The Benefits of DOF Separation in Mid-air 3D Object Manipulation

Daniel Mendes*, Filipe Relvas, Alfredo Ferreira, Joaquim Jorge
INESC-ID Lisboa, Técnico Lisboa, Universidade de Lisboa

Abstract

Object manipulation is a key feature in almost every virtual environment. However, it is difficult to accurately place an object in immersive virtual environments using mid-air gestures that mimic interactions in the physical world, although being a direct and natural approach. Previous research studied mouse and touch based interfaces concluding that separation of degrees-of-freedom (DOF) led to improved results. In this paper, we present the first user evaluation to assess the impact of explicit 6 DOF separation in mid-air manipulation tasks. We implemented a technique based on familiar virtual widgets that allow single DOF control, and compared it against a direct approach and PRISM, which dynamically adjusts the ratio between hand and object motions. Our results suggest that full DOF separation benefits precision in spatial manipulations, at the cost of additional time for complex tasks. From our results we draw guidelines for 3D object manipulation in mid-air.

Keywords: 3D user interfaces, immersive virtual environments, spatial interactions, mid-air object manipulation, DOF separation

Concepts: •**Human-centered computing** → **Gestural input; Empirical studies in HCI;**

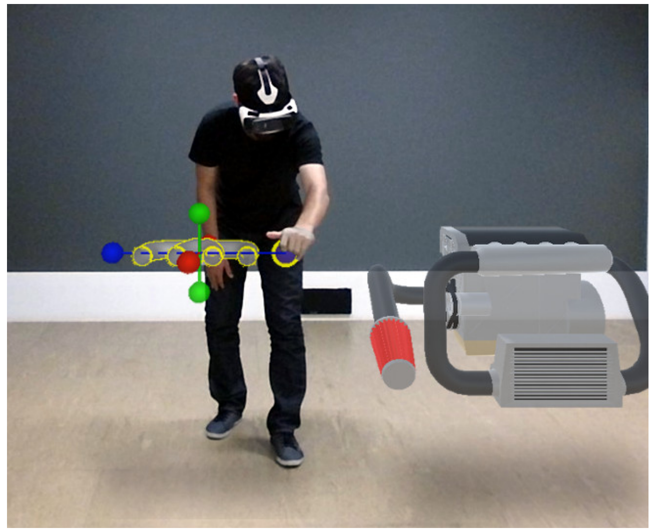


Figure 1: *User moving an object with virtual handles in mid-air.*

1 Introduction

We are witnessing a huge interest in virtual reality (VR), mainly due to the recent technological advances that made head-mounted displays (HMD) affordable and widely available. Immersive virtual environments (IVE) that were made possible with such technologies are being used for several purposes, like engineering, architecture, game development and so forth, offering unique capabilities. To interact within those virtual environments (VE), the ability to manipulate virtual objects is a key feature. While direct approaches that mimic interactions in the physical world are the most natural, it is still difficult to place a virtual object in the desired place with a high degree of accuracy. These difficulties may arise from different factors, such as limited human dexterity for mid-air gestures and lack of precision from tracking systems.

Given its importance, object manipulation in virtual environments has been subject of research for long, covering different kinds of interaction paradigms. For both mouse and touch interfaces, separation of degrees-of-freedom (DOF) led to better users' performance when compared to direct approaches, mainly due to the required mapping between the 2D input and 3D output. While in mid-air interactions this dimensional difference between input and output does not exist, new techniques need to be developed in order to compensate for the lack of precision inherent from mid-air gestures. Indeed, previous work showed that even in mid-air, users tend decompose arbitray rotations into sequences of single DOF tasks [Veit et al. 2009]. Our work shows that full and explicit DOF separation may also be useful for purely mid-air object manipulation within IVEs, since it may prevent unintended and unexpected transformations.

In this paper we present a study to assess the impact of DOF separation in mid-air object manipulation. For this purpose, we conducted an evaluation comparing three manipulation techniques based on existing literature. One follows a direct approach, the second scales users' movement and the third is our implementation of mid-air virtual handles for DOF separation, as depicted in Figure 1. In the remainder of the document, we survey the most relevant related work, then we present our user evaluation, detailing the implemented techniques, interactive setup based on non-intrusive positional tracking and results. From the attained results, we draw guidelines for future object manipulation techniques in mid-air. Finally, we conclude our work and point out directions for future research.

2 Related Work

Techniques for manipulating virtual objects have been subject of research in the past few decades. From mouse to mid-air, passing through touch enabled surfaces, several approaches have been proposed, ever trying to more effectively position and orient objects in virtual environments.

2.1 Traditional Mouse-based Manipulations

To overcome the difference between input and output DOF, most mouse-based techniques for virtual object manipulation rely on some sort of widgets, which reduce the simultaneous DOF being

*email: danielmendes@tecnico.ulisboa.pt

VRST '16,, November 02 - 04, 2016, Garching bei Munchen, Germany
ISBN: 978-1-4503-4491-3/16/11.
DOI: http://dx.doi.org/10.1145/2993369.2993396

controlled. Houde [Houde 1992] proposed the handle box, a bounding box surrounding the object being manipulated, with a lifting handle to move the object up and down, and four rotation handles, to rotate the object about its central axis. Conner et al. [Conner et al. 1992] also resorted to virtual handles to develop 3D widgets for performing 9DOF transformations on virtual objects. The handles are used to constrain geometric transformations to a single plane or axis. Focusing only in rotations, Ken Shoemake proposed Arcball [Shoemake 1992], where users can draw an arc on the screen projection of a sphere to change object's orientation. These techniques attained such popularity, they are still used in common commercial applications for creating or editing 3D virtual models, like Unity3D[1] or SketchUp[2].

2.2 Interacting with Multi-touch Surfaces

Distinctly from mouse techniques, touch enabled surfaces allow users to directly touch objects displayed. Trying to create more natural interactions since it has been shown that rotation and translation have a parallel and interdependent structure in the human mind [Wang et al. 1998], researches initially proposed techniques for controlling several DOF at the same time [Hancock et al. 2009; Reisman et al. 2009]. Nonetheless, reduction of simultaneous DOF controlled have been later suggested [Martinet et al. 2010] and followed by several authors. Thus, techniques that allow manipulations with high DOF, but with no more than one controlled at each moment, have been later proposed. Eden [Kin et al. 2011] and LTouchIt [Mendes et al. 2011] resorted to a direct drag approach to translate objects, while separating horizontal and vertical movements, and perform rotations through virtual widgets. tBox [Cohé et al. 2011], a 3D transformation widget that appears as a wireframe bounding box, favors independent 9DOF control. Similarly, Gimbal Box [Bollensdorff et al. 2012] also uses a cube shaped widget to separate DOF manipulation, and its authors concluded that adapted widgets are superior to other approaches for multi-touch interactions. Semi-immersive stereoscopic tabletops present different challenges, since imagery appears on a volumetric space. Benko and Feiner [Benko and Feiner 2007] decomposed 3DOF tasks into a set of 2DOF and 1DOF tasks, using a balloon metaphor. Triangle cursor [Strothoff et al. 2011] follows a similar approach but uses two touches to define a triangle's base, with a cursor in the top vertex. To manipulate virtual objects in full 9DOF, Toucheo [Hachet et al. 2011] relies on widgets on the multi-touch surface.

2.3 Mid-air Manipulations

Having an input with higher DOF, most current mid-air approaches for 3D virtual object manipulation try to mimic physical world interactions [Hilliges et al. 2009; Wang et al. 2011; Song et al. 2012; Araújo et al. 2013]. However, fine-grained manipulation and precision tasks are hard to perform with these techniques, due to limited human accuracy, which is sometimes aggravated by input devices' resolution.

To overcome the lack of precision with object positioning techniques, Kiyokawa et al. [Kiyokawa et al. 1997] proposed manipulation aid consisting of discrete placement constraints (snapping) and collision avoidance mechanisms. Without imposing placement restrictions, Frees et al. [Frees and Kessler 2005] proposed PRISM (Precise and Rapid Interaction through Scaled Manipulation). This technique scales the hand movement down to increase precision. Switching between precise and direct mode occurs according to the current velocity of the user's hand. When moving an object from one general place to another, the user is not necessarily interested in being precise and moves relatively rapidly. When users are focused on accurately moving an object to very specific locations, they normally slow their hand movements down and focus more on being precise. PRISM increases the control/display ratio, which causes the cursor or object to move more slowly than the user's hand, reducing the effect of hand instability and creating an offset between the object and the hand. User evaluation's results show faster performance and higher user preference for PRISM over a traditional direct approach for translation tasks. The authors later extended the previous work, by adding support in PRISM for object rotation, which uses the angular speed of the hand [Frees et al. 2007]. Although extending transformations to additional 3 DOF, authors concluded that this approach for rotations is confusing to users.

Also focusing on precise positioning of 3D virtual objects in IVEs, Osawa [Osawa 2008] proposed a position adjustment that consists in a scale factor for slowing hand movement, similar to PRISM, and viewpoint adjustment, that automatically approaches the viewpoint to the grabbed point so that the object being manipulated appears larger. Through a user evaluation, these techniques showed improvements for small targets. Veit et al. showed that interactions that ease task's decomposition can lead to significant improvements in performance for orientation tasks [Veit et al. 2009]. However, their approach for DOF separation was based upon a planar surface for restricting hand's movements, which might not be always feasible for exclusively mid-air scenarios. The 7-Handle manipulation technique [Nguyen et al. 2014] consists of a triangle shaped widget with seven points. User evaluation results showed that 7-Handle is only better suited than the traditional direct 6 DOF approach for manipulating large objects.

As presented, most mouse based interfaces rely on widgets for object manipulation, focusing on reducing the simultaneous DOFs being controlled. For touch enabled surfaces, albeit allowing users to directly interact objects displayed, researchers found out that DOF separation led to better performance on object manipulation tasks, having turned once again to virtual widgets to clearly and undoubtedly select transformations and axes. Even when interacting with stereoscopic imagery above tabletops, the only technique that allow full 9 DOF manipulations resorts to widgets. To improve users' accuracy in mid-air interactions, researchers already tried to either scale down hand motions or move the viewpoint closer to object being manipulated, but without regard to DOF separation. On the other hand, an approach based on virtual widgets has already been proposed. However, this technique does not have promising results, possibly because widgets it used are very different from those used in mouse and touch based interfaces, being more complex, not allowing controlling a single DOF at a time and not using common reference frames, such as object or world axes. Clear DOF separation in mid-air scenarios, using familiar virtual widgets, might improve users' performance in object manipulation tasks.

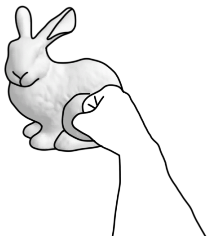
3 Evaluation of DOF Separation in Mid-air

Since DOF separation showed positive results in mouse and touch interaction for virtual 3D object manipulation, we conducted a user evaluation to assess if it also benefits spatial interactions in IVEs.

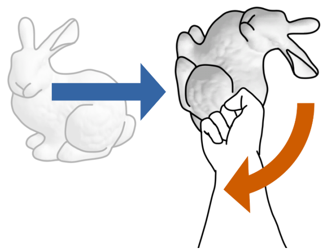
3.1 Object Manipulation Techniques Implemented

For our evaluation, we implemented three techniques based on the literature. The first is a direct approach in which all transformations performed by the user hand are directly applied to the object, the second follows scaled transformations based on the user movement's speed, and the third consists of spatial widgets for separating DOF. All implemented techniques provide 6 DOF transformations: three for translation and three for rotation.

[1]Unity3D: http://unity3d.com, last visited June 30th 2016.
[2]SketchUp: http://sketchup.com, last visited June 30th 2016.



(a) *Initial State.*



(b) *Simultaneous translation and 90 degrees rotation.*

Figure 2: *6DOF Technique.*

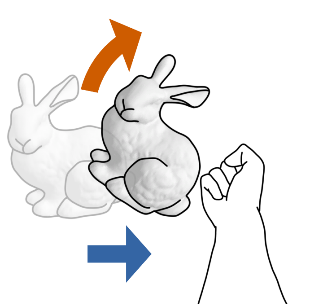
3.1.1 6DOF or Direct Manipulation

To mimic interactions with physical objects as closely as possible, direct manipulation uses all 6 DOF information from users' hands [Wang et al. 2011]. It is often used as a baseline for evaluations of other techniques [Frees et al. 2007; Mendes et al. 2014; Nguyen et al. 2014]. This technique consists of grabbing an object directly, moving it to a new location and/or rotating it, and then releasing. After being grabbed, the object directly follows the movement of the hand: dragging changes object's position and wrist's rotation controls object's rotation. All transformations are simultaneously applied to the object, as pictured in Figure 2. The grabbed point in the object will remain the center of all transformations during the entire manipulation, until the object is released.

3.1.2 PRISM

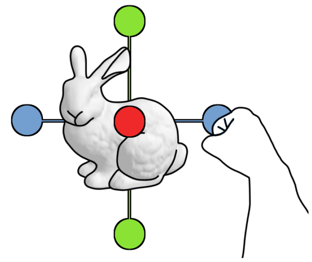
We implemented the PRISM technique as presented by Frees et al. [Frees and Kessler 2005]. This technique aims in improving accuracy of direct manipulation, switching between a precise and a direct mode according to the current velocity of users' hands. Hand's movement in each coordinate axis is scaled down when users move their hands slower than a pre-defined threshold in that axis. We used the threshold value proposed by the original authors. This scaling results in an offset between the hand and the object being manipulated, that can be canceled by moving hands faster than the same threshold. We also included rotations later proposed by same authors [Frees et al. 2007], which follows the same premise from translations, scaling down slow wrist rotations. As suggested by the authors, resulting offsets are represented by a white line for translations, and two sets of axis for rotations, as shown in Figure 5. Similarly to 6DOF technique, both translations and rotations can be performed simultaneously, as exemplified in Figure 3.

(a) *Initial State.*

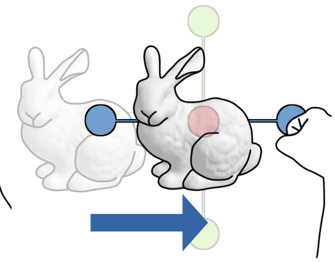


(b) *Simultaneous scaled translation and rotation, with accumulated offset.*

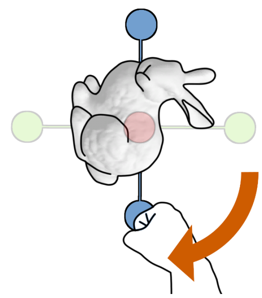
Figure 3: *PRISM Technique.*



(a) *Initial State.*



(b) *Translation.*



(c) *Rotation.*

Figure 4: *Widgets Technique.*

3.1.3 Widgets for DOF Separation

Widget based manipulations are widely used in mouse and keyboard 3D user interfaces. Our implementation, as opposed to those described in sections 3.1.1 and 3.1.2, strictly follows DOF separation. Not only translation and rotation operations are treated independently, users can only manipulate 1 DOF at a time. We used a representation similar to that introduced by Conner et al. [Conner et al. 1992], illustrated in Figure 4. Users can grab the sphere connected to the desired axis and move the hand along the axis to trigger object translation. For rotations, the approach is similar, but the hand movement is performed around the target axis. The decision to either perform a translation or rotation, is made based on the hand's path after 10 cm. Selected transformation and axis remain locked until a release gesture.

3.2 Methodology

All user sessions followed the same structure, each lasting approximately 45 minutes. We started by introducing the experiment the participant was about to perform, followed by a brief description of the techniques being evaluated. The techniques were performed in alternated order, assuring that each one was experienced in every possible permutation, in order to avoid biased results.

For each technique we played a video showing how to apply transformations to the object with it. After the video, participants had a training period of three minutes, or less if they considered themselves to be already acquainted, to explore the approach in a dedicated environment, showed in Figure 5. Following the practice period, we asked participants to perform six tasks, described in the next section. After completing each technique's tasks, participants fulfilled a questionnaire regarding distinct aspects of the interaction. The experiment concluded with a profiling questionnaire.

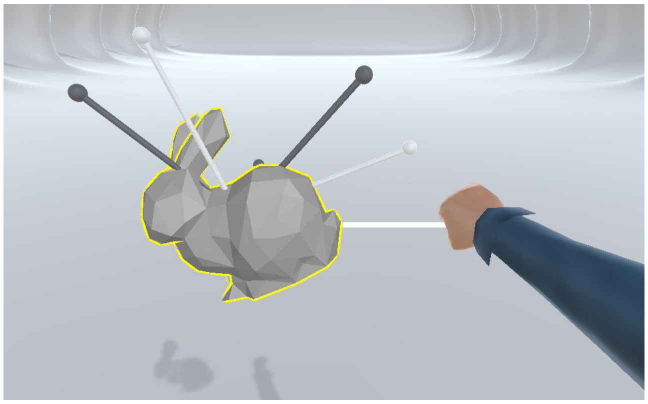
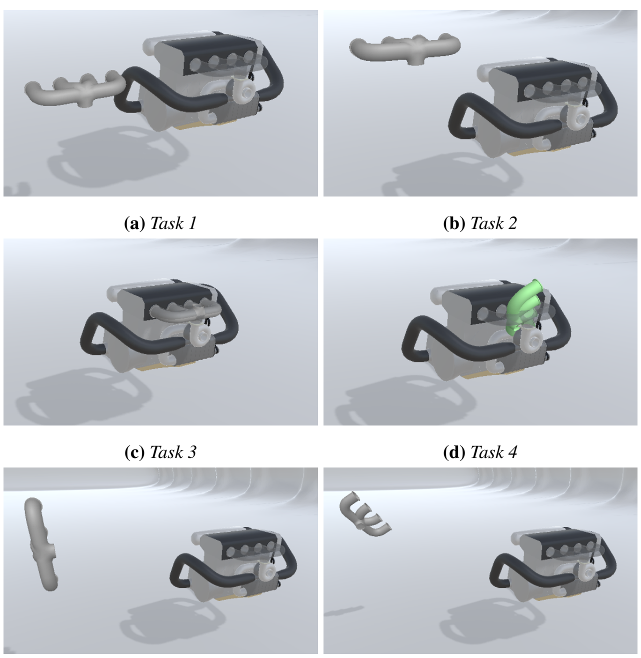


Figure 5: *Interacting with an object in our virtual environment during the training period, with PRISM technique.*

3.3 Tasks

As we mentioned in the previous section, we requested participants to complete a set of six tasks for each technique. All consisted in a docking task [Frees et al. 2007; Martinet et al. 2010; Mendes et al. 2014], where participants had to put the exhaust pipes in the right place of a car engine[3]. That component of the engine was the only object in our virtual environment that could be grabbed and transformed. Engine's model had a semi-transparent replica of the pipes showing the only possible target position and orientation, as depicted in Figure 6. To prevent excessively long sessions, each task was limited to a maximum of three minutes. After reaching time limit we informed participants they could stop, and we considered the attained position and orientation as final.



(a) *Task 1* **(b)** *Task 2* **(c)** *Task 3* **(d)** *Task 4* **(e)** *Task 5* **(f)** *Task 6*

Figure 6: *Tasks performed by the participants.*

[3]Original 3D model of the used engine uploaded to Sketchup's 3D Warehouse by user M-Speed.

For the first task (Figure 6a), the object to be manipulated begun with the correct position along both YY and ZZ scene axes and orientation, only with an incorrect position according to the X coordinate. Similarly to the previous task, the second task (Figure 6b) object started with the correct orientation, however its position was incorrect along all three coordinates. The third task (Figure 6c) consisted in only rotating the object around the Z axis, while the fourth task (Figure 6d) implied rotation around an arbitrary axis, requiring no translation as well. The fifth task (Figure 6e) required the object to be rotated around the Z axis and translated along both XX and YY axes. Finally, in the last task (Figure 6f), participants had to apply full 6 DOF transformations to the object. Although some tasks required only one kind of transformation (translation or rotation), none was restricted, as we did not intend to modify any technique in order to accommodate a specific task.

3.4 Setup and Prototype

Our setup comprises non-invasive and affordable full body user tracking with three depth cameras Microsoft Kinect v2. One of them was placed facing the user while the remaining ones lied on each side, 90 degrees from the first one. Since Microsoft Kinect fails in providing reliable hand orientation data, we developed an wireless custom made device to better acquire such data, pictured in Figure 7. It uses an IMUduino[4], an Arduino based circuit board, that incorporates an IMU and a Bluetooth LE modules. The IMU is composed of gyroscope, accelerometer and digital compass sensors for accurate 3 DOF orientation tracking. We attach the device to the user's dominant hand using an acrylic clip, which assures it does not fall when the hand is opened. A pressure pad detects if the hand is open or closed. For the visualization component, we used a Gear VR with a Samsung Galaxy S6, connected via Wi-Fi to our tracking server.

We developed our prototype using Unity3D engine, with gravity and objects' collision disabled. For improved user feedback while grasping, the object becomes transparent, revealing the penetrating portion of the hand, as suggested by previous research [Prachyabrued and Borst 2014]. To guide participants during evaluation tasks, we make the object gradually turn green as it approaches the target position and orientation, as it can be seen in Figure 6d.

Figure 7: *Our custom made device for tracking hand's rotation and its open / grab state.*

[4]IMUduino: http://femto.io/products/imuduino, last visited June 30th 2016.

3.5 Apparatus and Participants

The experiment was performed in our laboratory with a controlled environment (Figure 8), using the setup detailed in the previous section. We counted with the participation of 21 people (5 female), between the ages of 18 and 50 years old, with the great majority (62%) between 18 and 25. Most had at least a BSc degree (86%), while the remainder are finishing it. More than half (52%) had never experienced a VR setting, and 43% use some kind of gesture recognition systems more than once a month, such as XBox Kinect, Wii Remote or Playstation Move. Only 28% of participants use 3D modelling systems at least once a month.

3.6 Results and Discussion

During our experiment, we collected both objective data, through logging mechanisms, and subjective data, asking participants to fill out questionnaires. We used Shapiro-Wilk test to assess data normality. We then ran the repeated measures ANOVA test with a Greenhouse-Geisser correction to find significant differences in normal distributed data, and Friedman non-parametric test with Wilcoxon-Signed Ranks post-hoc test. In both cases, post-hoc tests used Bonferroni correction (corrected sig. = sig. $\times$ 3).

3.6.1 Objective Data

We measured time taken by participants to fulfil each task, as well as object placement error. Time taken for all tasks, in seconds, is depicted in the graph of Figure 9. Regarding errors, we registered both position error, in millimeters (Figure 10), and rotation error, in degrees (Figure 11).

For the translation only tasks, we found statistically significant differences in completion time (Task 1: $\chi^2(2)$=25.368, p<.0005; Task 2: F(1.611,30.604)=9.025, p=.002). For the first task, post-hoc test revealed Widgets approach (avg=25s) to be faster than both 6DOF (avg=59s, Z=-3.542, p<.0005) and PRISM (avg=90s, Z=-3.823, p<.0005), and 6DOF to be faster than PRISM (Z=-3.267, p=.003). In the second task, PRISM (avg=102s) was significantly slower than Widgets (avg=49s, p=.008) and 6DOF (avg=71s, p=.028). For position error, differences were also found (Task 1: F(1.851,24.066)=17.474, p<.0005; Task 2: F(1.359,14.946)=6.653, p=.015), with Widgets (Task 1 avg=3.3mm; Task 2 avg=5.2mm) outperforming 6DOF (avg=15.0mm, p<.0005) in the first task and PRISM on both first (avg=10.7mm, p=.002) and second (avg=12.2mm, p=.003) tasks. The technique used also influenced rotation error (Task 1: $\chi^2(2)$=24.500, p<.0005; Task 2: $\chi^2(2)$=15.000, p=.001), with Widgets (Task 1 avg=0.0^o, Task 2 avg=0.0^o) achieving lower error than 6DOF (Task 1: avg=11.7^o, Z=-3.724, p<.0005; Task 2: avg=9.8^o) and PRISM (Task 1: avg=7.3^o, Z=-3.408, p=.003; Task 2: 7.1^o, Z=-2.803, p=.015).

Figure 8: *Participants during evaluation sessions.*

Widgets might have outperformed both 6DOF and PRISM in the first task, due to its DOF separation. Since this task required translating the object along a single axis, the ability to manipulate with such constraint allowed users to avoid unexpected rotations and translations, thus preventing error. The same principle applies to time completion, because users did not need to correct mistakes. Similarly, the second task saw better results with Widgets in both translation and rotation error, although the time taken by users had no significant difference against 6DOF. We believe this occurred because transformation separation found in the Widgets technique made it impossible to take a direct path, requiring users to move in all three axes separately.

In the second pair of tasks we focused on rotations. Significant differences for execution time were only found for the third task ($\chi^2(2)$=20.985, p<.0005), in which the use of Widgets (avg=27s) reduced time needed when compared to 6DOF (avg=53s, Z=-3.053, p=.006) and PRISM (avg=58s, Z=-3.823, p<.0005). For both tasks, position error revealed significant differences according to the technique used (Task 3: $\chi^2(2)$=16.545, p<.0005; Task 4: F(1.619,14.575)=6.586, p=.012). Widgets (Task 3 avg=0.0mm, Task 4 avg=9.7mm) led to better positioning than 6DOF in both tasks (Task 3: avg=13.3mm, Z=-3.296, p=.003; Task 4: avg=15.7mm, p=.008) and than PRISM in the third task (avg=16.6mm, Z=-3.059, p=.006). Rotation error was also significantly affected by the techniques (Task 3: $\chi^2(2)$=20.118, p<.0005, Task 4: $\chi^2(2)$=16.545, p<.0005). Once again, Widgets (Task 3 avg=1.8^o, Task 4 avg=5.3^o) performed better than 6DOF in both tasks (Task 3: avg=8.8^o, Z=-3.547, p<.0005; Task 4: avg=8.7^o, Z=-2.868, p=.012) and than PRISM in the third task (avg=7.1^o, Z=-3.574, p<.0005).

Alike the first pair, third and fourth tasks revealed advantageous results for Widgets in both translation and rotation error. Even though the focus of these tasks shifted from translation to rotation only, the ability to separate transformations proved to be, once again, significant. The increased completion time found in the fourth task, was a consequence of rotations around all axes. Users felt confused and unable to easily figure out the necessary rotations to reach the desired orientation.

The last pair of tasks required both translations and rotations. In both cases, techniques had an effect on the time participants took to complete tasks (Task 5: F(1.422,27.021)=12.645, p<.0005; Task 6: $\chi^2(2)$=27.900, p<.0005). While in the fifth task PRISM (avg=102s) was outperformed by both Widgets (avg=72s, p=.004) and 6DOF (avg=63s, p=.003), in the sixth Widgets (avg=135s) took longer than 6DOF (avg=55s, Z=-3.920, p<.0005) and PRISM (avg=112s, Z=-2.520, p=.036). 6DOF was also faster than PRISM in the final task (Z=-3.323, p=.003). In both tasks, there were differences regarding error in object positioning (Task 5: $\chi^2(2)$=8.533, p=.014, Task 6: F(1.671,23.391)=5.232, p=.017). Widgets (avg=6.6mm) reduced distance to target in the fifth task when compared to 6DOF (avg=15.1mm, Z=-2.809, p=.015) and PRISM (avg=21.4mm, Z=-3.010, p=.009). In the last task, 6DOF (avg=11.4mm) allowed users to place the object closer to its target position than PRISM (avg=21.2mm, p=.048). Analysing rotation error, we only found significant differences in the fifth task ($\chi^2(2)$=22.625, p<.0005), in which Widgets (avg=1.1^o) attained better results than 6DOF (avg=9.2^o, Z=-3.823, p<.0005) and PRISM (avg=8.9^o, Z=-3.464, p=.003).

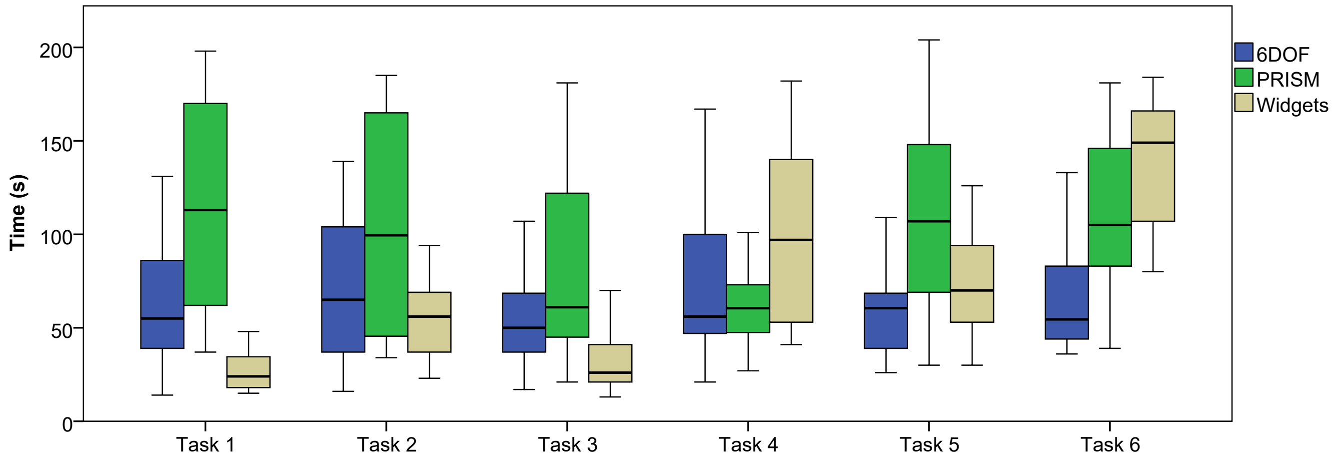


Figure 9: *Time to complete the six tasks using the three techniques, in seconds. The graphic presents the median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers).*

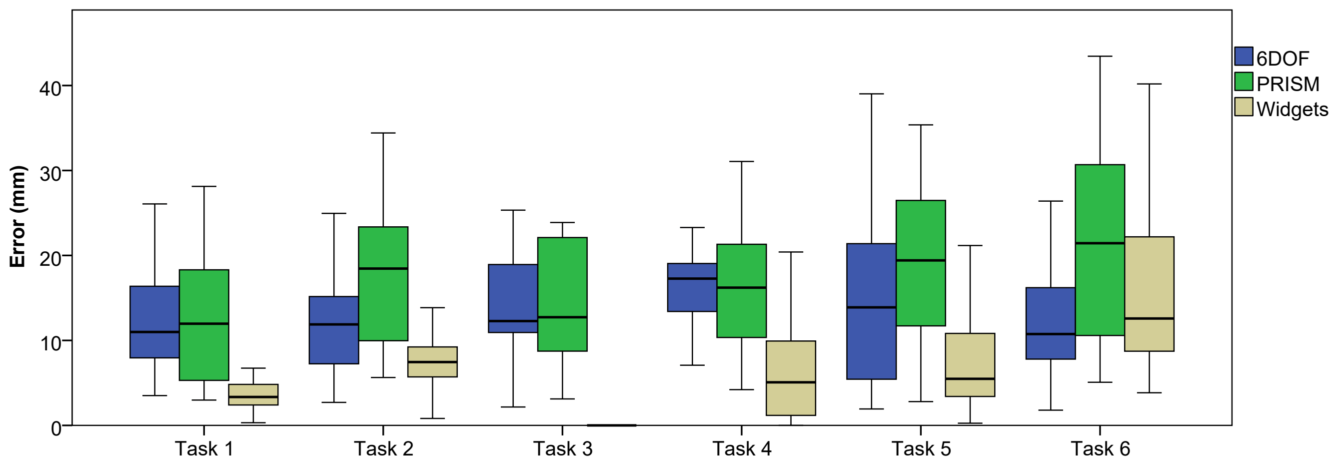


Figure 10: *Position error attained in the six tasks using the three techniques, in millimeters. The graphic presents the median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers).*

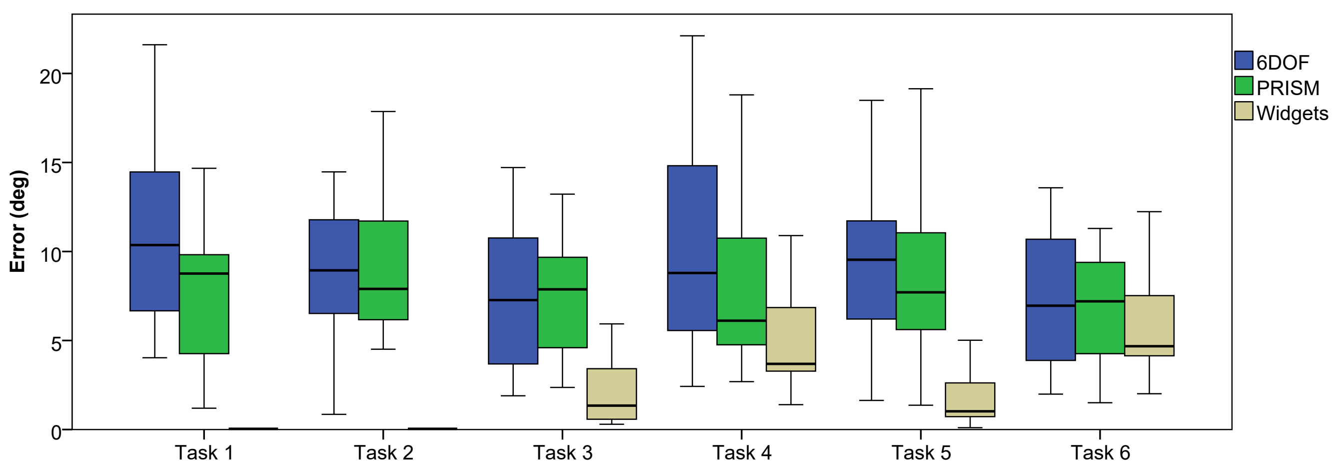


Figure 11: *Rotation error attained in the six tasks using the three techniques, in degrees. The graphic presents the median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers).*

	6DOF	PRISM	Widgets
Easiness*	4 (1)	2 (1)	4 (1)
Translation	4 (2)	4 (1)	4 (1)
Rotation*	3 (2)	2 (1)	4 (2)
Fun*	3 (2)	2 (1)	4 (2)

* indicates statistical significance

Table 1: *Participants preference for each technique, regarding different criteria (Median, Inter-quartile Range).*

Final tasks had an increase in complexity, since they both required participants to apply translations and rotations to the object. The time participants took to complete these tasks was negatively affected due to the necessary increased number of operations. As a consequence, translation and rotation error presented worse results when compared to previous tasks, because the time limit prevented participants to make final adjustments.

It is also worth of notice that both 6DOF and PRISM did not have major variations along all tasks, with no regard to its difficulty. For these techniques, after grabbing an object all tasks are alike, since there is no constraint in transformations being applied to the object. Taking the first and last task as an example, we used a Paired-Samples T Test and no significant differences were found in time, translation error or rotation error. Moreover, PRISM and 6DOF consistently shared similar results. As the authors pointed out, PRISM rotations are confusing for some users, which might have had a negative impact in tasks overall performance.

3.6.2 Subjective Data

Using questionnaires, we asked the participants how they felt about each technique. This included general easiness of use, translation and rotation difficulty and fun factor. Participants were given a Likert Scale from 1 to 5 to answer our questions, being 5 the favorable value. Answers are depicted in Table 1.

Analysing attained results, we identified significant differences in ease of use ($\chi^2(2)$=19.547, $p<.0005$), rotation difficulty ($\chi^2(2)$=25.352, $p<.0005$) and fun factor ($\chi^2(2)$=13.216, p=.001). Participants strongly agreed that PRISM was generally harder (Widgets: Z=-3.716, $p<.0005$, 6DOF: Z=-3.157, p=.006) and less fun to use (Widgets: Z=-3.057, p=.006, 6DOF: Z=-2.463, p=.042). Widgets appealed more to participants to perform object rotation than 6DOF (Z=-2.863, p=.012) and PRISM (Z=-3.874, $p<.0005$). Also, participants agreed that it is easier to rotate objects using 6DOF than PRISM (Z=-2.708, p=.021). There was no difference in translation difficulty, even though PRISM sacrifices directness and time over enhanced precision. The Widgets approach, although requiring more effort for complex movements, was as appealing to participants as other techniques. It is as fun as direct manipulation, but with increased final placement.

3.7 Discussion

Users found the widget-based approach as easy-to-use as 6DOF, and easier than PRISM. Overall, error attained in object placement using Widgets was smaller than with other approaches. However, this increased positioning sacrificed speed in more complex tasks. The results between 6DOF and PRISM are similar to the six DOF task from original PRISM evaluation [Frees et al. 2007], as we did not impose any minimum requirements for distance or angle between the object and its target placement. Indeed, in our evaluation PRISM's translation was praised by participants, as its operation was easy to understand and its benefits were clear. The main issue with this technique was found on rotations, where some users complained about it being confusing, as previously stated in [Frees et al. 2007]. Since there is no complete DOF separation in PRISM and none in 6DOF, as opposed to the Widgets approach, extra hand tremor or tracker noise occasionally caused unwanted transformations. This was mostly noted when users desired to only translate the object and an accidental rotation occurred. Distinctly from 6DOF, users found it difficult to return to the correct orientation with PRISM when this disturbance was too strong, which had a severe impact on performance of all tasks.

Since our tracking solution considerably differs from that used in [Frees et al. 2007], we experimented with different values for PRISM's scaling constant hoping to find better suited ones. However, we ended using those originally proposed, as mentioned in Section 3.1.2. Because this constant simultaneously affects when scaling is applied and how much movement is scaled, we could not identify a better compromise. We also found the method used to calculate hand's speed very prone to be negatively impacted by tracker noise. Instead of using information from two consecutive frames, it uses the difference between the current hand's position and that from 500ms ago. However, this does not totally prevent noise, but potentially reveals it half a second later.

4 Guidelines for Mid-air Object Manipulation

As a result of our evaluation, we were able to draw some guidelines for object manipulation in IVE. These should aid researchers and developers in creating better techniques that can combine the better aspects of each evaluated approach:

- Direct manipulation (6DOF) is well suited for coarse transformations. It allows fast and natural interactions, although not offering accurate placement;
- It should be possible to perform translation and rotation operations independently. We found that, in both 6DOF and PRISM, unwanted transformations happen when a simple translation or rotation is in order, which negatively impacts performance;
- Single DOF separation is very desirable for precise transformations, typically for fine-grain adjustments. This separation, more than separating translation and rotation, constrains transformations to a single dimension, preventing additional unwanted actions;
- Scaled transformations, as proposed in PRISM, are appealing only for translation. Separated scaled rotation in each coordinate axis confused participants, but they found scaled translations to be helpful in improving accuracy. Combining scaled translations with other approaches might improve their overall performance.

5 Conclusion and Future Work

Object manipulation is one of the most relevant tasks in virtual environments. While mid-air gestures in immersive virtual environments allow natural interactions, it is still difficult to place an object accurately, with the desired position and orientation. In this work, we conducted an evaluation to assess the benefits of DOF separation in mid-air, after it has been proved useful in other interaction paradigms by previous research. We concluded that indeed DOF separation through virtual widgets led to error reduction, at the cost of increased time for more complex tasks. Drawn from our results, we also proposed a set of guidelines to help developing better manipulation techniques.

As future work, we believe it might be interesting to combine Widgets and a direct approach, in a similar fashion to the 7-Handle technique [Nguyen et al. 2014], but with the more familiar design of the widgets we used. This will possibly allow for quick transformations in complex tasks, while keeping transformation separation advantages for final adjustments. Additionally, adding scaled translation in widget manipulation is also worth of consideration. This way, the benefits of PRISM's translation might be added to both translation and rotation, since these operations are applied only through hand translation. Further experiment with PRISM's scaling constant value and hand's speed calculation, for the latter to be less noise sensitive using a moving average for instance, may also lead to better performance. Finally, defining custom arbitrary transformation axis might be a compelling addition to the Widgets approach. By doing so, object manipulation keeps the DOF separation benefits, but reducing substantially the number of operations needed.

Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with ref. UID/CEC/50021/2013, through projects TECTON-3D (PTDC/EEI-SII/3154/2012) and IT-MEDEX (PTDC/EEISII/6038/2014), and doctoral grant SFRH/BD/91372/2012.

References

ARAÚJO, B. R. D., CASIEZ, G., JORGE, J. A., AND HACHET, M. 2013. Mockup builder: 3d modeling on and above the surface. *Computers & Graphics 37*, 3.

BENKO, H., AND FEINER, S. 2007. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *3D User Interfaces (3DUI) 2007*, IEEE.

BOLLENSDORFF, B., HAHNE, U., AND ALEXA, M. 2012. The effect of perspective projection in multi-touch 3d interaction. In *Proceedings of Graphics Interface 2012*, Canadian Information Processing Society.

COHÉ, A., DÈCLE, F., AND HACHET, M. 2011. tbox: a 3d transformation widget designed for touch-screens. In *Proceedings of CHI 2011*, ACM.

CONNER, B. D., SNIBBE, S. S., HERNDON, K. P., ROBBINS, D. C., ZELEZNIK, R. C., AND VAN DAM, A. 1992. Three-dimensional widgets. In *Proceedings of I3D 1992*, ACM.

FREES, S., AND KESSLER, G. D. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *Virtual Reality 2005*, IEEE.

FREES, S., KESSLER, G. D., AND KAY, E. 2007. Prism interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer-Human Interaction 14*, 1, 2.

HACHET, M., BOSSAVIT, B., COHÉ, A., AND DE LA RIVIÈRE, J.-B. 2011. Toucheo: multitouch and stereo combined in a seamless workspace. In *Proceedings of UIST 2011*, ACM.

HANCOCK, M., TEN CATE, T., AND CARPENDALE, S. 2009. Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of ITS 2009*, ACM.

HILLIGES, O., IZADI, S., WILSON, A. D., HODGES, S., GARCIA-MENDOZA, A., AND BUTZ, A. 2009. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of UIST 2009*, ACM.

HOUDE, S. 1992. Iterative design of an interface for easy 3-d direct manipulation. In *Proceedings of CHI 1992*, ACM.

KIN, K., MILLER, T., BOLLENSDORFF, B., DEROSE, T., HARTMANN, B., AND AGRAWALA, M. 2011. Eden: a professional multitouch tool for constructing virtual organic environments. In *Proceedings of CHI 2011*, ACM.

KIYOKAWA, K., TAKEMURA, H., AND YOKOYA, N. 1997. Manipulation aid for two-handed 3-d designing within a shared virtual environment. In *HCI (2)*.

MARTINET, A., CASIEZ, G., AND GRISONI, L. 2010. The effect of dof separation in 3d manipulation tasks with multi-touch displays. In *Proceedings of UIST 2010*, ACM.

MENDES, D., LOPES, P., AND FERREIRA, A. 2011. Hands-on interactive tabletop lego application. In *Proceedings of ACE 2011*, ACM.

MENDES, D., FONSECA, F., ARAUJO, B., FERREIRA, A., AND JORGE, J. 2014. Mid-air interactions above stereoscopic interactive tables. In *3D User Interfaces (3DUI) 2014*, IEEE.

NGUYEN, T. T. H., DUVAL, T., AND PONTONNIER, C. 2014. A new direct manipulation technique for immersive 3d virtual environments. In *ICAT-EGVE 2014*.

OSAWA, N. 2008. Two-handed and one-handed techniques for precise and efficient manipulation in immersive virtual environments. In *Advances in Visual Computing*. Springer.

PRACHYABRUED, M., AND BORST, C. W. 2014. Visual feedback for virtual grasping. In *3D User Interfaces (3DUI) 2014*, IEEE.

REISMAN, J. L., DAVIDSON, P. L., AND HAN, J. Y. 2009. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of UIST 2009*, ACM.

SHOEMAKE, K. 1992. Arcball: a user interface for specifying three-dimensional orientation using a mouse. In *Graphics Interface*, vol. 92.

SONG, P., GOH, W. B., HUTAMA, W., FU, C.-W., AND LIU, X. 2012. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of CHI 2012*, ACM.

STROTHOFF, S., VALKOV, D., AND HINRICHS, K. 2011. Triangle cursor: interactions with objects above the tabletop. In *Proceedings of ITS 2011*, ACM.

VEIT, M., CAPOBIANCO, A., AND BECHMANN, D. 2009. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of VRST 2009*, ACM.

WANG, Y., MACKENZIE, C. L., SUMMERS, V. A., AND BOOTH, K. S. 1998. The structure of object transportation and orientation in human-computer interaction. In *Proceedings of CHI 1998*, ACM.

WANG, R., PARIS, S., AND POPOVIĆ, J. 2011. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of UIST 2011*, ACM.