

# HOW DO USERS SELECT STEREOSCOPIC 3D CONTENT?

*Haiyue Yuan, Janko Čalić, Ahmet Kondož*

{h.yuan, j.calic, a.kondoz@surrey.ac.uk}

I-Lab, Multimedia Communications Research,  
Centre for Vision, Speech and Signal Processing, University of Surrey

## ABSTRACT

A large number of research has looked into the three dimensional (3D) related technology especially in 3D content capture, production, compression, and delivery. However, few research has addressed the design practice of interaction techniques for stereoscopic 3D video interaction. This paper presents a study of object selection in 3D from users' perspective. The aim is to learn the user preferences and user behaviour of selecting stereoscopic 3D content. The study suggests that interaction techniques 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 technique using Wiimote can outperform the volume-based interaction technique using the mouse and keyboard for object selection accuracy.

**Index Terms** — 3D interaction, 3D stereoscopic video, Selection, User behaviour, User preferences, User study

## 1. INTRODUCTION

The development of 3D display technologies has introduced a proliferation of 3D video related product into our everyday lives. TV viewers can immerse into 3D experience in their own living room and movie lovers can enjoy more vivid immersive experience in the theater. A significant amount of ongoing related research has addressed the issues regarding to 3D content capture, production and delivery. 3D video provides an additional viewing dimension and thus offers more immersive experience to the audiences. Given this crucial characteristic of 3D video medium, surprisingly little attention has been dedicated towards developing intuitive interactive techniques for stereoscopic 3D video interaction.

The aim of our research is to study user practices and user preference of completing selection task for stereoscopic 3D content. In our previous pilot study [1] of user requirement elicitation for stereoscopic 3D video interaction, we revealed that that object selection is the fundamental requirement for the proposed interactive system. This is consistent with the findings from previous literatures. Selection has been considered as one of the primary techniques for interactive applications especially in 3D virtual environment ([2]). However there is few research has been focusing on studying user preference and user behaviour of selection in virtual environment. In this paper, we present a user study that investigate the impact of user attitudes, interaction techniques, depth profiles, and dominant eye on object selection in stereoscopic 3D environment. Our intention is to learn how users select the stereoscopic 3D object in virtual environment, and generate transferable benefits for the 3D stereoscopic video interaction in our future work.

## 2. RELATED WORK

There have been a number of studies that introduced advanced interactive 2D video user interfaces, facilitating intuitive interaction with 2D video content. DimP (Direct Manipulation Player) developed by [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 time-line slide. In addition, other features such as motion trajectories and annotations were presented by [4], providing more categories for direct interaction with 2D video content. However, the research towards intuitive interaction for stereoscopic 3D video content is still little addressed. 3D interaction with computer generated (CG)/animated content has been extensively studied in virtual reality and 3D user interface research communities. [5] outline that 3D interaction consists of three common tasks: object manipulation, viewpoint manipulation, and application control. Many researchers have looked into the possible benefits from this emerging stereoscopic techniques to improve 3D interaction. Most of the research evaluate the stereo benefits for completing individual tasks such as selection or positioning. Previous research ([6, 7, 8]) reveal that stereoscopic viewing can help interaction in terms of improving user performance, and depth perception. As the advantageous benefits from stereoscopic technique, a lot of research has been dedicated to develop intuitive interaction techniques for 3D stereoscopic CG content. [9] present an interactive 3DTV interface with an intelligent remote controller, which enables the user to change the viewpoint from the controller according to visual attention model. Similarly, [10] introduce view control interface in 3D stereo environment using Wiimote. [11] developed a gaze-based interaction application, which is based on the calculation of degree of eye rotation and pupil center distance to interact with 3D content. Furthermore, [12] introduced the concept of interscopic interaction which means the visualization of 3D data is using stereoscopic techniques whereas the user interaction is performed via 2D graphical user interfaces. [13] investigated user preferences of haptic interaction with 3D stereoscopic object on a 2D surface.

## 3. EXPERIMENT

We conducted an experiment to ask participants to complete the object selection task in 3D using two interaction techniques, and two different user attitudes within twenty different depth profiles. The interaction techniques 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 1). In order to build a controlled ex-

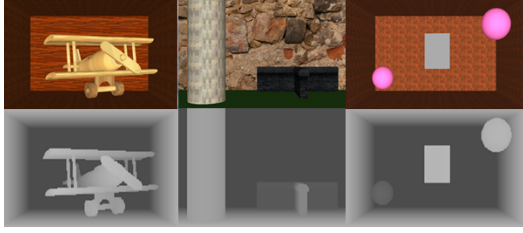


Figure 1. Two examples of the 3D scenes with their depth maps used in user preference study

perimental environment, we used 3D stereoscopic CG(Computer Generated) content in this study.

### 3.1. 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 1920x1080 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 one mouse, one keyboard, two Wiimotes with motion plus and one Wii sensor bar in the experiments.

### 3.2. Participants

There were 15 participants recruited for this experiment. They are all research students in the same research lab, 1 female and 14 male. All the participants have previous experience of watching 3D stereoscopic video and playing 3D game. Dolman method known as hole-in-the-card test was used to determine participants' dominant eye, 5 of them are left eye dominant, and 10 of them are right eye dominant. Before conducting the experiment, each participant took a Randots stereo acuity test, and all participants had accepted stereo perception.

### 3.3. Procedure

The whole experiment consists of two parts. We implemented volume based selection technique as a virtual cursor metaphor in part 1. We implemented ray-based selection technique to design an prototype of virtual laser pointer in part 2. Each part contained

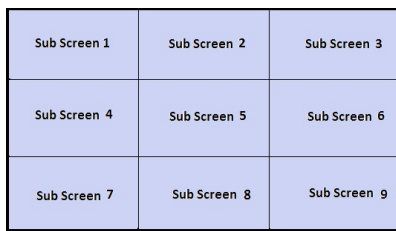


Figure 2. Sub-screens

2 sets. For the first set, each participant was asked to take time to choose 1 object out of 9 random objects which he/she like the most, and then put the selected object into the randomly generated equal distance destination. For the second set, each participant was required to do the 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 sub screens (see Figure 2). For each trial, 1 object was random generated to its respective sub screen within each depth profile. Therefore each participant completed the task for 2x2x20x3 trials for the whole experiment. It took around 30 minutes per participant to complete each part of the experiment, and one hour for the whole experiment. In addition, in order to cancel the possible learning effect, we divided all participants into three groups, each group has assigned different order of using different interaction techniques to finish the experiment.

## 3.4. Results

The results in this section present the impact of different parameters on the user preferences of object selection in 3D.

### 3.4.1. User Attitude Impact

The participant was asked to choose the object in two different attitude, one was to **Take Time** to choose the object which he/she like 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

Table 1. Impact of user attitudes to task completion time and accuracy.

	Completion Time	Accuracy
Attitude	Mean (std)	Mean (std)
Take Time	3.73s(3.00)	.727(1.862)
A.S.A.P	2.61s(1.36)	.645(1.637)
ANOVA Test	F1,1784=101.7, p=.00	F1,1784=.99, p=.319

used to analyze the statistic difference between two attitudes regarding the task completed time and task completed accuracy respectively. ANOVA showed a significant main effect ( $F_{1,1783}=101.7$ ,  $p=.000$ , see Table 1) of user attitude on the task completed time. It is not surprising that participant spent about one more seconds in average to completed the task in Take Time attitude than in A.S.A.P attitude. For the accuracy, there was no significant difference between two groups ( $F_{1,1783}=.99$ ,  $p=.319$ ), which indicated that the user attitude did not have significant impact on the accuracy of completing the task. In addition, we investigated the

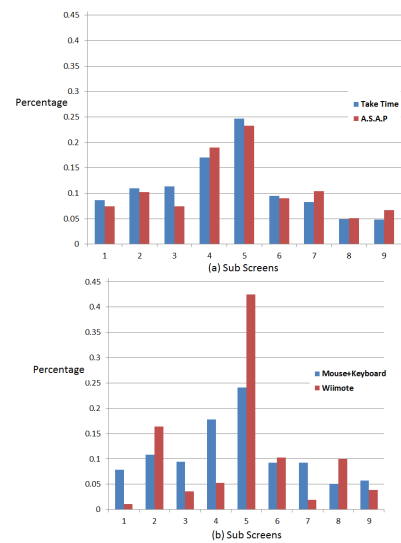


Figure 3. Percentage of chosen objects across sub-screen: a) User attitudes, b) Interaction techniques.

impact of user attitude towards the matter of where the participant wants to select the object in both 2 dimension and third dimension. 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 sub screen. The corresponding distribution of object chosen percentage across sub screens is depicted in Figure 3 (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 two groups. Significant correlation between two groups ( $r=.9483$ ,  $p=.0001$ ) indicated that the user attitude did not affect participants' choices of object selection in 2D domain. Furthermore, we took a look at the user preferences of object selection in third dimension. We clustered the position of objects in third dimension into three categories: near, middle and far. We measured the percentage of chosen objects against all the objects that are in the same depth cluster (please see Figure 4 (a)). For both scenarios, participants preferred objects in front. The

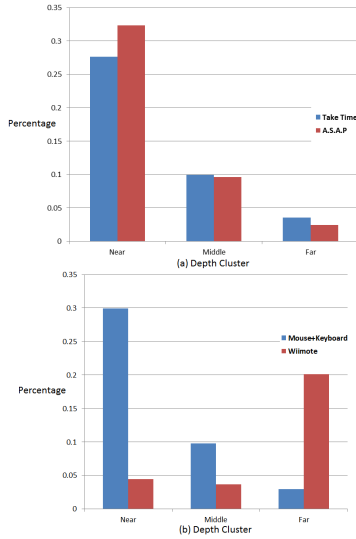


Figure 4. Percentage of chosen objects in the third dimension: a) User attitudes, b) Interaction techniques.

pairwise correlation test indicated the significant correlation between two groups ( $r=.9996$ ,  $p=.0017$ ). Therefore user attitude did not have effect on the user preference of object selection in third dimension. Above analysis was based on volume selection interaction technique using mouse and keyboard. Similar results have been found for ray casting selection interaction technique using Wiimote.

#### 3.4.2. Interaction Technique Impact

In this part, two interaction techniques were used to find out how do they affect participants' preferences of 3D object selection. The dependent variable was task completed time and accuracy respectively, the independent variable was interaction technique, which contain **M+K** (Mouse+Keyboard) and **Wiimote** two categories. ANOVA indicated no significant difference ( $F_{1,1775}=.07$ ,  $p=.7891$ , see Table 2) of task completed time between two techniques. For the accuracy analysis, significant difference ( $F_{1,1775}=17.61$ ,  $p=.000$ ) between two interaction techniques suggested that using Wiimote can offer higher accuracy of object positioning. The comparison of object chosen rate in 2D between two interaction

Table 2. Impact of the interaction technique to the task completion time and its accuracy.

	Completion Time	Accuracy
Technique	Mean (std)	Mean (std)
M + K	3.17s(2.41)	.69(1.75)
Wiimote	3.22s(2.22)	.22(.61)
ANOVA Test	$F_{1,1775}=.07$ , $p=.79$	$F_{1,1775}=17.61$ , $p=.00$

techniques across 9 sub screens is shown in Figure 3 (b) The correlation analysis found correlation between two interaction techniques ( $r=.7523$ ,  $p=.01$ ) 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 user preference of object selection in third dimension revealed that the participant was more willing to choose further objects using Wiimote (see Figure 4 (b)). No significant correlation have been found between two techniques in this case ( $r=-.664$ ,  $p=.5373$ ). 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 technique had significant impact on the user preference of object selection in third dimension, and less impact on the user preference of object selection in 2D.

#### 3.4.3. Depth Profile Impact

There were 20 different depth profiles in this study, we conducted ANOVA test across different groups (user attitude group and interaction technique 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 3, there was no significant difference within depth profiles between different user attitudes, and between different interaction techniques. In addition, we compared the corre-

Table 3. ANOVA test across depth profiles

	Completion Time	Accuracy
Attitude	<b>F19,1784(p)</b>	<b>F19,1784(p)</b>
Take Time	1.52 (.08)	.67(.849)
A.S.A.P	1.22 (.2337)	.7(.8198)
Technique	<b>F19,1775(p)</b>	<b>F19,1775(p)</b>
M + K	1.25 (.2071)	.71(.807)
Wiimote	.89 (.595)	0.82(.679)

lation of object chosen rate each profile for 20 different depth profiles across different groups. For the majority of the depth profiles, participants had similar preferences 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 techniques. Numbers in bold in Table 4 indicates significant correlation between groups for each corresponding depth profile. The results indicated that for different depth profiles, user attitudes had less impact than interaction techniques on the user preference of object selection. This is consistent with the previous findings from 3.4.1 and 3.4.2.

#### 3.4.4. Dominant Eye Impact

In this part, we are looking at the impact of the dominant eye on the user preferences of object selection. We expected a significant

Table 4. Correlation Test for Each Depth Profile

	Attitude	Technique		Attitude	Technique
DP	r(p)	r(p)	DP	r(p)	r(p)
1	<b>.89(.00)</b>	<b>.96(.00)</b>	11	<b>.97(.00)</b>	0(1.0)
2	<b>.97(.00)</b>	<b>.98(.00)</b>	12	<b>.99(.00)</b>	.10(.79)
3	<b>.68(.04)</b>	-.39(.29)	13	.20(.61)	-.59(.09)
4	<b>.64(.06)</b>	.48(.19)	14	<b>.86(.00)</b>	<b>.78(.01)</b>
5	.44(.24)	-.18(.64)	15	<b>.74(.02)</b>	.34(.37)
6	.42(.26)	0(1.00)	16	<b>.96(.00)</b>	.57(.11)
7	<b>.96(.00)</b>	.29(.45)	17	.39(.30)	-.20(.61)
8	.46(.21)	.58(.10)	18	.57(.11)	<b>.66(.05)</b>
9	<b>.89(.00)</b>	<b>.87(.00)</b>	19	<b>.95(.00)</b>	<b>.94(.00)</b>
10	<b>.85(.00)</b>	.32(.40)	20	<b>.72(.03)</b>	.14(.72)

DP stands for Depth Profile

difference of user choices between left dominant eye participants and right dominant eye participants. The dependent variable was the relative horizontal distance between chosen object and center of the screen, where minus distance indicated that the object is located at the left side of the center 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. A robust linear regression test has been imple-

Table 5. 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				

mented, the result ( $t = 4.17$ ,  $p = .0000$ ,  $F = 17.52$ ,  $p = .0000$ , please see Table 5) suggests a significant difference between dominant eyes. 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 the participant with left dominant eye would choose the object more close to the left hand side than the participant with right dominant eye.

#### 4. DISCUSSION AND CONCLUSION

This paper presents a study that addresses the issues from the participants' perspective to have a better understanding of interacting with stereoscopic 3D content. We have looked into the impact of user attitudes, interaction techniques, depth profiles and influence of the dominant eye on participants' preferred location for selection in three dimensions. Regardless of the user attitudes and interaction techniques, participants have similar preference towards locations in the 2D domain, i.e. 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 technique is not so significant. Nevertheless, different interaction techniques 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 the interaction techniques, where volume based selection is more challenging when reaching the objects far from the user, while the ray-casting selection provides more freedom of navigation 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 centered design approach and enhancing the user experience.

## References

- [1] H. Yuan, J. Čalić, and A. Kondoz, "User requirements elicitation of stereoscopic 3d video interaction," in *Proc. of ICME*. 2012, IEEE.
- [2] D.A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev, *3D User Interfaces: Theory and Practice*, Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.
- [3] P. Dragicevic and et al, "Video browsing by direct manipulation," in *Proceedings of CHI '08*, New York, NY, USA, 2008, pp. 237–246, ACM.
- [4] D.B. Goldman and et al, "Video object annotation, navigation, and composition," in *Proceedings of UIST '08*, New York, NY, USA, 2008, pp. 3–12, ACM.
- [5] D.A. Bowman, S. Coquillart, B. Froehlich, M. Hirose, Y. Kitamura, K. Kiyokawa, and W. Stuerzlinger, "3d user interfaces: New directions and perspectives," *Computer Graphics and Applications, IEEE*, vol. 28, no. 6, pp. 20–36, nov.-dec. 2008.
- [6] S. Zhai, W. Buxton, and P.I. Milgram, "The silk cursor: investigating transparency for 3d target acquisition," in *Proceedings of CHI '94*, New York, NY, USA, 1994, pp. 459–464, ACM.
- [7] J. Boritz and K.S. Booth, "A study of interactive 6 dof docking in a computerised virtual environment," in *Virtual Reality Annual International Symposium, 1998. Proceedings.*, IEEE 1998, 18-18 1998, pp. 139–146.
- [8] M. Fujimoto and Y. Ishibashi, "The effect of stereoscopic viewing of a virtual space on a networked game using haptic media," in *Proceedings of ACE '04*, New York, NY, USA, 2004, pp. 317–320, ACM.
- [9] Min-Chul P., Sung Kyu K., and Jung-Young S., "3d tv interface by an intelligent remote controller," in *3DTV Conference, 2007*, may 2007, pp. 1–4.
- [10] M. Tamai, Wanmin W., K. Nahrstedt, and K. Yasumoto, "View control interface for 3d tele-immersive environments," in *Proc. of ICME*, 23 2008-april 26 2008, pp. 1101–1104.
- [11] Jeongseok K. and Yong-Moo K., "3d gaze estimation and interaction," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008, may 2008, pp. 373–376.
- [12] F. Stenicke and et al, "Interscopic user interface concepts for fish tank virtual reality systems," in *Virtual Reality Conference, 2007. VR '07. IEEE*, march 2007, pp. 27–34.
- [13] D. Valkov and et al, "2d touching of 3d stereoscopic objects," in *Proceedings of CHI '11*, New York, NY, USA, 2011, pp. 1353–1362, ACM.