When Spatial Devices are not an Option: Object Manipulation in Virtual Reality using 2D Input Devices

Martin Bellgardt, Niklas Krause, Torsten W. Kuhlen

Visual Computing Institute, RWTH Aachen University, Germany

Abstract: With the advent of low-cost virtual reality hardware, new applications arise in professional contexts. These applications have requirements that can differ from the usual premise when developing immersive systems. In this work, we explore the idea that spatial controllers might not be usable for practical reasons, even though they are the best interaction device for the task. Such a reason might be fatigue, as applications might be used over a long period of time. Additionally, some people might have even more difficulty lifting their hands, due to a disability. Hence, we attempt to measure how much the performance in a spatial interaction task decreases when using classical 2D interaction devices instead of a spatial controller. For this, we developed an interaction technique that uses 2D inputs and borrows principles from desktop interaction. We show that our interaction technique is slower to use than the state-of-the-art spatial interaction but is not much worse regarding precision and user preference.

Keywords: 3D-Interaction, Virtual Reality, Graphics Tablet, Mouse

1 Introduction

Systems that immerse users in a virtual reality (VR) have existed for decades. It is a crucial part of those systems to block stimuli from the outside world and replace them with computer generated stimuli to achieve a sense of immersion. In the recent years, immersive systems in the form of head mounted displays (HMDs) have become very popular, since due to renewed commercial interest, low-cost versions of such devices have become available. These systems, like the HTC Vive, Oculus Rift or Valve Index, to name just a few, are cheap enough to enable completely new applications while providing a degree of immersion that is comparable to previous state-of-the-art systems. However, for an immersive system to generate a virtual reality, interaction is equally important. Fortunately, most aforementioned systems come with spatial interaction devices that provide the means to design powerful interaction metaphors for nearly any application.

When it comes to interaction metaphors for VR, a very important concept to consider is presence. While presence and immersion are often used interchangeably, presence is more than just the stimuli that a user receives through their natural senses. Presence is the feeling of existing within the virtual world. Without the user being able to manipulate objects within the virtual world, and without those manipulations feeling natural, the user feels like a ghost floating through a world that is not aware of them. This would be an example of

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2020 in B. Weyers, C. Lürig, D. Zielasko (Hrsg.):
GI VR / AR Workshop 2020, 24. - 25, September 2020, Trier.
Copyright © 2020 bei den Autoren.
http://dx.doi.org/10.18420/vrar2020_9

immersion without presence. For that reason, most research on VR user interfaces is focused on designing interactions that feel natural[Kul09].

Conversely, some new applications that have been recently enabled by the aforementioned low-cost VR hardware call for a different design goal that often conflicts with that of presence. Especially for professional applications to be used for longer time periods in a working environment, comfort, i.e., the lessening of fatigue, should be an important design goal. Since most spatial interactions are to be performed in mid air, fatigue becomes non-negligible with prolonged usage of most systems. Additionally, many use-cases, especially when concerned with abstract data, have little to no benefit from presence[BPZ+17]. This highlights a demand for spatial interaction techniques that sacrifice presence for comfort, i.e., that feel less natural but are easier to use over a prolonged period of time.

To design an interaction technique for comfort, the first factor we considered is the usage scenario and the interaction devices used. The first key aspect to note is that the most reasonable degree of comfort is achieved when the user is sitting at a desk. This is the scenario in which most office workers spend most of their time. Second, the interaction devices that are generally used in desktop environments are designed in a way that allows the user to rest their arms on the desk. This holds true for the most commonly used interaction devices, i.e., keyboard and mouse but also for more specialized tools like graphics tablets. The classical six-degrees-of-freedom (6-DOF) devices that come with most consumer HMDs are not designed for this purpose. While the user, in theory, can rest their hands on the desk while holding these controllers, they would need to lift them in order to fully utilize all six degrees of freedom. For this reason, we conclude that, when designing for comfort, alternatives to 6-DOF controllers should be considered. Moreover, such alternatives could be beneficial to people suffering from disabilities that limit their ability to lift their arms.

In this work, we explore the applicability of this proposition by showcasing an implementation of a standard spatial interaction task without 6-DOF controllers. We wanted to use a task that is both common in professional applications and very convenient to achieve with 6-DOF devices, to act as a worst case scenario. Consequently, we chose object translation and rotation, as our interaction task. This task appealed to us, since it is essentially about manipulating virtual objects along six degrees of freedom that directly correspond to the 6-DOF controller, making it predestined for this task. In the following, we attempt to show that this task, even though deliberately chosen in favor of 6-DOF devices, can be achieved with different devices without sacrificing too much usability.

This brings us to the main contributions of this work. First, we present an interaction technique that uses a 2D input device, i.e., a mouse or a graphics tablet, to perform translation and rotation of virtual objects in an immersive environment. Second, we present the results of a user study that compared our interaction technique to the state-of-the-art technique using a 6-DOF device. It has to be stressed that this comparison does not aim to improve on the state-of-the-art, but to measure the sacrifices that have to be made when using the more convenient 2D devices.

2 3D object manipulation using 2D devices

Interaction techniques in 3D have been widely studied. For example, [JH13] provide an overview of many interaction techniques for interactive 3D environments. Although these techniques are not focused on immersive environments, they provide a good basis for this work, as many of them use 2D input devices. The main challenge to overcome when using 2D devices to control 3D environments is mapping the limited number of degrees of freedom to the many degrees of freedom required in 3D. One way to do this is mentioned in [CRI09], where different modes are used to switch between, e.g., translation and rotation via an icon based interface. However, as stated by [ISS02], the use of 2D interfaces for mode selection is not optimal. The user's attention shifts away from the object they want to interact with and towards the menu which could be in a completely different place. Additionally, the mode selection requires an additional mouse click (or a comparable interaction for different devices). This is why the authors propose a different method of interaction. Manipulation handles are visual representations of the manipulations that can be applied to an object. This way, the user can directly select the operation they want to perform and apply it to the object with only one action. For this reason, we see manipulation handles as the best solution to the degrees of freedom problem and utilize them in this work, as seen in Figure 1.

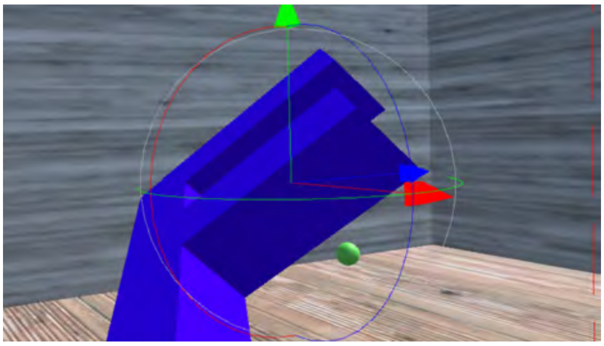

Figure 1: The manipulation handles we use in this work. They represent rotation and translation operations.

In this work we want to showcase interactions in an immersive 3D environment. This poses an additional challenge, which does not exist when 3D environments are displayed on a 2D screen. The manipulation handles, mentioned above, rely on a way for the user to precisely select them. In the 2D case, this is usually done with a mouse cursor, utilizing the projection of the 3D scene to precisely determine the object the user is pointing to. Since immersive systems use two projections, one for each eye, this technique is not applicable as the selected object would not be well defined. Instead, we decided to use a ray casting technique, as is very common in immersive environments. Moreover, we wanted the ray to be defined in a way that gives it similar behavior to a mouse cursor.

[PFC$^+$97] introduce a technique called occlusion selection. They define an interaction ray that starts from the user's eye and shoots through the user's hand. The name stems from the fact that this technique can only select non-occluded objects, as all potential interaction rays are identical to visual rays. Unfortunately, this technique still requires a spatial interaction device. To circumvent this, one might consider to use eye tracking, as proposed by [TJ00]. In this case, the ray's direction is defined by the center of the user's vision. While this technique completely frees up the user's hands, it requires a lot of focus and the user cannot look around anymore. Additionally, most eye tracking hardware is lacking precision. Instead, we chose to implement a technique that is very similar to the one proposed in [AA09]. While they also use the user's eye position as the ray's origin, they determine its direction via the rotation of the user's wrist. The only difference is, that we do not use the user's wrist rotation but the 2D position controlled by our 2D devices.

Unsurprisingly, our implementation also uses a point close to the user's eyes as the origin for our interaction ray. We considered the fact that, according to [MOB03], each person usually has an eye which is dominant, i.e., which is chosen over the other if conflicting signals are received. However, for the sake of simplicity, we chose the ray's origin to be the point between the user's eyes. To find the direction of the ray, we map the screen coordinates, the 2D devices operate in, to an invisible quad which is always in front of the user's face. With the origin point and the resulting point on this quad, the interaction ray is well defined.

We experimented with different graphical representations of the ray. First, we attempted to simply show the line, as is common with 6-DOF controllers. Unfortunately, seeing a ray that emanates from between one's eyes is very irritating, as informal testing showed. We settled on displaying only the impact point of the interaction ray with a small sphere. To make the sphere easily visible, we assigned it the color green, as this color was not used anywhere else in the application. In a more general application, one might need to find more sophisticated ways of highlighting it.

3 Experimental Setup

To test our interaction technique, we formulate the following hypotheses:

- H1: Manipulations with the 6-DOF controller are faster than with the mouse or the graphics tablet.
- H2: The 6-DOF controller requires less individual steps to perform manipulations.
- H3: The graphics tablet, mouse and 6-DOF controller achieve similar precision.
- H4: The percieved usablity of the 2D input devices is similar to that of the 6-DOF device.

H1 and H2 are based on the fact that the 6-DOF controller is able to manipulate all degrees of freedom at once, while the 2D devices can only manipulate one at a time. As shown by H3 and H4, the points where we expect the 2D device to achieve comparable performance are precision and subjective preference. Note that we do not include a hypothesis regarding fatigue over prolonged usage. Unfortunately, properly testing such a hypothesis would require a much more sophisticated study than was possible within the scope of this work.

For our experiment, we used an Oculus Rift HMD and an Oculus Touch Controller as our 6-DOF base-line device. The other input devices tested were a standard PC mouse and a Wacom Intuos 3 graphics tablet. The participants were seated at a regular office desk that contained all the devices listed above, ready to use. After filling out an informed consent form, the participants received a written explanation of the task, the HMD and the input devices. The participants were then prompted to wear the HMD and begin with the practice scene. During practice, the participants were given up to 10 minutes of time to try out all the interaction devices in a setting that was very similar to the actual tasks. Using a menu, which was only available in this practice scene, they were able to freely choose the device they wanted to practice. The participants were also free to ask any questions during practice and, at the end, were prompted again if any questions arose.

Next, the participants performed the main tasks while their performance was being measured. There were 3 distinct scenes to complete with all of the 3 interaction devices, resulting in 9 tasks total. The order of the tasks, i.e, the order in which the devices were used and the order of the scenes for each device, was randomized. In each task, the users were supposed to fit a three dimensional letter into a predefined position (see Figure 2), a task setup which is very similar to the one used in [CRI09]. The target position was displayed to the user as an outline of the same object, making even slight deviations easily perceivable. To get the object into the defined outline, it had to be both translated and rotated. The users were allowed to look at the scene as long as they wanted. When they were ready to start manipulating the object, they pushed a button to start the timer. When the users were contempt with the placement of the object, they pressed another button to stop the timer and to proceed to the next scene. We measured the time the users took to test H1. We also measured the number of operations used to test H2, where an operation is defined as one instance of pressing the mouse button, touching the surface of the graphics tablet with the pen or pulling the trigger on the 6-DOF controller. Finally, we measured the distance between the object's center and the target's center, as well as the angular difference to test H3.

After all tasks were completed, the participants removed the HMD and were presented with a questionnaire. The questionnaire contained the questions from the standardized SUS questionnaire[BKM08], as well as a few additional questions about their subjective preference, to test H4. The free-text part of the questionnaire was asked in an interview and noted down by the experimenter.

4 Results

The study was conducted with 30 participants of which 20 identified as male and 10 identified as female. 16 participants were between the age of 18 and 24, the rest were older than 25

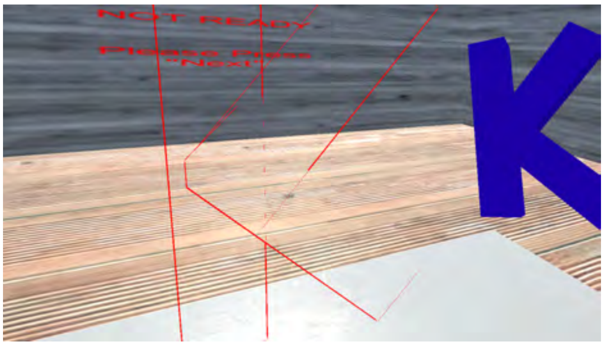

Figure 2: During the tasks, the participants had to fit a giant letter, in this case a K, into the predefined shape, outlined in red.

Table 1: The preferences the participants indicated in the questionnaire regarding three different aspects.

	6-DOF	Graphics Tablet	Mouse
precise	4	18	8
fast	25	0	5
comfortable	27	2	1

years. All participants reported to use a smartphone regularly but only one participant used VR Devices on a regular basis.

The results of our quantitative evaluation can be seen in Figure 3. When using the 6-DOF devices the participants took on average 40.9 seconds to position the object, as compared to 60.8 seconds with the mouse and 65.1 seconds with the tablet. The lowest average number of operations used was recorded for the mouse with 21.4 operations, compared to 24.6 and 26.8 for 6-DOF controller and Tablet, respectively. The average error, both regarding distance and angular difference were very similar for all three devices. For each of our four quantitative measurements, we performed a one-way Welch-ANOVA to check if the means for the three devices are different. We found that the means for the time taken were significantly different with $p < 0.0005$ and the means for the number of operations were significantly different with $p < 0.05$.

The tablet received the lowest average SUS score of 75.3. The mouse received an average SUS score of 80.3, while the highest average SUS score of 82.0 was assigned to the 6-DOF controller. Using a Welch-ANOVA, we found those differences to be not significant. In addition to the SUS, we asked the participants for their specific preference regarding precision, speed and comfort. The answers to this question are summarized in Table 1.

During the qualitative feedback part of the study, 19 participants complimented the precision of the graphics tablet. Conversely, 13 participants found the use of the graphics tablet to be unintuitive. It was commonly mentioned that the graphics tablet needed additional practice because it was not visible under the HMD. Analogously, there were 14 positive comments about the precision of the mouse. 7 of those participants stated that the mouse was precise but not as precise as the tablet. However, 6 users stated that the mouse was not precise, as the ability to adjust the sensitivity of the mouse was missing. Additionally, it was commented 14 times that the usage of the mouse was intuitive and/or easy to use.

Feedback about the 2D devices in general included that they were difficult to use because the interaction depended on head movement. 9 users liked the fact that they could manipulate the degrees of freedom individually while 8 disliked it. Overall, 12 times the usage of the 2D devices was described as inefficient. On the other hand, the 6-DOF controller was generally seen as very fast and 24 users described it as intuitive and/or easy to use.

5 Discussion

Unsurprisingly, the users performed the tasks significantly faster when using the 6-DOF controller. Hence, we can accept H1. Surprisingly, the results were not as clear for the number of operations used. While the means were shown to differ significantly between the three devices, the device with the lowest number of operations was the mouse, so we need to reject H2. The most likely explanation is that we underestimated the number of corrections that the participants performed to achieve a precise result.

The participants generally achieved a high degree of precision when aligning the object, as the distance error was below 2 cm and the angular difference below 2 degrees most of the time. Unfortunately, although the Welch-ANOVA we performed showed no significant difference between the means, this result cannot be interpreted as evidence for their equivalence. However, we found that the 0.95 confidence intervals of all means intersect for both distance and angular difference. This means that there are values which could represent the true mean with a confidence of 0.95. For that reason, we assume the means to be equivalent and accept H3.

To check for H4, we first considered the SUS scores of the three techniques. Even though the score of the tablet is the lowest, we found no significant difference between the means and all 0.95 confidence intervals to intersect. Based on the SUS, we could accept H4, but the preferences presented in Table 1 show a different situation. When forced to make a choice, nearly all participants prefer the 6-DOF controller with regard to speed and comfort. Since the participants were objectively faster with the 6-DOF controller, the preference for speed is unsurprising. As far as comfort is concerned, the participants might have focused on the intuitiveness and effort to achieve the task instead of fatigue. As mentioned earlier, fatigue could not be measured in this study, since the time the participants used the technique was not long enough. Surprisingly, most participants preferred the graphics tablet with regard to precision, even though we could not find evidence of the precision of the graphics tablet being higher. This can also be seen from the free text feedback, where 19 participants perceived the grapics tablet as precise. Regarding the other feedback from the participants, we do not see a general rejection of the idea of using the 2D devices, although there was some negative feedback. Although these results are inconclusive, we choose to accept H4 under reservations.

6 Conclusion and Future Work

We have presented an interaction technique for 3D object manipulation in immersive environments using classical 2D devices. We chose 3D object manipulation as a worst case scenario, since it is a task that 6-DOF devices are almost ideal for. Our results showed that, using 2D devices, participants were able to perform the task of object manipulation with equal precision as with a 6-DOF controller but at a reduced speed. Generally, the performance of the 2D devices was lower than that of the 6-DOF controller, as was expected. However, the 2D devices did not under-perform strongly and we have shown that they might be a viable alternative in situations where 6-DOF controllers are not applicable. Such applications might include situations where the system is used for a long period, so the usage of 6-DOF devices induces too much fatigue, or by individuals with disabilities that limit their ability to lift their arms. Naturally, it has yet to be shown that 2D devices outperform 6-DOF devices in terms of fatigue, which could be a subject of future studies. Additionally, one could measure the effect on immersion, which we believe to be severe, since the interaction space is heavily limited in our scenario.

Additionally, we have shown that users perceived the graphics tablet to be very precise, even though this precision could not be measured. This hints at potential applications where users might even prefer the graphics tablet over a 6-DOF controller, if precision is their main concern. To improve on this observation, the visual representation of the graphics tablet could be refined to make it more usable in immersive environments. One potential approach might be to render the graphics tablet and pen at its precise location in the real world.

Other devices might be considered as well. Hand tracking has gained traction in the recent years and, although we consider it spatial interaction, it might be interesting to see how it compares to our interaction technique. A hand tracking system might even be used to simulate a 2D device on a flat surface, without the need for additional hardware but with all benefits regarding fatigue.

Finally, our study application was very artificial. Showcasing the usage of 2D interaction devices in a more realistic application might provide further valuable insights. Looking at an immersive scenario where a 2D interaction device is beneficial, i.e., where the task only requires two degrees of freedom, might be another possible direction.

7 Acknowledgements

The authors would like to thank the German Research Foundation DFG for the kind support within the Cluster of Excellence "Internet of Production" - Project-ID: 390621612.

References

[AA09] F. Argelaguet and C. Andujar. Efficient 3d pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, 29(6):34–43, Nov 2009.

[BKM08] Aaron Bangor, Philip T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594, 2008.

[BPZ+17] Martin Bellgardt, Sebastian Pick, Daniel Zielasko, Tom Vierjahn, Benjamin Weyers, and Torsten W. Kuhlen. Utilizing Immersive Virtual Reality in Everyday Work. In *IEEE Workshop on Everyday Virtual Reality, WEVR*, 2017.

[CRI09] Luca Chittaro, Roberto Ranon, and Lucio Ieronutti. 3d object arrangement for novice users: The effectiveness of combining a first-person and a map view. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, pages 171–178, New York, NY, USA, 2009. ACM.

[ISS02] P: Issacs, J. Shrag, and P. Strauss. The design and implementation of direct manipulation in 3d. *SIGGRAPH 2002 Course Notes*, 01 2002.

[JH13] Jacek Jankowski and Martin Hachet. A Survey of Interaction Techniques for Interactive 3D Environments. In M. Sbert and L. Szirmay-Kalos, editors, *Eurographics 2013 - State of the Art Reports*. The Eurographics Association, 2013.

[Kul09] A. Kulik. Building on realism and magic for designing 3d interaction techniques. *IEEE Computer Graphics and Applications*, 29(6):22–33, Nov 2009.

[MOB03] Alistair P. Mapp, Hiroshi Ono, and Raphael Barbeito. What does the dominant eye dominate? a brief and somewhat contentious review. *Perception & psychophysics*, 65:310–7, 03 2003.

[PFC+97] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, I3D '97, pages 39–ff., New York, NY, USA, 1997. ACM.

[TJ00] Vildan Tanriverdi and Robert J. K. Jacob. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 265–272, New York, NY, USA, 2000. ACM.

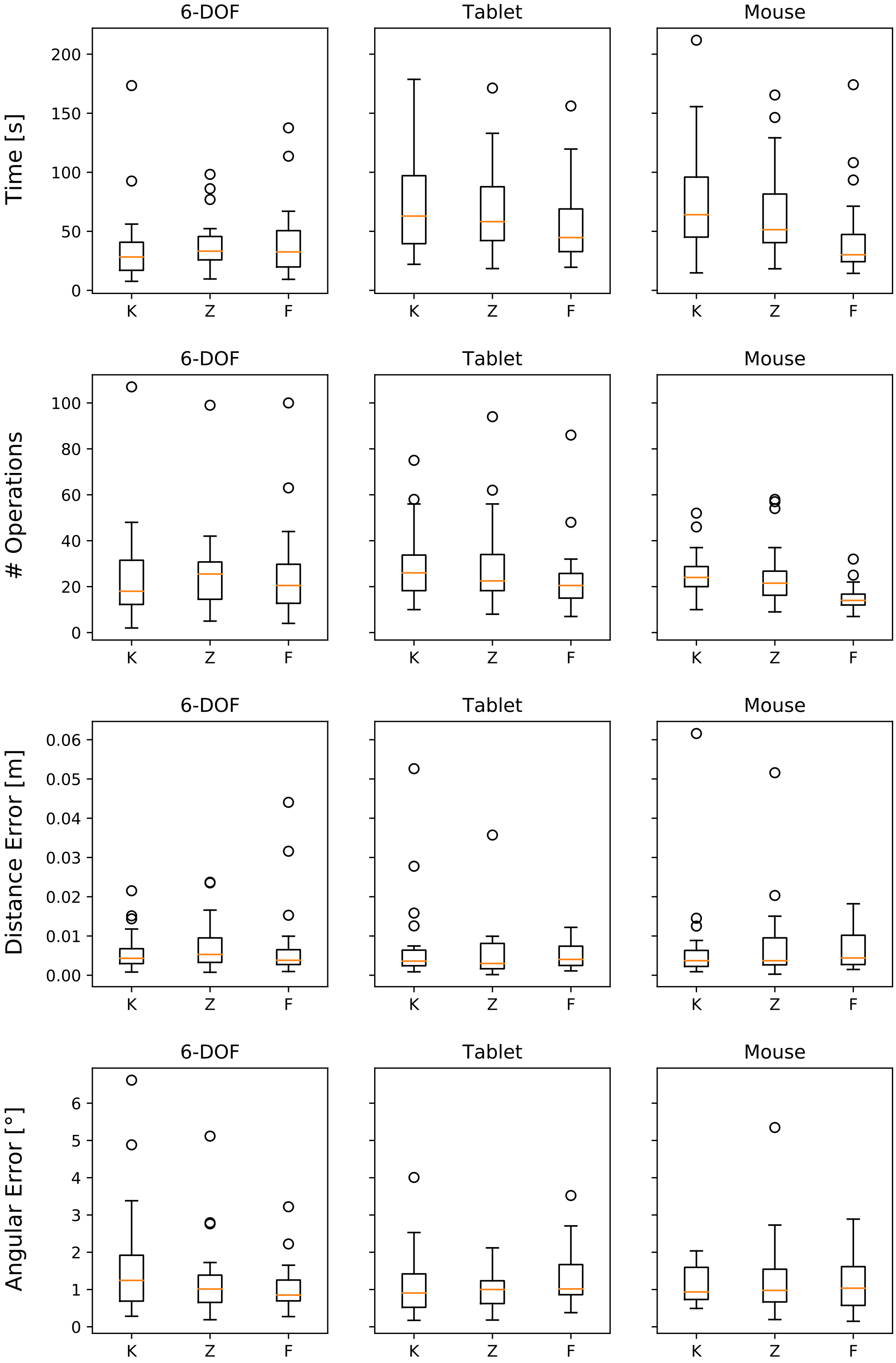

Figure 3: The results of our 4 quantitative measures for all 3 scenes (K, Z, F) and all 3 conditions. 6 outlier values were removed for readability.