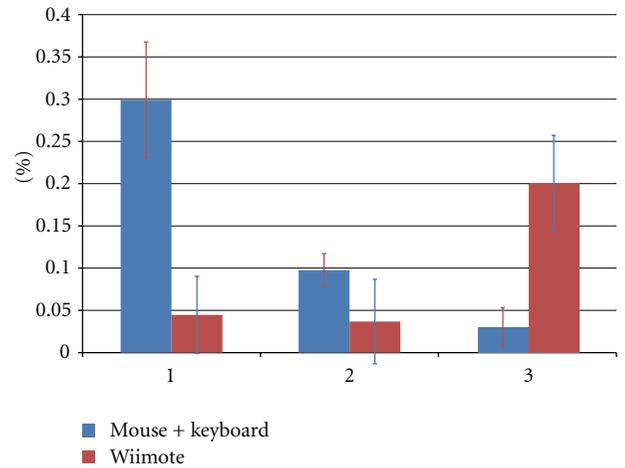
Attitude	Task completed time	Accuracy
	Mean (std)	Mean (std)
Take time	3.73 s (3.00)	0.727 (1.862)
A.S.A.P	2.61 s (1.36)	0.645 (1.637)
ANOVA test	$F_{1,1784} = 101.7, P = 0.00$	$F_{1,1784} = 0.99, P = 0.319$

the object in  $X$ - $Y$  plane and along  $z$ -axis respectively (please see Figure 4). The chosen rate of each sub screen was indicated by the percentage of chosen objects, which was the number of chosen objects divided by the total number of objects allocated in this subscreen. The corresponding distribution of object chosen percentage across sub screens is depicted in Figure 7(a).

Sub screen 4 and sub screen 5 had highest percentage for both user attitude scenarios. Sub screen 2, sub screen 6, and sub screen 7 had around 10 percent of chosen rate. In addition, we did a pairwise correlation test between the two groups. Significant correlation between the two groups ( $r = 0.9483, P = 0.0001$ ) indicated that the user attitude did



(a) Depth cluster



(b) Depth cluster

FIGURE 8: Percentage of chosen objects in the third dimension: (a) user attitudes; (b) interaction modalities.

not affect participant's choices of object selection in  $X$ - $Y$  plane.

Furthermore, we took a look at the participant's preference of object selection along  $z$ -axis. We clustered the position of objects along  $z$ -axis into three categories: near, middle, and far. In OGRE units, 0 at  $z$ -axis indicates that the scene has 0 binocular disparity, which can be referred as screen level. Below 0 units refers to negative binocular disparity, which indicates that the scene is behind the screen. Above 0 units refers to positive binocular disparity, which indicates the scene is in front of the screen. The definition of "near," "middle," and "far" with equivalent OGRE units and disparity in pixels are shown in Table 4. We measured the percentage of chosen objects against all the objects that are in the same depth cluster (please see Figure 8(a)).

For both scenarios, participants preferred objects in front. The pairwise correlation test indicated the significant correlation between the two groups ( $r = 0.9996, P = 0.0017$ ). Therefore, user attitude did not have effect on the participant's preference of object selection in third

TABLE 4: Clustering units.

	Near	Middle	Far
OGRE units ( $u$ )	$u \geq 2$	$-2 < u < 2$	$u \leq -2$
Disparity ( $d$ )	$d \geq 8$	$-10 < d < 8$	$d \leq -10$

TABLE 5: Impact of the interaction modality on the task completion time and its accuracy.

Modality	Task completed time		Accuracy
	Mean (std)		Mean (std)
Mouse + keyboard	3.17 s (2.41)		0.69 (1.75)
Wiimote	3.22 s (2.22)		0.22 (0.61)
ANOVA test	$F_{1,1775} = 0.07, P = 0.79$		$F_{1,1775} = 17.61, P = 0.00$

TABLE 6: ANOVA test across depth profiles.

	Task completed time	Accuracy
Attitude	$F_{19,1784} (P)$	$F_{19,1784} (P)$
Take time	1.52 (0.08)	0.67 (0.849)
A.S.A.P	1.22 (0.2337)	0.7 (0.8198)
Modality	$F_{19,1775} (P)$	$F_{19,1775} (P)$
Mouse + keyboard	1.25 (0.2071)	0.71 (0.807)
Wiimote	0.89 (0.595)	0.82 (0.679)

TABLE 7: Correlation test for each depth profile.

DP	Attitude	Modalities	DP	Attitude	Modalities
	$r (P)$	$r (P)$		$r (P)$	$r (P)$
1	<b>0.89 (0.00)</b>	<b>0.96 (0.00)</b>	11	<b>0.97 (0.00)</b>	0 (1.0)
2	<b>0.97 (0.00)</b>	<b>0.98 (0.00)</b>	12	<b>0.99 (0.00)</b>	0.10 (0.79)
3	<b>0.68 (0.04)</b>	-0.39 (0.29)	13	0.20 (0.61)	-0.59 (0.09)
4	<b>0.64 (0.06)</b>	0.48 (0.19)	14	<b>0.86 (0.00)</b>	<b>0.78 (0.01)</b>
5	0.44 (0.24)	-0.18 (0.64)	15	<b>0.74 (0.02)</b>	0.34 (0.37)
6	0.42 (0.26)	0 (1.00)	16	<b>0.96 (0.00)</b>	0.57 (0.11)
7	<b>0.96 (0.00)</b>	0.29 (0.45)	17	0.39 (0.30)	-0.20 (0.61)
8	0.46 (0.21)	0.58 (0.10)	18	0.57 (0.11)	<b>0.66 (0.05)</b>
9	<b>0.89 (0.00)</b>	<b>0.87 (0.00)</b>	19	<b>0.95 (0.00)</b>	<b>0.94 (0.00)</b>
10	<b>0.85 (0.00)</b>	0.32 (0.40)	20	<b>0.72 (0.03)</b>	0.14 (0.72)

DP stands for depth profile.

dimension. Above analysis was based on volume selection-based interaction modality using mouse and keyboard. Similar results have been found for ray-casting selection-based interaction modality using Wiimote.

**4.5.2. Interaction Modality Impact.** In this part, two interaction modalities were used to find out how do they affect participants' preference of 3D object selection. The dependent variable was task completed time and accuracy, respectively, the independent variable was interaction modality, which contain mouse+keyboard and Wiimote two categories. ANOVA indicated no significant difference ( $F_{1,1775} = 0.07, P = 0.7891$ , see Table 5) of task completed time between the two modalities.

For the accuracy analysis, significant difference ( $F_{1,1775} = 17.61, P = 0.000$ ) between the two interaction modalities suggested that using Wiimote can offer higher accuracy of object positioning.

The comparison of object chosen rate in 2D between two interaction modalities across 9 sub screens is shown in Figure 7(b). The correlation analysis found correlation between two interaction modalities ( $r = 0.7523, P = 0.01$ ) Although it was not highly correlated, sub screen 5 had the highest chosen rate for both scenarios and sub screen 2 and sub screen 6 had similar chosen rate.

The analysis of participant's preference of object selection in third dimension revealed that participant was more willing to choose further objects using Wiimote (see Figure 8(b)). No significant correlation have been found between two modalities in this case ( $r = -0.664, P = 0.5373$ ). The reason of such bias of object selection in third dimension was because of the interaction techniques. The informal post experiment interview also backed up this result. It was easier to use laser pointer like metaphor to reach anywhere in the scene. The interaction modality had significant impact on the preference of object selection along Z-axis, and less impact on the preference of object selection in X-Y plane.

**4.5.3. Depth Profile Impact.** There were 20 different depth profiles in this study, we conducted ANOVA test across different groups (user attitude group and interaction modality group) to investigate the relationship between depth profiles in terms of task completed time and accuracy, respectively. The dependent variable was task completed time and accuracy, respectively, the independent variable was depth profile. As seen from Table 6, there was no significant difference within depth profiles between different user attitudes and between different interaction modalities.

In addition, we compared the correlation of object chosen rate each profile for 20 different depth profiles across different groups. For the majority of the depth profiles, participants had similar preference of object chosen rate across 9 sub screens no matter they take time to select the object or select the object as soon as possible. Only few significant correlation has been found for different interaction modalities. Numbers in bold in Table 7 indicate significant correlation between groups for each corresponding depth profile.

The results indicated that for different depth profiles, user attitudes had less impact than interaction modalities on the user preference of object selection. This is consistent with the previous findings from 4.5.1 and 4.5.2.

**4.5.4. Dominant Eye Impact.** One of the previous works by [24] was about touch interaction with 3D stereoscopic content. The major finding indicated that dominant eye can significantly influence participants' choice of where to interact with the 3D stereoscopic object. Inspired by their work, the aim is to look at the impact of dominant eye in our case of object selection.

The dependent variable was the relative horizontal distance between chosen object and centre of the screen,

TABLE 8: Dominant eye impact.

Distance	Coef	Std	$t$	$P$
Eye	0.586	0.14	4.19	0.000
Constant	-0.669	0.11	-6.17	0.000
$F = 17.52, P = 0.000$				

where minus distance indicated that the object is located at the left side of the centre and vice versa. The independent variable was dominant eye, where left eye dominant was indicated by dummy variable 0 and right eye dominant was indicated by dummy variable 1. Therefore, the hypothesis model is shown as below (1), where  $Y$  refers to distance, and  $X$  refers to eye dominance:

$$Y_i = \alpha + \beta X_i + \epsilon_i. \quad (1)$$

In order to test the difference between left dominant eye and right dominant eye, we have the null hypothesis (2) that there is no statistic difference of distance between participants with different dominant eye, and alternative hypothesis (3) as below:

$$H_0 : \beta = 0; \quad (2)$$

$$H_a : \beta \neq 0. \quad (3)$$

A robust linear regression test has been implemented, and the results (i.e.,  $t = 4.17$ ,  $P = 0.0000$ ,  $F = 17.52$ ,  $P = 0.0000$ , please see Table 8) suggest that we cannot reject the null hypothesis, which indicate that there is a significant difference between dominant eyes:

$$Y_i = -0.669 + 0.586 * X_i + \epsilon_i. \quad (4)$$

Therefore as given in (4), if the participant is left eye dominant (i.e., eye = 0), the relative horizontal distance is  $-0.669$ . On the contrary, if the participant is right eye dominant (i.e., eye = 1), the relative horizontal distance is  $-0.083$ . The results indicated that participants with left dominant eye would choose the object more close to the left, hand side than the participants with right dominant eye.

**4.6. Discussion.** Selection is one of the essential building blocks of interaction in virtual environments. Large amount of work has focused on the selection techniques that facilitate accurate and comfortable object selection in interactive applications. However, little has been done to address the user preference of selective location in virtual environments as well as the impact from different parameters that influence users' choice of object selection. This work addresses these issues from the user's perspective to understand better their behaviour. We have looked into the impact of user attitudes, interaction modalities, depth profiles, and influence of the dominant eye on user's preferred location for selection in three dimensions.

Two tasks were studied: "take your time to select" and "select as soon as possible." These are two distinct user

attitudes towards the task. The expected results before conducting the user study was that the user would have different choices of locations for different attitudes. In addition, it was expected that different interaction modalities might increase the arbitrariness of the results. However, the experimental results revealed surprising findings that there were certain patterns of user preferences and user behaviours.

Regardless of the user attitudes and interaction modalities, participants have similar preference towards locations in the 2D domain, that is, the middle area of the screen is the hot spot for object selection, while the bottom right of the screen has lowest rate of selection.

When it comes to the location in the third dimension, the impact of user attitudes using the same interaction modality is not so significant. Nevertheless, different interaction modalities result in entirely contrasting user preferences of object selection in the third dimension. Using mouse and keyboard, participants prefer to select objects that are closer to the audience, while the chosen rate of object selection gradually decreases as the depth increases. On the contrary, using Wiimote ray-casting approach, the highest chosen rate of object selection is at the deeper end, while the lowest selection rate happens at the front. This is in accord with the characteristics of its underlying interaction techniques, where volume-based selection is more challenging when reaching the objects far from the participants, while the ray-casting selection provides more freedom of navigation in 3D. The investigation of various depth profiles in this study did not provide evidence that its impact would affect participants' preference of object selection in 3D. The analysis of dominant eye impact indicated that participants with left dominant eye would select the object more relatively close to the left of the display. These results can be applied in the design and production of stereoscopic 3D video interaction systems and gaming, enabling user centred approach and enhancing the user experience.

## 5. Conclusion and Future Work

This paper presents a set of user studies that focus on user requirements in 3D video interaction and the user preferences related to object selection in 3D. The results as well as design recommendations are listed below.

- (i) Changing of the angle of view, textual information, zoom in/out, and dynamic video browsing are the interactive functionalities that can facilitate intuitive interaction with 3D video content. Object selection should be considered as the fundamental requirement in the design of the 3D video interaction.
- (ii) Participants have consistent behaviour of object selection over different user attitudes while using the same interaction modality.
- (iii) Participants have significantly different preferences related to object selection, especially in the third dimension while using different interaction modalities.

- (iv) The choice of location for object selection in the third dimension significantly depends on interaction modality.
- (v) The area around the centre of the screen has the highest rate of selection regardless of user attitudes, interaction modalities across depth profiles.
- (vi) The virtual laser pointer based on ray-casting approach to selection using Wiimote can offer higher accuracy of object positioning when compared with the volume-based selection using the mouse and keyboard modality.
- (vii) The participants with left dominant eye prefer selecting the objects relatively closer to the left side of the display.

In order to develop this research further, we will focus on two domains. One will investigate methodologies that will enable interactions with 3D video content proposed in this paper. The second one will conduct experiments to quantify user experience of different interaction modalities aimed at completing proposed interaction tasks as well as investigate the impact of depth to 3D video interaction. These studies will provide the understanding and guidelines of intuitive interaction with stereoscopic 3D video content from users perspective.

## References

- [1] H. Yuan, J. Čalić, and A. Kondoz, "User requirements elicitation of stereoscopic 3d video interaction," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '12)*, 2012.
- [2] T. Karrer, M. Weiss, E. Lee, and J. Borchers, "Dragon: a direct manipulation interface for frame-accurate in-scene video navigation," in *Proceedings of the twenty-sixth annual SIGCHI conference on CHI (CHI '08)*, pp. 247–250, ACM, New York, NY, USA, 2008.
- [3] P. Dragicevic, G. Ramos, J. Bibliowicz et al., "Video browsing by direct manipulation," in *Proceedings of the 26th Annual SIGCHI Conference on CHI (CHI '08)*, pp. 237–246, ACM, New York, NY, USA, April 2008.
- [4] D. B. Goldman, C. Gonterman, B. Curless, D. Salesin, and S. M. Seitz, "Video object annotation, navigation, and composition," in *Proceedings of the 21st ACM Symposium on User Interface Software and Technology (UIST '08)*, pp. 3–12, October 2008.
- [5] D. A. Bowman, S. Coquillart, B. Froehlich et al., "3D user interfaces: new directions and perspectives," *IEEE Computer Graphics and Applications*, vol. 28, no. 6, pp. 20–36, 2008.
- [6] S. Zhai, W. Buxton, and P. I. Milgram, "The Silk cursor: investigating transparency for 3D target acquisition," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '94)*, pp. 459–464, April 1994.
- [7] J. Boritz and K. S. Booth, "Study of interactive 6 DOF docking in a computerized virtual environment," in *Proceedings of the IEEE Virtual Reality Annual International Symposium (VRAIS '98)*, pp. 139–146, March 1998.
- [8] G. S. Hubona, P. N. Wheeler, G. W. Shirah, and M. Brandt, "The relative contributions of stereo, lighting, and background scenes in promoting 3d depth visualization," *ACM Transactions on Computer-Human Interaction*, vol. 6, pp. 214–242, 1999.
- [9] M. Fujimoto and Y. Ishibashi, "The effect of stereoscopic viewing of a virtual space on a networked game using haptic media," in *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '04)*, pp. 317–320, June 2005.
- [10] R. J. Teather and W. Stuerzlinger, "Guidelines for 3D positioning techniques," in *Proceedings of the Conference on Future Play (Future Play '07)*, pp. 61–68, ACM, November 2007.
- [11] M. C. Park, S. K. Kim, and J. Y. Son, "3D TV interface by an intelligent remote controller," in *Proceedings of the 1st International Conference on 3DTV (3DTV-CON '07)*, pp. 1–4, May 2007.
- [12] M. Tamai, W. Wu, K. Nahrstedt, and K. Yasumoto, "A view control interface for 3D tele-immersive environments," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 1101–1104, June 2008.
- [13] J. Ki and Y. M. Kwon, "3D gaze estimation and interaction," in *Proceedings of the 3DTV-Conference: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON '08)*, pp. 373–376, May 2008.
- [14] F. Stenicke, T. Ropinski, G. Bruder, and K. Hinrichs, "Interscopic user interface concepts for fish tank virtual reality systems," in *Proceedings of the IEEE Virtual Reality Conference (VR '07)*, pp. 27–34, March 2007.
- [15] J. Schöning, F. Steinicke, A. Krüger, uger, K. Hinrichs, and D. Valkov, "Bimanual interaction with interscopic multi-touch surfaces," in *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (INTERACT '09)*, pp. 40–53, Springer, Berlin, Germany, 2009.
- [16] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: nonlinear mapping for direct manipulation in vr," in *Proceedings of the 9th annual ACM symposium on UIST (UIST '96)*, pp. 79–80, ACM, New York, NY, USA, 1996.
- [17] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons, 2002.
- [18] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, *3D User Interfaces: Theory and Practice*, Addison Wesley Longman, Redwood City, Calif, USA, 2004.
- [19] A. Steed, "Towards a general model for selection in virtual environments," in *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI '06)*, pp. 103–110, March 2006.
- [20] R. J. Teather and W. Stuerzlinger, "Pointing at 3D targets in a stereo head-tracked virtual environment," in *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI '11)*, pp. 87–94, March 2011.
- [21] A. Douglas Bowman, *Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application [Ph.D. thesis]*, Atlanta, Ga, USA, 1999, AAI9953819.
- [22] OGRE, <http://www.ogre3d.org/>.
- [23] WiiYourself! Native C++Wiimote Library v1.15 RC3, <http://wiiyourself.gl.tter.org/>.
- [24] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs, "2D touching of 3D stereoscopic objects," in *Proceedings of the 29th Annual CHI Conference on Human Factors in Computing Systems (CHI '11)*, pp. 1353–1362, May 2011.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

