INVESTIGATION AND EVALUATION OF POINTING MODALITIES FOR INTERACTIVE STEREOSCOPIC 3D TV

Haiyue Yuan, Janko Čalić, Anil Fernando, Ahmet Kondoz
I-Lab, Centre for Vision, Speech and Signal Processing, University of Surrey
h.yuan, j.calic, w.fernando, a.kondoz@surey.ac.uk

ABSTRACT

The recent proliferation of stereoscopic three-dimensional (3D) video technology has fostered a large body of research into 3D video capture, production, compression and delivery. However, little research has been dedicated to the design practices of stereoscopic 3D video interaction. Interaction tasks such as pointing and selection are critical to the consumer’s experience of the 3D video technology. This paper presents investigation of pointing modalities in the context of stereoscopic 3D television (TV). Adopting the ISO 9241-9 standard for multi-directional tapping task, the conducted user study compares and evaluates three pointing modalities: standard mouse-based interaction, virtual laser pointer implemented using Wiimote, and hand movement modality using Kinect. The results suggest that the virtual laser pointer modality is more advantageous than other modalities in terms of user performance and user comfort. In addition, this paper discusses the impact of disparity levels on the pointing tasks.

Index Terms—Stereoscopic 3D, Wiimote, Kinect, Interaction, Pointing, Fitt’s Law, ISO 9241-9

1. INTRODUCTION

The recent development of 3D display technology has brought a whole new experience to the end user, triggering the proliferation of 3D multimedia services such as 3D movies in cinemas and broadcasting of 3D TV. Among the emerging 3D display technologies, stereoscopic 3D displays with the compatible 3D video content have been introduced to the consumer electronic devices market and have become increasingly accessible to the general public. The emergence of 3D video content has raised a lot of interest in the research community, with a considerable amount of work focusing on 3D content capture, production, and delivery. On the other hand, there has been very little research conducted focusing on meaningful user interaction with stereoscopic 3D video content. Having this in mind, the main aim of our research is to study user practices and propose technical solutions and design guidelines and design recommendations to developers of interactive stereoscopic 3D TV.

This paper focuses on the interaction modalities for the interactive 3D TV. In a previous study of user requirements for the interactive 3D TV [1], it is suggested that the interaction modalities such as hand movement with gesture recognition, virtual laser pointer, and 3D mouse can facilitate intuitive interaction with 3D video content. A considerable amount of research has focused on the development of interaction modalities using state-of-the-art consumer electronic devices such as Microsoft Kinect and Nintendo Wiimote. The consumer device manufacturers have mainly focused on the intuitive interaction modalities for interactive TV. The smart TV produced by Samsung captures the hand motion and gesture to offer smart interaction. The LG Magic Remote from LG smart TV enables users to point and click on the smart content, offering similar functionalities to the computer mouse. However, very little research has been conducted addressing the interaction modalities for interactive 3D TV in terms of user performance, user experience and user satisfaction. The existing 2D of 3D CG interaction modalities have distinctive differences to the 3D stereoscopic video systems. Thus, the thorough comparison between them is required in order to provide understanding of user requirements and experience of such interaction modalities.

The following section of the paper outlines the related work in this area. Fitts’ law model and ISO 9241-9 standard are presented in Section 3 of the paper. Section 4 describes the experiment evaluation and Section 5 highlights the design recommendations and concludes the paper.

2. RELATED WORK

There has been a large body of research conducted on 3D interaction with CG content. 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, rotating, etc. Viewpoint manipulation refers to navigation in the virtual reality environment, as well as manipulate the zooming parameters, while the application control integrates the 2D control user interface with 3D environment to enhance the compatibility of 2D user interface. A lot of research has been dedicated to development of intuitive interaction modalities for 3D stereo-
scopic CG content in virtual reality and 3D user interface communities. Park et al. [2] presented an interactive 3DTV interface with an intelligent remote controller, which enables the user to change the viewpoint by choosing one of the candidate viewpoints from the remote controller. The idea is using theory of human visual attention to generate reference viewpoints, which is updated and displayed on the perceptive remote controller. Furthermore, Steinicke et al. [3] introduced the concept of interscopic interaction which means the visualisation of 3D data is using stereoscopic techniques whereas the user interaction is performed via 2D graphical user interfaces. A recent research output of the MUSCADE project [4] introduced a Samsung smart mobile phone with an intelligent remote controller to switch view between 3D video content and 3D CG content.

Mackenzie and Jusoh [5] described an empirical evaluation of two remote pointing devices in comparison with a standard mouse. The results indicated that effectiveness and comfort of the remote pointing devices are significantly worse than mouse. Therefore the further development is required to facilitate better pointing efficiency. However, with the development of the hardware, the interest of developing pointing modalities is increasing especially for large distance interaction. Jota et al. [6] compared grab, point, and mouse interaction modalities for a large-scale displays. They found that point modality achieve overall better results, while grab and mouse were better for specific tasks such as close range or far range tests. Yoo et al. [7] presented a 3D user interface combine gaze and hand gestures captured by a time of flight (TOF) depth camera. The user study indicated that the interaction using hands in air could cause the fatigue. Furthermore, Gallo and Minutolo [8] described an empirical study of comparing a laser style modality using Wiimote and a image-plane pointing modality using digital gloves. Their main findings suggested that the image-plane pointing modality could cause hand and arm fatigue to affect the usability. The laser-style pointing provided better user performance and user satisfaction.

3. FITTS’ LAW & ISO 9241-9 STANDARD

Fitts’ law [9] is a model to describe the relationship between movement time, distance, and accuracy for people engaged in rapid aimed movements [10]. According to Fitts’ law, the movement time (MT) has the relation with the movement distance (D) and target width (W).

\[ MT = a + b \times \log_2(D/W + 1) \]  

(1)

where a and b in Eq. 1 are constants empirically determined through the experiments. The log term is referred as index of difficulty (ID) in Eq. 2, the unit for ID is bits. The formulation for ID is derived from Shannon formulation Eq. 2.

\[ ID = \log_2(D/W + 1) \]  

(2)

The dependent measure of Fitts’ law is throughput (TP). TP is defined in the Eq. 3 and carries the units of ‘bits per second’.

\[ TP = IDe/MT \]  

(3)

where IDe represents effective index of difficulty. IDe is determined in the Eq. 4 by the effective distance De and effective target width We.

\[ IDe = \log_2(De/We + 1); \]  

(4)

We is derived from the observed distribution of selection coordinates in participants’ trials as described in Eq. 5. De is calculated as the mean movement distance from the start-of-movement position to the end-point position.

\[ We = 4.133 \times SD \]  

(5)

where SD is the standard deviation of the end-point positions. It is a convention to use a sub-range of end-point positions, which corresponds to around 96% of the distribution as the effective width. This range is approximately equivalent to 4.133 standard deviations of the end positions.

ISO 9241 In our study, we adopted ISO 9241 standard [11] that presents the requirements for ergonomic design of visual display terminals used for office work. Part 9 of this standardization describes the requirements for non-keyboard input devices. In Annex B of Part 9, there are several performance tests: one-directional tapping test, multi-directional tapping test, dragging test, path following test, tracing test and so on. We used multi-directional tapping task in this study. The primary dependent measure for ISO 9241-9 standard is throughput (TP), which we have mentioned in section 3. In addition, the post-experiment qualitative evaluation of the pointing devices is included in the ISO9241-9 to have participants subjectively assess the aspects of operation, fatigue, comfort, and overall usability.

4. EXPERIMENTAL EVALUATION

A multi-directional tapping task is applied to evaluate the pointing modalities for interactive 3D TV. We use Fitts’ law model to compare three pointing modalities: mouse modality, virtual laser pointer modality using Wiimote, and hand movement modality using Kinect. The rest of this section presents the experiment design and the experimental results.

4.1. Participants & Apparatus

We recruited fifteen subjects to participate in the study as volunteers. They were aged ages from 22 to 30. All participants were right handed. Each participant filled in a pre-experiment questionnaire before conducting the experiment. All of them have previous experience of watching 3D movies. 2 of them watch the 3D movies more than 5 times a year. 6 participants watch the 3D movies less than 5 times a year but more
than once a year. The rest of the participants watch the 3D movies less than once a year. All participants had previous experience using PC and Laptop to play games. 6 participants played game with Wiimote before, while only 3 participants used Kinect before. In addition, each participant took a Randot stereo acuity test, and all of them had accepted stereo perception. The experiment took place in a laboratory equipped with 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 produced and rendered the stereoscopic 3D content using OGRE (Open Source 3D Graphics Engine) [12]. A Mouse, a Wiimote with Motion Plus and a Kinect was used in this experiment. WiiYourself library was used to access Wiimote usage data. Microsoft Kinect Software Development Kit (SDK) was used to read skeleton information and tracking.

### 4.2. Pointing Modality Implementation

The mouse modality was implemented by using the mouse to control the pointer on the screen to complete the pointing task. We implemented the virtual Laser pointer modality using a Wiimote with MotionPlus and a Kinect. The Kinect is placed at center position under the bottom of the display, and the Wiimote is hand held by the participant. The Kinect tracked the right hand of the participant. The 3D coordinates of the tracked right hand located the source of the virtual laser ray. Wiimote with MotionPlus was used to detect the degree of pitch and yaw of Wiimote, which can indicate the orientation of the virtual laser ray The combination of the 3D coordinates and orientation enables user to hold the Wiimote as a virtual laser pointer to move the pointer on screen to complete the pointing task. The Hand Movement Modality was implemented using the Kinect only. The Kinect is placed at the center position on top of the display. For hand movement modality, the participants control the pointer on the screen by moving their right hands.

### 4.3. Procedure & Design

The pointing task was designed based on the ISO 9241-9 multi-directional tapping task. The participants were presented with 12 circular targets, arranged in a circle in the center of the screen (see Figure 1). Each target is separated by 30 degree. Participants were instructed to point each highlighted target as soon as possible using the input device. Following the ISO 9241-9 paradigm, the order of the highlighted target started with the “top most” target and always went across the circle. For both virtual laser pointer and hand movement modality, we implemented automatic selection mechanism in order to keep consistency from different participants. Pointing and selecting the target ended the current trial and began the next one. Figure 1 show the screen-shots of the multi-directional tapping task used in the experiment. Before the experiments started, participants were offered the opportunities to have trials in order to get familiar with the Wiimote and Kinect.

This study used a $6 \times 5 \times 3$ within-subjects design. The independent variables were 6 conditions of different targets size and distance, 5 disparity levels, and 3 pointing modalities. The dependent variables were movement time (s), and throughput (bits per second). The following present more detailed description for independent and dependent variables. We have defined two different sizes of targets and three different distances between targets. The details are given in Table 1. The different combinations of the target size and the distance between targets forms 6 conditions (see Table2).

In addition, 5 disparity levels were utilised: -20 pixel, -10 pixels, 0 pixels, 10 pixels, and 20 pixels (1 pixel approximately equals 0.05 mm on the screen). All targets were presented at a consistent depth in each condition, so that the participants had the same visual depth for all targets in each condition. However, the depth of the targets was varied between each condition. Taking into account of three pointing modalities, there were 30 $(2 \times 3 \times 5)$ scenarios for each pointing modality. Each of the 15 participants completed 12 trials over 30 scenarios for each pointing modality. There

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Target Width</th>
<th>Target Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Small</td>
<td>Small</td>
</tr>
<tr>
<td>2</td>
<td>Small</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>4</td>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td>5</td>
<td>Large</td>
<td>Medium</td>
</tr>
<tr>
<td>6</td>
<td>Large</td>
<td>Large</td>
</tr>
</tbody>
</table>

---

### Table 1. Target Width and Target Distance

<table>
<thead>
<tr>
<th></th>
<th>OGRE</th>
<th>Meter</th>
<th>Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Target Width</td>
<td>0.5</td>
<td>0.05</td>
<td>94</td>
</tr>
<tr>
<td>Large Target Width</td>
<td>1.0</td>
<td>0.1</td>
<td>188</td>
</tr>
<tr>
<td>Small Target Distance</td>
<td>1.5</td>
<td>0.15</td>
<td>282</td>
</tr>
<tr>
<td>Medium Target Distance</td>
<td>2.3</td>
<td>0.23</td>
<td>423</td>
</tr>
<tr>
<td>Large Target Distance</td>
<td>4.4</td>
<td>0.44</td>
<td>819</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Screen-shots of multi-directional tapping task with 12 targets in different target size and target distance
were three different modalities in total. Therefore there were $3 \times 2 \times 3 \times 5 \times 12 \times 15 = 16200$ trials. We divided 15 participants in to three groups. The order of the presentation of the pointing tasks to each group was counterbalanced using a $3 \times 3$ Latin square to offset any learning effects. The dependent variables in this experiment were movement time (MT) and throughput (TP), which can be computed using the Eq. 3, Eq. 4, and Eq. 5 in section 3.

### 4.4. Experimental Results & Discussion

#### 4.4.1. Fitting the Fitts’s Law

We gathered the data from all the participants, and took the average values for each condition. There were 30 conditions for each pointing modality. According to the Eq. 1, and Eq. 2, we can get the following Eq. 6 as the Fitts’ law model.

$$MT = a + b \times IDe$$

(6)

The least-squares linear regression is used to find the fitness of the model in terms of the intercept (a) and slope (b) parameters of the Fitts’ law Eq.6. The following Eq.7, Eq.8, and Eq.9 represent the Fitts’ law equation for the mouse, virtual laser pointer, and hand movement modality respectively.

$$MT = 0.2482 + 0.1998 \times IDe, r^2 = 0.8288$$

(7)

$$MT = 0.3796 + 0.2114 \times IDe, r^2 = 0.9026$$

(8)

$$MT = 0.4995 + 0.2854 \times IDe, r^2 = 0.8884$$

(9)

The $r^2$ of the linear regression between MT and IDe for all pointing modalities approximately equals to 0.9, which indicates a good fit to the Fitts law model. As stated in [10], the intercept should be less than 0.4 seconds to prove the legitimacy of the experimental design. The value of the intercept for hand movement modality is 0.4995. The cause of the large slope is the dwell time due to the hand fatigue during the experiment. The analysis of the post-experiment questionnaire also proves the hand and arm fatigue problem in case of hand movement modality. The value of the intercept for virtual laser pointer modality is 0.3796 seconds, which is close to the accepted threshold of 0.4 seconds. The reason is the design of self calibration mechanism for the Wiimote. The MotionPlus used with Wiimote incorporates a dual-axis tuning fork gyroscope, and a single-axis gyroscope to compute the rotational motion. From time to time, the gyroscope produces accumulated error. However, an automatic calibration mechanism was implemented at the start of each condition of the experiment to eliminate the accumulated error. This is a systematic bias inherent to the design of the self-calibration mechanism of virtual laser pointer modality. Although it is imperceptible to the participants, the delay does affect the outcome of the experiment in terms of MT and TP in the further analysis. Apart from this bias, we have not experienced any additional irregularities caused by the implementation of the employed modalities. In the following data analysis, we eliminated the bias of virtual laser pointer to validate the results.

#### 4.4.2. Effect of Different Factors

Mouse has been the most frequently used input device for PC and laptop. The participants have extensive experience of using the mouse to do various tasks. All the participants can be considered as expert users of this modality. On the contrary, only few participants have the previous experience of using Wiimote or Kinect. In order to avoid the bias caused by the previous experience, we offered trial sessions to the participants to get familiar with Wiimote and Kinect. In addition, we have looked at the data of participants with different level of experience with modalities. We did not find any significant effect of previous experience to the MT and TP.

Results were analysed using ANOVA (analysis of variance) and Turkey-Kramer multiple comparisons at 5% significance level. The statistical report is presented in Table 3.

Table 3. Statistical report. (Significant effects are marked *for p<0.05, **for p<0.01 and ***for p<0.001)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Movement time F(p)</th>
<th>Throughput F(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P)ointing modality</td>
<td>577.34(<em><strong>), 0.4976(</strong></em>), 60.43(***), 0.62(0.65)</td>
<td>1.22(0.15)</td>
</tr>
<tr>
<td>(C)ondition</td>
<td>284.29(<em><strong>), 60.43(</strong></em>), 1.77(0.09)</td>
<td>1.75(0.11)</td>
</tr>
<tr>
<td>(D)isparity level</td>
<td>5.48(*), 0.62(0.65)</td>
<td>1.54(0.12)</td>
</tr>
<tr>
<td>P x C</td>
<td>12.62(**), 1.77(0.09)</td>
<td>1.75(0.11)</td>
</tr>
<tr>
<td>P x D</td>
<td>3.14(*), 1.75(0.11)</td>
<td>1.75(0.11)</td>
</tr>
<tr>
<td>C x D</td>
<td>1.81(0.06), 1.54(0.12)</td>
<td>1.54(0.12)</td>
</tr>
</tbody>
</table>

![Fig. 2. Comparison of MT with different conditions](image)

Overall, there was a significant main effect of both pointing modality and condition on the MT and TP. Figure 2 presents the comparison of MT between modalities for different conditions. Participants spent less time using both mouse and Wiimote to complete the task compare with using Kinect for each condition. The comparison between mouse and virtual laser pointer modality suggest that on average participant can achieve quicker pointing action using virtual laser pointer.

Furthermore, the post hoc Tukey-Kramer test was applied
to compare the correlation of MT between pointing modalities. The results suggest that there is significant difference of MT between different pointing modalities. As well as the pointing modality, the Tukey-Kramer test was employed to compare MT between conditions. The results suggest that except the difference between condition 3 and condition 6, the analysis confirms that there is a significant effect of condition to the MT. In addition, the interaction effect to the MT suggest that the MT across different conditions is dependent on the pointing modalities used. Furthermore, refer to the Table 3, the analysis revealed a significant effect of pointing modalities to the TP. In addition, the analysis of interaction effect between the conditions and the pointing modalities to the TP was depicted in Figure 3. The experiments demonstrated that the virtual laser pointer and mouse has similar TP, which was always greater than hand movement modality. The larger the value of TP represents the better assessment of pointing modalities.

The post hoc Tukey-Kramer test of TP between pointing modalities proves the significance of pointing modalities towards the TP. The pairwise comparison of TP between conditions revealed that there is no significant difference of TP between condition 1 and condition 6, and between condition 3 and condition 6. Despite these two pairs, there is significant impact of condition on the TP for the rest of the compared pairs.

4.4.3. Post Experiment Evaluation

Each participant took a post-experiment questionnaire to assess the modality with its input device. The mean and standard deviations of subjective scores on the thirteen questions of the questionnaire are shown in Figure 4. As we can see from the results, the physical effort required for hand movement modality has scored 3.72 which is much higher than for the mouse and virtual laser pointer modality. This leads to the higher score of arm, shoulder and neck fatigue corresponding to the Question 9, 10, and 11. In addition, both of mouse and virtual laser pointer have better scores than hand movement modality.

5. CONCLUSION

In this paper, we presented an investigation and evaluation of pointing modalities for interactive stereoscopic 3D TV using ISO 9241-9 standard multi-directional tapping task. The Fitts’ law model and qualitative evaluation were used to compare the pointing modalities. We found that the target size, the target distance, and the pointing modality has significant impact on movement time and throughput. However the analysis of disparity level failed to find any significant effect. The results suggest that the mouse and virtual laser pointer modalities tend to achieve very similar user performance. Therefore, due to the user familiarity with the hand-held remote controller as a device used to interact with the TV, we did find the virtual laser pointer modality more appropriate modality for 3D stereoscopic content. The practicality of a table setup with a mouse required to interact with a TV in your living room is somewhat unfeasible. We suggest an upgrade of the current remote controller with more precise pointing that can offer more advanced and intuitive interaction with future TV. Furthermore, the investigation of hand movement modality looked into the issue of defining the individual interaction space. The analysis indicated that either too small or too big interaction space could result worse overall usability of input device. According to the main findings of this study, the im-
Applications and design recommendations for interactive 3D TV can be summarised as follows:

- The virtual laser pointer modality is more accessible and applicable to provide better user performance in terms of movement time and throughput than controller-free hand movement modality.
- A standard remote TV controller with additional functionality can facilitate virtual laser pointer modality and offer better overall usability for 3D content.
- In case of hand movement modality, the results of interaction space experiment provide initial insights and understandings to controller-free hand interaction design for interactive TV.

6. ACKNOWLEDGMENT

Thanks to the support by the European Commission FP7 program, under the MUSCADE Integrated Project.

References