Do Stereo Display Deficiencies Affect 3D Pointing?

Mayra Donaji Barrera Machuca
SIAT, Simon Fraser University
Vancouver, CANADA
mbarrera@sfu.ca

Wolfgang Stuerzlinger
SIAT, Simon Fraser University
Vancouver, CANADA
w.s@sfu.ca

CHI'18 Extended Abstracts, *April 21–26, 2018, Montreal, QC, Canada*

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188540.

Abstract

Previous work has documented that limitations of current stereo display systems affect depth perception. We performed an experiment to understand if such stereo display deficiencies affect 3D pointing for targets in front of a screen and close to the user, i.e., in peri-personal space. Our experiment compares isolated movements with and without a change in visual depth for virtual targets. Results indicate that selecting targets along the depth axis is slower and has less throughput than laterally positioned targets.

Author Keywords

3D Pointing, cursor, selection, Fitts' Law.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces – input devices, interaction styles

Introduction

With the availability of 3D TVs and high quality virtual reality (VR) and augmented reality (AR) devices, there has been an increase in applications that allow users to directly manipulate virtual 3D objects [10]. Examples include using the controller to pick 3D objects in a virtual environment or selecting a specific option in a floating menu. To enable better spatial perception, most such systems use stereo display, i.e., a display that shows two different images to the users' eyes from viewpoints that correspond to the eye positions in a human head. Yet, as common stereo displays show the images at a fixed focal depth, this can cause problems with perceiving conflicting spatial cues. Especially for objects in peri-personal space, the human vision system does not perform well when perceiving target depth on stereo displays, due to several known issues, such as the vergence-accommodation conflict [6], diplopia [2], age related near field vision problems [17] and personal stereo deficiencies [5]. Previously proposed models for 3D selection [12] do not take the effect of visual depth changes correctly into account.

In this research, our primary goal is to understand the effect of current stereo display systems on 3D pointing. We focus only on virtual hand/wand interactions, i.e., selection of 3D targets within arms' reach. All our targets are in peri-personal space, up to about 70 cm away from the user, which requires users to successfully perceive target depth for selection. This task also corresponds to real world reaching movements. We hypothesize that stereo deficiencies affect selection increasingly when there is a change in visual depth between targets. Thus, movements directly towards or away from the viewer, should have worse performance than lateral, i.e., side-to-side movements. Our results show that movements with a change in visual depth suffer in time and throughput.

Related Work

Depth Perception in Peri-personal Space in VR

Non-pictorial depth perception in peri-personal space is based on stereopsis, convergence and accommodation and motion parallax [15]. Previous work has established that depth cue conflicts affect depth perception in common stereoscopic displays [14]. More specifically, the eyes need to focus on the display at a fixed distance in stereo display systems, whereas in the real world they need to (con)verge at different distances to correctly perceive the stereoscopic effect. Renner et al. [15] identified a mean under-estimation, about 74% of the true distance, in virtual environments. This difference is independent of the VR display type, but could be a consequence of each individuals' vision system [5], the vergence angle of the eyes [6] and age [17]. However, previous work has not identified the specific distance cues that cause this effect [9].

3D Selection in Stereo Display Systems

Virtual hand/wand techniques require users to intersect the target with their hand or a wand in peri-personal space, i.e., applies only to distances within arms' reach, and also to targets that are outside the screen. To successfully select a target at such close distances, users need to correctly perceive the position of the target and then move their hand there. Lin and Woldegiorgis [11] reviewed previous work and found that although stereo displays are beneficial for depth-related tasks performed in the near-field, compared to the real world, distance perception is compressed in a stereoscopic view.

Fitts Law and 3D Pointing

Fitts' Law [4] is a widely use model to quantify performance in pointing tasks [16]. The movement time (MT) is predicted by $MT = a + b * \log_2(D / W + 1) = a + b * ID$, where D and W are the target distance respectively size, while a and b are empirically derived via linear regression. The logarithmic term is known as the index of difficulty (ID) and indicates the overall pointing task difficulty. Fitts' Law shows the relationship between the time to hit a target and target distance from the

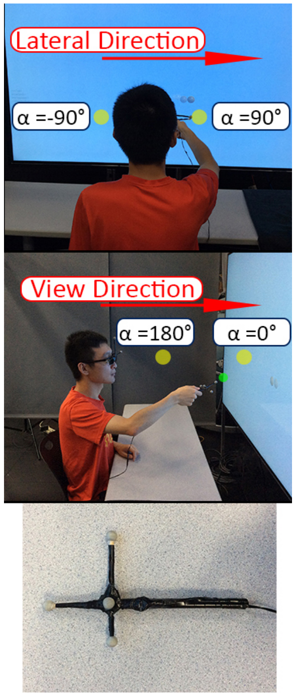

Figure 1: Back and side view of the experimental setting with a close-up of the used input device. Users select targets in front of the screen by moving the wand to them and clicking a button. Movements were either from/towards the screen along the view direction or left/right along the lateral direction. The yellow spheres illustrate exemplary virtual target positions, and the smaller green sphere the virtual cursor inside the virtual environment.

screen and target size. In other words, a small and far away target is more difficult to select than a big and close one. A refined version of Fitts' law specified by the ISO 9241-400 standard [7] combines speed and accuracy into a measurement known as throughput (TP) to make the measurement less dependent on user strategies. TP is predicted by $TP = \log_2(A_e / 4.133 * SD_x + 1) / MT$, where A_e is the amplitude of movement and SD_x is standard deviation. Yet, the traditional formulation of Fitts' Law is not sufficient to predict 3D movement times or throughput in stereo displays [12].

Motivation

Our goal is to identify how the deficiencies of current stereo display systems affect pointing performance. We focus on how vision impairments can affect visually-guided motions [3]. For example, human lateral target discrimination across the field of view is better than depth discrimination. Using typical values for stereo acuity, depth discrimination varies non-linearly between 0.2 and 1 mm within 30-70 cm. Converting typical visual (lateral) acuity into distances yields smaller values, e.g., 0.15 mm at 50 cm. This difference alone gives rise to the hypothesis that selection performance in peri-personal space will be negatively impacted when there is a change in depth between targets. Therefore, movements in the view direction, i.e., between targets with different visual depths, should have worse performance than lateral movements, between targets in the same visual depth. A dependency on the visual target depth is not part of the standard Fitts' law model, where ID only depends on D and W. Therefore, we expect that pointing tasks with the same ID, but with different visual depths will exhibit different movement times and throughput.

User Study

Methodology: **Participants**: We recruited twelve paid participants from the university community (6 female). All participants measured normal when tested for their stereo viewing capability. **Apparatus**: We used a Windows PC with an NVidia GTX970 to display the 3D scene at a resolution at 3840x2160. It consisted of an open space with no additional pictorial depth cues. Eight 250 Hz OptiTrack were used for 3D tracking of the head and a handheld wand (Figure 1). A green 1 cm sphere was used as the virtual cursor. We placed the virtual cursor 2 cm above the wand tip to avoid diplopia with the real wand and it mimicked the physical movement of the pointing wand. The system requires users to intersect a target with the virtual cursor and to click a button on the wand to indicate selection. To provide motion cues for depth perception, the system used head tracking. We used the built-in stereo capabilities of an 85″ 4K stereo Samsung TV, capable of displaying stereo images at 120 Hz, to project the 3D targets outside the screen. End-to-end system latency was about 140 ms. All targets were displayed at a position that matched the eye height of the user (Figure 1): one pair in the view direction with different visual depths but the same positions (α = 0°, 180°), and a pair in the lateral direction with the same visual depths but different positions (α = -90°, 90°). Targets were two yellow spheres placed at a specific distance from the user. Target depth was measured relative to the screen. When intersected by the virtual cursor, targets highlighted. Users only interacted with one of these pairs at any given time. The active target was visible, while the inactive target was invisible. Upon selection, the two targets alternated. **Procedure**: First, participants were tested to see if they could

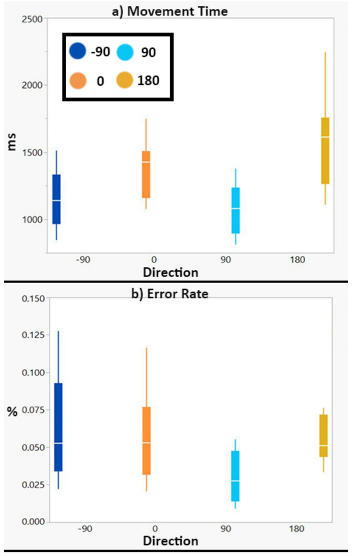

Figure 2. (a) Movement Time, (b) Error Rate, for each condition, showing 95% confidence intervals.

merge stereo targets correctly. Then participants were seated at the middle of the screen behind the table to keep their body parallel to the screen (Figure 1). Participants sat 75 cm away from the display in a chair, 41 cm high. The chair did not allow for movement or rotation. To avoid affecting the view direction, participants were instructed to only move their arm, while keeping their head and body in (approximately) the same position. To eliminate any potential effect of vertical disparity, the visual height of the targets was calibrated to match each participants' eye-level. Subsequently, participants were instructed on the task, and encouraged to practice it. In the training phase, the target shape was different from the one used in the actual experiment. The task was effectively a 3D version of the ISO 9241-400 standard task, with targets positioned along a single axis. Users had to reciprocally select the 3D targets. The distance between targets and their size changed depending on the condition. The conditions were randomly selected without replacement from the available options. Participants needed to select the current target as quickly and accurately as possible. We emphasized that the movement had to be continuous from target to target until they saw a 60 second resting prompt. This prompt occurred between changes of movement axes. If a part of the virtual cursor was inside the target when participants clicked the button, we recorded a successful selection; otherwise, the software recorded a miss. Participants had to select a specific number of targets for each combination of width and distance, i.e., a set. Once all sets for a movement direction were done, the targets changed to the other direction.

Design: The study used a 4x3x3 within-subjects design. The independent variables were movement direction (-90°, 90°, 0° and 180°), target separation (10, 20, and 30 cm) and target size (1.5, 2.5, and 3.5 cm). Overall, participants saw the targets at seven different visual depths. Average target size is approximately constant between changes in movement direction. The dependent variables were movement time (ms), error rate (percentage of targets missed), and throughput (bits per second, bps). There were 11 trials recorded per target ID. Each target circle represented a different index of difficulty, combinations of 3 distances and 3 sizes. This yields nine distinct IDs ranging from 1.94 bits to 4.39 bits. Each participant completed 3 sets of each ID, a total of 594 trials (3 x 11 x 2 x 3 x 3), for a total of 7128 recorded trials overall.

Results: The results were analyzed using repeated measures ANOVA at the 5% significance level. See Figure 2 and 3. We only excluded double clicks (2.4% of the data). As the data was not normally distributed we used an Aligned Rank Transform (ART) for nonparametric analysis before the ANOVA. There was a significant main effect of movement direction on ***movement time*** *($F_{3,32} = 32.93$, $p<.0001$)*. A post-hoc comparison found a significant difference between ɑ=180°, ɑ=0° and the other 2 movement directions, p <.0001. There was a significant main effect between movement directions on ***error rate*** *($F_{3,32} = 5.01$, $p<.0058$)*. A post-hoc comparison found a significant difference between ɑ=90° vs. all the other movement directions, p <.05. (Effective) throughput was computed using the ISO 9241-400 standard adapted to 3D motions [7]. There was a significant main effect of movement direction on ***throughput*** *($F_{3,32} = 55.41$, $p<.0001$)*. A post-hoc comparison found a significant

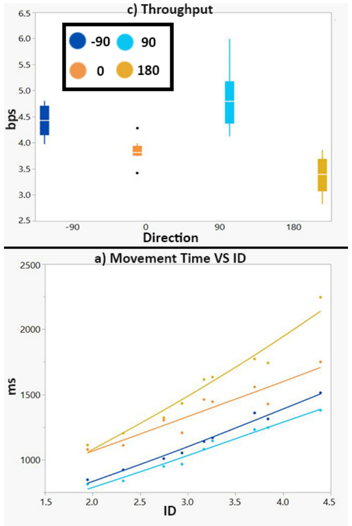

Figure 3. (a) Throughput and (b) MT vs. index of difficulty, for each condition, showing 95% confidence intervals.

difference between ɑ=180°, ɑ=0° and the other two directions, p <.0001. Finally, we analyzed the ***movement paths*** using target re-entry events ($F_{3, 32}$ = 44.31, p<.0001), speed ($F_{3, 32}$ = 32.48, p<.0001), ballistic ($F_{3, 32}$ = 57.09, p<.0001) and correction times ($F_{3, 32}$ = 3.29, p<.03). There was a significant main effect of movement direction on all these measures. We also identified an interaction effect in speed between azimuth and distance for the groups 90-180:0.1-0.3, p=<0.0001, and 0-180:0.1-0.2, 90-180:0.1-0.2, -90-180:0.1-0.3, p=<0.001.

Discussion

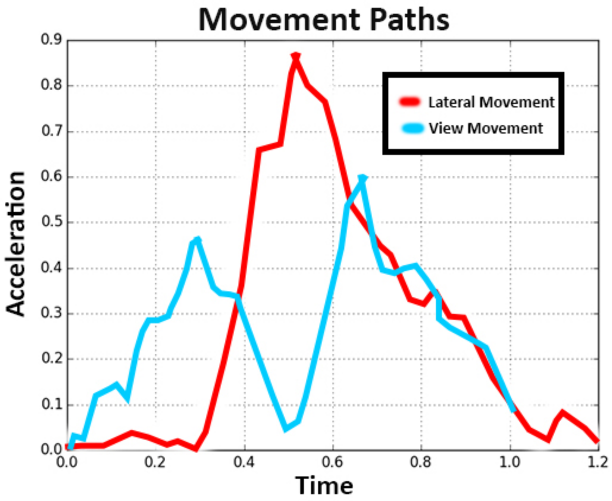

Figure 4. Exemplar second sub-movement in the view direction, and typical movement in the lateral direction.

Movement in the lateral direction had noticeable better performance than movements in the view direction. For example, for an ID of 2.94 the difference is almost 0.4 s between the slowest targets in both directions (180°= 1.432 s vs. -90° = 1.052 s). For movements in the lateral direction this relationship has the same slope, while for movements in the view direction it has not (Figure 3b). For example, a task with an ID of 4.39 has a difference of almost 0.5 s between view direction targets (0° = 1.750 s vs. 180° = 2.245 s), but between lateral direction targets the difference is only around 0.14 s (-90° = 1.513 s vs. 90° = 1.380 s), whereas for an ID of 1.94 the movement times for targets in the same movement direction are similar (0° = 1.079 s vs. 180° = 1.112 s and -90° = 0.847 s vs. 90° = 0.813 s). This effect is also visible in the throughput measure, where movements in the view direction have consistently about 20% bps less than movements in the lateral direction. When analyzing the movement paths, movements in the lateral direction follow previously identified patterns [13]. Yet, movements in the view direction towards the screen have a much-extended correction phase (as identified by the post-hoc test on correction times), which could be interpreted as evidence for participants experiencing issues with depth perception due to stereo display deficiencies (Figure 4). Observations during the experiment also confirm that for movements in the view direction, participants sometimes completely misperceived the target depth, i.e., they made gross depth-estimation errors. Only after they identified their error, they started a second sub-movement to reach the correct position. We can identify such behaviors in the data, as for depth movements 15% of the correction phases had a high speed (more than 20% of maximum). In contrast, only 6% of the lateral movements had such high-speed corrections. This effect is also evident in the post-hoc analysis for target re-entry, where both lateral movements are grouped together and both view direction movements are in different groups. These findings support our hypothesis, as selecting targets with similar IDs exhibit different performance if they are arranged laterally or in depth. Kopper et al.’s work on ray-based interaction [8] identified that target angular size affects performance. Yet, while targets at different visual depths vary in perceptual size, we observe differences (25% for depth movements) beyond any effect that can be explained through angular differences (9%). In summary, our results support the hypothesis that deficiencies of stereo displays affect selection performance.

Conclusion

In general, as the time for movements in the view direction were slower than movements in the lateral direction, our results establish that a change in visual depth between targets affects user performance negatively. Our results not only quantify this effect in movement time, but also confirm that the effect is also visible in terms of error rate and throughput. Based on our results, HMDs that provide correct accommodation (focus) cues might address this issue. In the future, we plan to run further experiments to model the effect of stereo display systems on pointing.

References

1. N. Antony & P. Keir. 2010. Effects of posture, movement and hand load on shoulder muscle activity. J Electromyography Kinesiol 20-2. 191-198.
2. G. Bruder, et al. 2013. Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. Proc. 3DUI'13. 115–118.
3. H. Hong & S. Kang. 2015. Measurement of the lens accommodation in viewing stereoscopic displays. J Soc Inf Disp 23-1. 19-26
4. P. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. 47-6. 381-391.
5. R. Hess, et al. 2015. Stereo vision: The haves and have-nots. I-Perception 6-3. 1-5.
6. D. Hoffman, et al. 2008. Vergance-accommodation conflicts hinder visual performance and cause visual fatigue. J. Vision 8-33. 1–30.
7. ISO 9241-400 Ergonomics of human system interaction - Part 9: Requirements for non-keyboard input devices: International Standard, International Organization for Standardization, 2007.
8. R. Kopper, et al. 2010. A human motor behavior model for distal pointing tasks. Int. J. HUM-COMPUT ST 68-10. 603–615.
9. C. Lin & B. Woldegiorgis. 2015. Interaction and visual performance in stereoscopic displays: A review. J Soc Inf Disp 23-7. 319-332.
10. MetaVision, http://www.metavision.com/
11. A. McDonald, et al. 2012. Spatial dependency of shoulder muscle demands in horizontal pushing and pulling. Applied Ergonomics 43-6. 971-978.
12. A. Murata & H. Iwase. 2001. Extending Fitts’ law to a three-dimensional pointing task. Human Movement Science 20-6, 791–805.
13. K. Nieuwenhuizen, et al. 2009. Insights from Dividing 3D Goal-Directed Movements into Meaningful Phases. Comput. Graph. Appl. 29-6, 44-53.
14. R. Patterson. 2007. Human Factors of 3-D Displays. J Soc Inf Disp 15-11, 861-871.
15. R. Renner, et al. 2013. The perception of egocentric distances in virtual environments - A review. Comput. Surv. 46-2 Article 23.
16. W. Soukoreff & S. MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research. HCI. Int. J. Hum.-Comput. Stud. 61-6, 751-789.
17. C. Zaroff, et al. 2003. Variation in stereoacuity: Normative description, fixation disparity, and the roles of aging and gender. IOVS 44-2, 891-900.