Exploring 3D Manipulation on large Stereoscopic Displays

Marco Speicher, Florian Daiber, Sven Gehring, Antonio Krüger
DFKI GmbH, Saarland Informatics Campus
66123 Saarbrücken, Germany
firstname.lastname@dfki.de

ABSTRACT

In the last years stereoscopic 3D has seen a drastic increase in popularity especially in terms of consumer-ready hardware and software. While the technology for input (smart-phone, Kinect, etc.) as well as output (passive/active stereoscopic and auto-stereoscopic displays etc.) is market ready, only few solutions for natural interaction with such devices exist. In this paper we propose an approach for mobile and gestural interaction with stereoscopic 3D content. We evaluate our technique in a 3D docking task on a large 3D display. In an experiment the interaction technique was evaluated for monoscopic and stereoscopic displayed data. Our results show that the translation and rotation precision benefits from the usage of stereoscopic 3D. Especially for tasks that require simultaneous rotation on all three axes stereoscopic displays outperform monoscopic 3D drastically.

Author Keywords

Public Displays; 3D User Interfaces; Gestural Interaction; Mobile Interaction; Stereoscopic Display

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Nowadays, more and more large-scale urban screen like media facades are used to display stereoscopic content[1,2]. This includes both displaying stereoscopic content on planar surfaces and mapping content onto the 3D surfaces of buildings creating a stereoscopic impression. Making such installations interactive is a cumbersome task, since although 3D interaction has a long research tradition, recent 3D technology lacks natural ways of interacting with the content. Due to recent advances in technology and due to their high availability, using consumer input devices such as smart-phones and depth sensors like the Microsoft Kinect[3] have the potential to close this gap. Most of the existing 3D user interfaces that are available have been developed for virtual reality (VR) environments. Unfortunately, they are often expert systems with complex user interfaces that require high instrumentation (e.g. optical tracking systems). In addition to media facades and installations in urban spaces, an increasing amount of the available public displays are capable of displaying stereoscopic 3D data. However, the manipulation of objects that are displayed with different parallaxes is still a challenging task [23]. In this paper we use smart-phones and depth sensors to overcome the aforementioned limitations and explore natural ways of interaction with 3D applications on large-scale public screens.

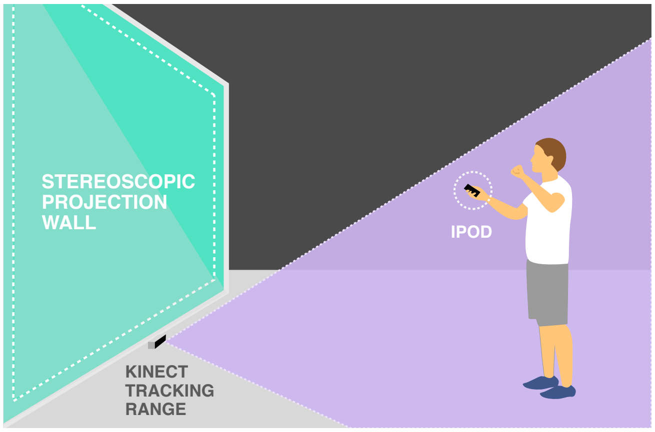

Figure 1. Experimental setup of the docking task.

We evaluate our concepts in a within-subjects experiment, where the participants manipulated 3D objects on a large stereoscopic projection via a Kinect and a mobile phone. We expect significant differences for object translation and rotation, as well as a difference when switching between display modes (stereo/mono), regarding task completion time, translation and rotation precision. The passive haptic property of the mobile device is expected to support the user's spatial orientation and control. The experiment give insights into how affordable input and output devices can be used for full 3D manipulation of objects. The results show that the average translation precision is higher in stereoscopic condition only with regard to the x- and z-axes. Furthermore, they indicate that stereoscopic content notably outperforms monoscopic presented content in tasks that require simultaneous rotation on all three axes. Observations indicate that the interaction technique was easy to use and efficient for experts, and more difficult for novice users.

---

[1] http://www.onionlab.com/diplopia-3d-anagliph-mapping/
[2] http://global.dstrict.com/hyperfacade/
[3] https://dev.windows.com/en-us/kinect

PerDis'16, June 20–June 22, 2016, Oulu, Finland.
ISBN 978-1-4503-4366-4/16/06...$15.00.
http://dx.doi.org/10.1145/2914920.2915018

RELATED WORK

Stereoscopic displays allow users to perceive 3D data in an intuitive and natural way. But interaction with stereoscopic content is still a challenging task, even in VR environments. Steinicke et al. [23] discussed potentials and limitations for using multi-touch interfaces with multi-touch enabled devices to interact with stereoscopic content. Daiber et al. [5] introduced and evaluated 3D interaction techniques using mobile device and Kinect separately on a large-scale stereoscopic display. Our interaction techniques combines their findings and the advantages of physical and virtual techniques with regard to stereoscopic content.

3D manipulation involves the control of 6 degrees-of-freedom (DOF): 3 for position (width, height, depth) and 3 for orientation (yaw, pitch, roll). It can take $10 - 30$ seconds to rotate 3D objects using current desktop interfaces and mice [12], which is much slower than direct object manipulation taking between one and two seconds [24]. Martinet et al. [18] introduced a new 3D manipulation technique based on separation of translation and rotation. Recent publications by Pietroszek et al. [21, 22] are also relevant. In particular, they investigated the use of mobile devices for target selection in 3D, and may provide a useful comparison for this work. Further, Liang et al. [17] investigated how mobile devices can be used as input for distant large 3D displays. Some of these concepts were adapted for the input technique on mobile devices in this work, e.g. tilting the device to rotate an object.

3D manipulation can be evaluated separately in selection, position or rotation tasks. But also as a combination of all in a docking task. Several docking task studies has been conducted to test the object manipulation capabilities of various input devices (e.g., [20, 1, 3, 16, 19, 25]). A docking task is a well established method to evaluate input devices and 3D manipulation techniques for 3D user interfaces [2]. In contrast to related work, our docking task was fairly complex, because the participants had to manipulate up to 6-DOF simultaneously. For example, rotations around more than one axis and translation in 3D space during one single trial. Therefore we choose a smart-phone as an interaction device as they carry various sensors (accelerometer/gyroscope) that have been proven to be suited for such task [11]. However, there is no standardized docking task described in literature.

An important prerequisite for the docking task is to define a metric in order to measure the performance in terms of speed and accuracy. Regarding 3D rotations, the most crucial problem is to define such a metric. Zhai and Milgram quantified the coordination in multiple DOF movement and evaluated their metric with 6-DOF input devices in a docking task [25]. Masliah and Milgram further proposed the m-metric, the product of the simultaneity and efficiency of a trajectory, as a measure for the allocation of control in a 6-DOF docking experiment [19]. In their paper, simultaneity was defined by a normalized error reduction function for each DOF separately.

Recent work suggested to avoid Euler angles in the definition of metrics for complex rotations [15]. They discussed three different kinds of metrics: Rotation Matrices, Euler Angles and Unit Quaternions. Rotation Matrices seem convenient at first view, but matrix representations of rotations generally suffer from several problems in practice, e.g. finite-precision computing and floating-point errors. Euler Angles instead are more compact, stable numerically and relatively computationally efficient. The three angles (yaw, pitch and roll) can be considered to be more intuitive to work with than Rotation Matrices. Due to their simplicity, these angles have been used in many path planning implementations. But unfortunately, if two or more axes happen to align, it will cause a loss of a DOF, which is commonly known as ”gimbal lock”-effect. Additionally, proper sampling, interpolation and distance metrics are serious problems using Euler Angles in the context of manipulation. To avoid such problems, a Unit Quaternion based metric was chosen in this work. Unit Quaternions are also very compact and efficient to work with, and although a slight numerical drift due to floating-point errors could occur, even so the quaternions can be easily renormalized.

For media architecture and media facade installations, it has become common to display stereoscopic content in public spaces. With *Diplopia*, Onionlab created a 3D anaglyph mapping installation for the International Mapping Festival of Girona 2015 in Spain. Diplopia is a stereoscopic installation that revolves around the concept of binocular vision, that is to say, how our eyes merge two separate images to perceive a single object. Two different points of view that arise as a consequence of binocular disparity[4]. With *HYPER FACADE*, the d’strict Arttech factory allows to create 3D media facades displaying stereoscopic content, transcending conventional LED and 2D projections [5]. With the Media Facade Toolkit, Gehring et al. provide a development and testing framework that supports 3D-shaped media facade surfaces as well as display in stereoscopic content [6]. The toolkit further supports the integration of various interaction devices supporting smart-phones as well as gestural interaction with depth sensors. In [14], Köster et al. present a framework for displaying interactive 3D content on media facades and other display environments. The core of this framework consists of a space simulator that is capable of efficiently rendering large amounts of objects and computing their physical behavior in real time while providing means for navigating through the 3D content. Gehring and Krüger supported interaction with 3D media facades by mapping the 3D surface to a 2D map representation using cartographic map projections [7]. Haslkov et al. investigated engaging experiences with horizontal 3D displays[8]. They analyzed how people interacted with and experienced 3D content on such displays.

Based on the related work presented in this section, a 3D manipulation technique was designed for state-of-the-art 3D input devices and evaluated in a comparative user study.

INTERACTION TECHNIQUE

Todays GUI toolkit of the operating systems require the user only to interact in two dimensions. For example, moving

---

[4] http://www.onionlab.com/diplopia-3d-anagliph-mapping/

[5] http://global.dstrict.com/hyperfacade/

mouse pointer in order to draw images or work with documents. In 3D interaction though, the user exchanges information with the computer system in 3D space. While one would expect this to be a more natural interaction type, as humans interact in three dimensions in the real world, no suited commonly used interaction technique so far exists. Thus, this work focuses only 3D interaction techniques. 3D manipulation techniques have a profound impact on the quality of the whole 3D interface: if the user cannot efficiently manipulate objects in virtual environments, other high-level tasks simply cannot be accomplished. Hence, understanding techniques for direct 3D manipulation is an important and necessary step toward developing effective virtual reality applications. So we present a concept of a manipulation technique which takes full advantage of different interaction spaces and leverage their corresponding benefits for the tasks they are best suited for.

The human hand is an ideal direct manipulation input device for virtual reality (VR) [13]. And direct hand manipulation remains the most natural and efficient input method for humans in 3D user interfaces. By equipping the dominant hand (DH) with a 3-DOF sensor, e.g. accelerometer or gyroscope of a mobile device, its position or orientation can be mapped onto the position or orientation of a virtual object in an immersive world. This creates the illusion that the user is able to move or rotate this object using his own hand and leads to a natural and intuitive interaction within immersive worlds. Furthermore, the use of a physical mobile device serves as a passive haptic prop to support the user's spatial orientation and control (c.f. passive real-world interface properties by Hinckley et al. [10]).

The object selection in our proposed interaction technique was realized by a simple toggle mechanism that was activated by a grip gesture of the non-dominant hand (NDH). The positioning of the virtual object was realized by moving the mobile device with the DH (see Figure 2). Tilting the mobile device in a non-isomorphic manner involved relative manipulation of the virtual's orientation. This indirect mapping was chosen due to the physiological constraints of the human hand for rotating a physical object. Initial pilot studies have shown that tracking hand movements using a depth camera is more accurate than using accelerometer of the mobile device for 3D translation of virtual objects. On the other hand, the gyroscope of the mobile device was better for object rotation than tracking the orientation of the user's hand with the depth camera. To stay consistent, the size of the 3D scene was chosen larger than average human arm length, such that the manipulation task requires clutching in order to get to the reference frame.

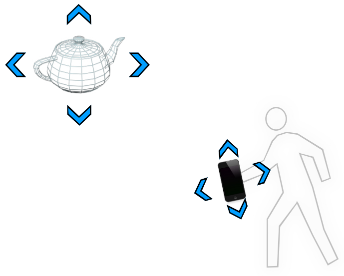

Figure 2. Mobile devices can be used to control objects from afar like using a TV remote control. This figure shows an user moves his mobile device, which results in positioning of the virtual teapot object in a 3D scene.

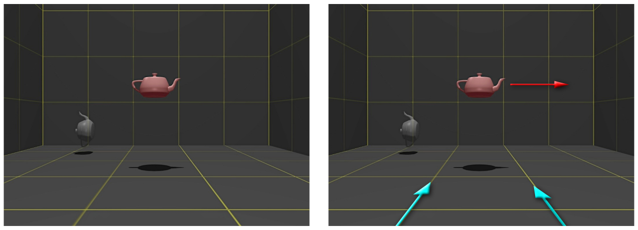

Figure 3. Initial state: red teapot represents the object to be manipulated, the transparent shows an example target position and orientation. Grid-line patterns and shadows were used for adapting artifacts from real world to support user's orientation in the virtual world.

The main benefit of using all 6-DOF without separation is that the participants were able to move and rotate the virtual object simultaneously. That will strongly influence the temporal aspect of the interaction, but also the frustration level in the beginning phase of the task, particularly for novices in 3D manipulation.

USER STUDY

The comparative user study in this paper was conducted in order to evaluate our method for mono- and stereoscopic displayed data and investigate the precision of this method in a docking task. Our first hypothesis to be verified by this experiment was, that the task precision for translation and rotation will be higher for mono- than stereoscopic content (H1). Further, we expected the task completion time will be worse for the monoscopic condition (H2). Moreover, the translation and rotation precision will be higher for simple than for the complex target rotations (H3).

The main approach was to evaluate our interaction technique for 3D object manipulation using affordable and ubiquitous input devices under different display conditions (mono- vs. stereoscopic). The 6-DOF interaction with a large projection wall was separated into 3-DOF from tracking the user's hand by a Microsoft Kinect for object positioning, and 3-DOF from orientation sensors an Apple iPod Touch as mobile device. Moreover, the passive haptic property of a mobile device in combination with peripheral-freedom of the Kinect was used to support the user's spatial orientation and control.

The main performance measures were precision in performing the translation and rotation, as well as task completion time. After performing all trials per display mode, the participants were asked to subjectively rate the workload of the currently passed input technique using the NASA TLX [9] rating scale. But before the experiment started, the participants were also asked to answer a short questionnaire to collect demographic data. The feedback from the participants in this study will help for a final design of further manipulation techniques using these input devices with regard to stereoscopic or monoscopic content.

Participants

All 12 participants (2 female and 10 male) were right-handed and aged between 21 and 35 ($M = 26.6$, $SD = 3.3$) and had variable experience with computer science, 3D modeling and graphic software. Furthermore, they all claimed to have former experience with stereoscopic visualization in different ways (i.e., cinema, TV, other studies, etc.). All of them owned smart-phones and were very familiar in using them, with an average usage of more than 6 hours per day.

Apparatus

All participants were placed in front of a 5 × 3 m^2 projection wall at a distance of 2.5 meters during the trials. The study setup was developed using a scalable rendering environment [4] for rendering the 3D content. The stereoscopic rendering was realized by two projectors with polarized filters. The participants took off the glasses for the monoscopic condition. The system was an Intel Core i5 4x 3.20 GHz CPU with 8 GB of RAM, and a NVIDIA GeForce GTX 660 Ti graphics card. The operating system was Windows 8 and the software was written in C++ and DirectX.

The experiment input devices were a mobile device (Apple iPod Touch 4th generation) and a depth camera (Microsoft Kinect V1). The Kinect detects a joint skeleton consisting of the 3D positions in space of the users hands, shoulders, head, etc. As result, users could interact with the system using hand or body gestures, like grabbing the hands or performing hand, head and body motions. The mobile device has a 3.5" multi-touch wide-screen-display and was connected through a wireless network. It is also equipped with an accelerometer, gyroscope and ambient light sensor (which was not used). Because of the huge amount of transferred data between the mobile devices and the rendering environment, it may be advisable to use an own closed WiFi network to avert network traffic problems.

Task

The goal of the docking task was to fit the virtual object in a second gray and light transparent reference object by changing position and orientation using the input devices. In general, a manipulation task is a combination of selecting, moving and rotating objects. In order to keep the scene as simple as possible, there was only one selectable object placed in the mid center of the scene. The selection itself was performed via direct target selection. The focus of this experiment was on evaluating an interaction technique for manipulating position and orientation of a virtual object on stereoscopic and monoscopic displays. Therefore, the participants were not able to manipulate the scaling of the object and the target. A trial was completed, when the Euclidean distance between the position vector of the object to be manipulated and the target, as well as the difference of the Unit Quaternions of the orientations, were under certain thresholds (5 units for translation and 30 degrees for rotation precision).

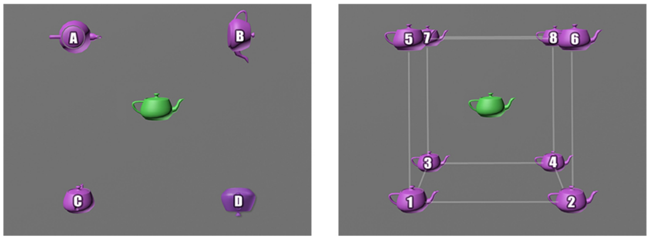

Figure 4. Left: rotation conditions; Right: position conditions

Design

We employed a 2 × 8 × 4 within-subjects factorial design. The factors in this design were *display mode*, *position* and *rotation* in 3D space. According to this, the participants performed 32 trials in each of the two conditions of the display mode (monoscopic and stereoscopic). The amount of the trials for each display mode was the product of 8 position and 4 rotation conditions. The number of positions resulted from the number of corners of a virtual cube, with the teapot in its center (see Figure 4). Hence, all target positions had the same distance to the initial starting position of the object to be manipulated. Furthermore, there were three simple (A,B,C) and one more complex rotation conditions (D). Here, a simple rotation means, the participants only needed to rotate around one axis to achieve the goal. In contrast, in the more complex condition they had to rotate around all three axes. The order of all conditions per factor was counterbalanced across participants, as well as the order of trials was randomized. Finally, each setting results in a total of $12 \times 2 \times 8 \times 4 = 768$ trials conducted.

Only minimal instructions were given to participants in how to use each controller. Thus, no explicit conceptual model was imparted to the participants. The participants were not told that the 3D motion is tracked from Kinect, so that some of the participants thought we would use the accelerometer of the mobile device. In addition, the users decided by themselves how they hold and use the mobile device and were not being influenced by the experimenter. The requirements were to perform each trial as precise as possible within 60 seconds, which results in 45 minutes per participant averagely, including the experimenter's introduction and three mandatory one minute rests. The whole trajectory of each task was logged for later analysis.

Procedure

First of all, the participants were asked to fill out a questionnaire with demographic data. After that, the task goal and the handling of the manipulation technique was introduced to the participant. The participant selected the object to be manipulated by picking it up with his NDH by grabbing it, which initiated at the same time the object manipulation phase. From now, the participant was able to start moving and rotating the virtual object, as long as it was selected.

The selection state was represented by the object color (yellow = selected, red = not). The participant had to grip his NDH to select the object and consequently activated the interaction stream. This indication of position preceded the positioning by specifying position, velocity and acceleration. The velocity was determined by the speed the participant moves his DH and mobile device. Though, the acceleration of each interaction was defined by a constant value. In the positioning phase, the distance and direction to the initial position as well as to the target position were essential parameters to determine the translation distance. The position of the virtual object (x,y,z) was manipulated bi-manually by moving the participant's DH, while grabbing the NDH at the same time. The participant hold the mobile device in the DH during the whole experiment, which finally amplified the immersion and haptic. Further, after the orientation of the virtual object was indicated, the rotation phase were initiated. The rotation phase parameters were target distance, initial and final orientation and, of course, the amount of rotation per axis. Finally, the orientation of the virtual object was effected by tilting the mobile device.

In order to prevent occlusion, the target model was visualized by a semi-transparent representation of the virtual object. When the participant decided that the object to be manipulated was close enough to the target location and also had similar orientation, the experimenter continued with the next trial. All trials for each display mode were conducted in sequence followed by break of 5 minutes before starting a new trial sequence. When the participant has performed all trials of the selected display mode, he were asked to fill out a NASA TLX questionnaire. After all questionnaires were filled out the experimenter disbanded the participant.

RESULTS

Our interaction technique was evaluated for mono- and stereoscopic displayed data. The numbers in the horizontal axis of the charts, which are presented in the following, symbolize the position condition (1-8), whereas the letters encode the rotation condition (A-D). In summary, 32 trials were performed for each display mode (mono vs. stereo).

Task Completion Time

Task completion time was the elapsed time for a user to perform a task, more precisely from the first to the last user interaction in a task. Univariate ANOVA analysis were conducted for position task condition and display mode, regarding task completion time. A significant difference for display mode condition was found ($F(1, 776) = 18.40, p < 0.001$). Furthermore, target position had significant effect on task completion time ($F(7, 776) = 2.04, p < 0.05$). In addition, univariate ANOVA analysis for rotation task condition was conducted, which showed also significant influence on task completion time ($F(3, 776) = 5.58, p < 0.002$).

Translation Task Precision

The translation task precision in this experiment was characterized by the Euclidean distance between object and target position. There was no significant difference for display mode condition regarding translation task precision. A pairwise univariate ANOVA was conducted, which showed significant difference ($F(3, 808) = 3.719, p < 0.01$) for rotation task condition with regard to translation precision, independent from display mode. Although there was no significant difference between target positions regarding translation precision, the average overall precision was higher in mono than in stereo, except for the two upper target positions in the back. This was mainly influenced by the worse precision in translation along y-axis, because translation around x- and z-axis was performed better in stereo than in mono (see Figure 5 for monoscopic condition). Translations along z-axis performing the worst with average precision of 12.13 in stereo and 10.96 in mono.

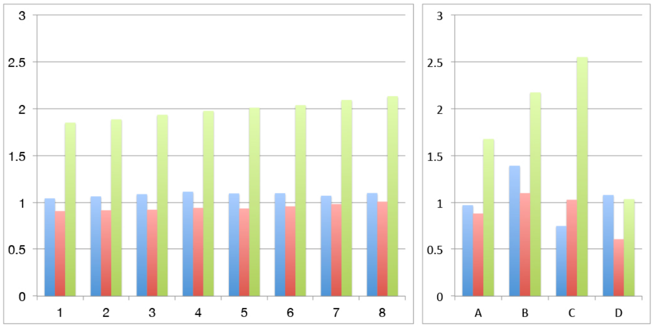

Figure 5. Error distance in monoscopic condition: Translation Task Precision per dimension (width, height, depth → blue, red, green). Left: target positions; Right: target rotations (A,B,C: simple; D: complex).

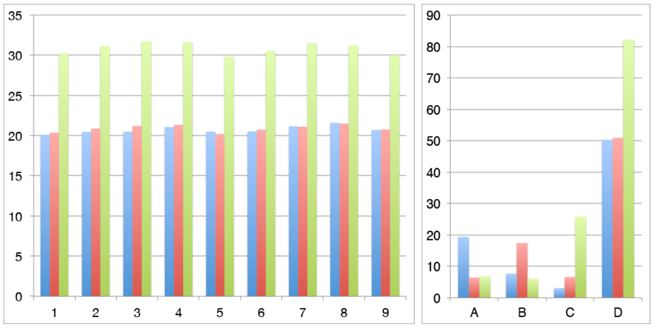

Figure 6. Error distance in monoscopic condition: Rotation Task Precision per dimension (yaw, pitch, roll → blue, red, green). Left: target positions; Right: target rotations (A,B,C: simple; D: complex).

Rotation Task Precision

For the rotation, based on related work a Quaternion metric was employed. Rotation precision was averagely higher in mono with significant differences for target positions and display mode, but without significant difference for target rotation. After inspecting the rotation precision per axis, the offset for the z-axis was better in stereo at all target positions and rotations. Rotations around the x-axis were performed better in mono with regard to the all target positions, but better in stereo condition regarding the rotations. Unfortunately, there were no significant differences for rotations around the y-axis. In summary, trials with simple one-dimensional rotations (A,B,C) were completed with evidently higher precision than the more complex three-dimensional (D) (see Figure 6 for monoscopic condition).

NASA TLX

Regarding the average overall workload for the sub-scales respectively, effort and physical demand dominated with the highest average values (5.81 and 5.88) in comparison to the other sub-scales. While mental demand and performance remained in the mid-field (5.73 and 5.00), the temporal demand had the lowest scaling average with 4.04. The average overall workload regarding the two display modes amounted 5.22 ($SD = 2.02$) for mono and 5.03 ($SD = 1.70$) for stereo. Altogether, there remained 5.09 ($SD = 1.79$) for both conditions in average. In conclusion, the trials performed in mono display mode resulted in minimal better performance with 5.17 than in stereo with 5.25. But regarding all other sub-scales, in particular mental demand (4.67 vs. 4.25) and effort (6.25 vs. 6.00), the trials performed worse in mono regarding NASA TLX results than in stereo. Attention should be paid to the lower physical demand in stereo with 5.67 in contrast to mono with 5.75.

Observations

Some of the participants mentioned that the object shadows were helpful in mono, especially in case of clipping. Other participants noticed the object shadows only after the half experiment was conducted. While translating the object, most of the participants thought they are using the accelerometer of the mobile device. A few participants have utilized a certain point, e.g. the home button, on the mobile device for a better orientation by mapping it to the front of the teapot. They mentioned, it would have been more intuitive for them. The better experimental results of those participants, in contrast to others, gave insights about how important the immersion and haptic factor is regarding 3D manipulation tasks. Another important observation during the study was a consistent ordering of object manipulation tasks performed by the participants with regard to their expertise level. Novices in the area of 3D modeling started generally with a coarse rotation followed by translating the object to the target location and finalized the trial with a finer rotation. The more expert users instead emphasized 3D trajectories while novices sticked with 1D trajectories.

DISCUSSION

Multidimensional input is often assumed to enable users to work quickly, but at the cost of precision due to the instability of the mid-air hand movement. Thus, the experiment was focused on precision rather than time. This section contains discussions about the experimental results and observations during the experiment. Those observations had a focus on how the participants went through the experiment, i.e. extract unusual or conspicuous behavior and sample trajectories.

In summary, the task completion time in average was better in stereoscopic display mode with regard to target position, and also regarding target rotation. Therefore we can accept hypothesis H2 to be confirmed. The experimental results indicated that translation and rotation task precision was significantly higher in the monoscopic than the stereoscopic condition, contrarily to other studies (e.g. [1]). The bad results of task precision for target positions in the lower left corner, lead to the assumption that the results were due to the partial occlusion by the non-dominant arm (all participants were right-handed). The occlusion of the object to be manipulated might also lead to perception issues in stereoscopic condition. With regard to target position, the translation precision was higher in monoscopic condition, and especially for target positions where the participants had to move the object downwards. The results also show that the translation task precision with regard to the rotation conditions was higher in the monoscopic condition. This partly confirms hypothesis H1. However, translation precision was higher in stereo for target positions in the upper back.

Considering rotations, the results were highly dependent on the complexity of the rotation. The difference between mono- and stereoscopic content considering 3D rotation task precision might be negotiable for applications that only deal with single-axis rotations. But this study shows that stereoscopic content notably outperforms monoscopic in complex situations that require simultaneous manipulation of all DOF. More precisely, when the participants had to rotate along more than one axis to achieve the task goal. The experimental results of rotation precision indicated that rotating around z-axis could be better performed in stereo than in mono. This could be concluded due to the physiological constraints of the human hand for rotating a physical object. Finally, the high standard deviations can be explained by the heterogeneity of the subjects and their skills regarding 3D modeling and stereo-vision.

The main benefit of using all six DOF without separation is that the participants were able to move and rotate the virtual object simultaneously. That was strongly influenced the temporal aspect of the interaction, but also the frustration level in the beginning phase of the task. This was proved by overall high task completion time and the evaluation of the NASA TLX results.

CONCLUSION

While the technology for input (smart-phones, Kinect, etc.) and output (stereoscopic displays) is market ready, few solutions for natural interaction with such devices exist. Nevertheless, there is still research needed. Therefore, the experiment described in this paper was conducted to investigate and evaluate natural interaction with stereoscopic 3D content. In this experiment, a mobile and gestural interaction technique for stereoscopic displayed 3D data was presented and evaluated in a docking task. The results gave insights into how affordable input and output devices can be used for complex 3D interaction with both, monoscopic and stereoscopic content. The experimental results showed that translation precision was higher in monoscopic condition. Furthermore stereoscopic content notably outperforms monoscopic in rotation around and translation along the z-axis.

While the bi-manual manipulation technique indicates to be efficient and easy to use for experts in 3D modeling, but more difficult to use for novice users. The overall conclusion is that monoscopy leads to more precise translation and rotation for 3D docking tasks using 6-DOF input devices, especially when simultaneous manipulations on all three axes are required. Future work, 3D object manipulation need to be addressed in more detail. In particular, the crucial problem of the evaluation of 3D rotation tasks should be investigated.

REFERENCES

1. James Boritz and Kellogg S Booth. 1998. A study of interactive 6 DOF docking in a computerised virtual environment. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*. IEEE, 139–146. DOI: http://dx.doi.org/10.1109/VRAIS.1998.658476
2. Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA. 3D interaction.
3. Michael Chen, S. Joy Mountford, and Abigail Sellen. 1988. A Study in Interactive 3-D Rotation Using 2-D Control Devices. *SIGGRAPH Comput. Graph.* 22, 4 (June 1988), 121–129. DOI: http://dx.doi.org/10.1145/378456.378497
4. Florian Daiber, Andrey Krekhov, Marco Speicher, Jens Krüger, and Antonio Krüger. 2013. A Framework for Prototyping and Evaluation of Sensor-based Mobile Interaction with Stereoscopic 3D. In *ACM ITS Workshop on Interactive Surfaces for Interaction with Stereoscopic 3D (ISIS3D). Workshop on Interactive Surfaces for Interaction with Stereoscopic 3D (ISIS3D-13), located at ITS 2013, October 6-9, St. Andrews, United Kingdom.* online, 13–16. http://www.dfki.de/~fdaiber/t3d/downloads/ISIS3D-proceedings/Daiber.pdf
5. Florian Daiber, Marco Speicher, Sven Gehring, Markus Löchtefeld, and Antonio Krüger. 2014. Interacting with 3D Content on Stereoscopic Displays. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. ACM, New York, NY, USA, Article 32, 6 pages. DOI: http://dx.doi.org/10.1145/2611009.2611024
6. Sven Gehring, Elias Hartz, Markus Löchtefeld, and Antonio Krüger. 2013. The Media FaçAde Toolkit: Prototyping and Simulating Interaction with Media FaçAdes. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 763–772. DOI: http://dx.doi.org/10.1145/2493432.2493471
7. Sven Gehring and Antonio Krüger. 2012. FaÇAde Map: Continuous Interaction with Media FaÇAdes Using Cartographic Map Projections. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 471–480. DOI: http://dx.doi.org/10.1145/2370216.2370286
8. Kim Halskov, Peter Dalsgaard, and Louise Bak Stolze. 2014. Analysing Engaging Experiences with a Tangible 3D Tabletop. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology (ACE '14)*. ACM, New York, NY, USA, Article 29, 10 pages. DOI: http://dx.doi.org/10.1145/2663806.2663831
9. S. G. Hart and L. E. Stavenland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati (Eds.). Elsevier, Chapter 7, 139–183. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000004342_1999205624.pdf
10. Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. ACM, New York, NY, USA, 452–458. DOI: http://dx.doi.org/10.1145/191666.191821
11. Ken Hinckley, Jeff Pierce, Mike Sinclair, and Eric Horvitz. 2000. Sensing techniques for mobile interaction. In *Proceedings of the 13th annual ACM symposium on User interface software and technology (UIST '00)*. ACM, New York, NY, USA, 91–100. DOI: http://dx.doi.org/10.1145/354401.354417
12. Ken Hinckley, Joe Tullio, Randy Pausch, Dennis Proffitt, and Neal Kassell. 1997. Usability Analysis of 3D Rotation Techniques. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology (UIST '97)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/263407.263408
13. Richard H Jacoby, Mark Ferneau, and Jim Humphries. 1994. Gestural interaction in a virtual environment. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 355–364. DOI: http://dx.doi.org/10.1117/12.173892
14. Marcel Köster, Michael Schmitz, and Sven Gehring. 2015. Gravity Games: A Framework for Interactive Space Physics on Media Facades. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*. ACM, New York, NY, USA, 115–121. DOI: http://dx.doi.org/10.1145/2757710.2757717
15. J.J. Kuffner. 2004. Effective sampling and distance metrics for 3D rigid body path planning. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, Vol. 4. 3993–3998 Vol.4. DOI: http://dx.doi.org/10.1109/ROBOT.2004.1308895
16. André Kunert, Alexander Kulik, Anke Huckauf, and Bernd Fröhlich. 2007. A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments. In *Proceedings of the 13th Eurographics conference on Virtual Environments (EGVE'07)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 43–52. DOI: http://dx.doi.org/10.2312/EGVE/IPT_EGVE2007/043-052
17. Hai-Ning Liang, Cary Williams, Myron Semegen, Wolfgang Stuerzlinger, and Pourang Irani. 2013. An Investigation of Suitable Interactions for 3D Manipulation of Distant Objects through a Mobile Device. *International Journal of Innovative Computing,*

Information and Control (IJICIC) 9, 12 (Dec. 2013), 4737–4752. http://www.ijicic.org/apchi12-291.pdf

18. Anthony Martinet, Géry Casiez, and Laurent Grisoni. 2010. The Effect of DOF Separation in 3D Manipulation Tasks with Multi-touch Displays. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology (VRST '10)*. ACM, New York, NY, USA, 111–118. DOI: http://dx.doi.org/10.1145/1889863.1889888
19. Maurice R. Masliah and Paul Milgram. 2000. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 25–32. DOI: http://dx.doi.org/10.1145/332040.332403
20. KasÖzacar, Kazuki Takashima, and Yoshifumi Kitamura. 2013. Direct 3D Object Manipulation on a Collaborative Stereoscopic Display. In *Proceedings of the 1st Symposium on Spatial User Interaction (SUI '13)*. ACM, New York, NY, USA, 69–72. DOI: http://dx.doi.org/10.1145/2491367.2491374
21. Krzysztof Pietroszek, Anastasia Kuzminykh, James R. Wallace, and Edward Lank. 2014. Smartcasting: A Discount 3D Interaction Technique for Public Displays. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 119–128. DOI: http://dx.doi.org/10.1145/2686612.2686629
22. Krzysztof Pietroszek, James R Wallace, and Edward Lank. 2015. Tiltcasting: 3D Interaction on Large Displays using a Mobile Device. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 57–62. DOI: http://dx.doi.org/10.1145/2807442.2807471
23. Frank Steinicke, Klaus H. Hinrichs, Johannes Schöning, and Antonio Krüger. 2008. Multi-Touching 3D Data: Towards Direct Interaction in Stereoscopic Display Environments coupled with Mobile Devices. In *Advanced Visual Interfaces (AVI) Workshop on Designing Multi-Touch Interaction Techniques for Coupled Public and Private Displays*. 46–49. http://viscg.uni-muenster.de/publications/2008/SHSK08
24. Yanqing Wang, Christine L. MacKenzie, Valerie A. Summers, and Kellogg S. Booth. 1998. The Structure of Object Transportation and Orientation in Human-computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 312–319. DOI: http://dx.doi.org/10.1145/274644.274688
25. Shumin Zhai and Paul Milgram. 1998. Quantifying Coordination in Multiple DOF Movement and Its Application to Evaluating 6 DOF Input Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 320–327. DOI: http://dx.doi.org/10.1145/274644.274689