PhD Dissertation Proposal

Manipulation of 3D Objects in Immersive Virtual Environments

Daniel Filipe Martins Tavares Mendes
danielmendes@tecnico.ulisboa.pt

September 2015

Abstract

Interactions within virtual environments (VE) often require manipulation of 3D virtual objects. For this purpose, several research have been carried out focusing on mouse, touch and gestural input. While existing mid-air gestures in immersive VE (IVE) can offer natural manipulations, mimicking interactions with physical objects, they cannot achieve the same levels of precision attained with mouse-based applications for computer-aided design (CAD).

Following prevailing trends in previous research, such as degrees-of-freedom (DOF) separation, virtual widgets and scaled user motion, we intend to explore techniques for IVEs that combine the naturalness of mid-air gestures and the precision found in traditional CAD systems. In this document we survey and discuss the state-of-the-art in 3D object manipulation. With challenges identified, we present a research proposal and corresponding work plan.

Contents

1 Introduction

Since the early days of virtual environments that moving, rotating and resizing virtual objects have been target of research. Considering three-dimensional virtual environments, these kind of manipulations are not trivial, mainly due to the required mapping between traditional input devices (2D) and the virtual environment (3D). Most common solutions resort to techniques that somehow relate the actions performed in the two-dimensional space of the input device (e.g. mouse cursor or touch) to three-dimensional transformations.

Since it is usual for people to interact with these kind of environments with traditional displays, 3D content is displayed in a 2D rendered image, which hinders content's perception. To overcome both the limitations of the input and the output devices, mainstream solutions for creating and editing 3D virtual content, namely computer-aided design (CAD) tools, resort to different orthogonal views of the environment. This allows a more direct two-dimensional interaction with limited degrees of freedom. Solutions that offer a single perspective view usually either apply the transformation in a plane parallel to the view plane, or resort to widgets that constraint interactions and ease the 2D-3D mapping. Research has shown that the first approach can sometimes result in unexpected transformations when users are allowed to freely navigate through the virtual environment, and that constrained interactions allow for more accurate manipulations.

Recent technological advances lead to an increased interest in immersive virtual reality settings. Affordable hardware for immersive visualization of virtual environments, such as the Oculus Rift head-mounted display (HMD), ease the perception of three-dimensional content. Moreover, advances in user tracking solutions make possible to know where users' head, limbs and hands are in space. This allows for more direct interactions, mimicking the ones with physical objects. First results of our work showed that mid-air direct interactions with 3D virtual content can reduce tasks' duration and are appealing to users.

Although mid-air interactions show promising results, the accuracy of human spatial interactions is limited. Moreover, the limited dexterity of mid-air hand gestures, aggravated by the low-definition of current HMDs, constrain precise manipulations. This precision is of extreme importance when creating or assembling engineering models or architectural mock-ups.

Inspired by previous research that focused on mouse- and touch-based interactions, we intend to explore different techniques to increase users' precision for mid-air interactions, when working with 3D virtual models. Our objective is to combine the naturalness of everyday physical interactions with the precision that is only possible with computer systems. We expect to develop techniques that, albeit being less natural than the ones that strictly copy physical world interactions, offer more controlled and precise manipulations.

In the remainder of this document we will present some background and the state-of-the-art regarding 3D object manipulations within virtual environments. With open challenges identified, we introduce those we intend to tackle, referring which are our research hypothesis and objectives. We will then propose an approach to our research for the following couple of years.

2 Background and Related Work

Objects' manipulation in virtual environments have been subject of previous research. In this section we will cover the more relevant works in this field, ranging from manipulation using traditional input devices, such as mouse and keyboard, to more recent approaches that rely on mid-air interactions to interact with objects displayed through stereoscopic imagery. But, before we do so, we will present the main research groups and individuals that have been carrying those research and the most relevant venues where this kind of work has been being published in.

2.1 Key players

Several people conducted research on how to improve interaction in 3D virtual environments. Below are presented some of the research groups that have created valuable contributions.

Natural Interaction Research group at Microsoft Research This group[1] aims to enrich and reimagine the human-computer interface. Their team explores a wide variety of interaction topics including sensing and display hardware, touch and stylus input, spatial and augmented reality, and user modeling. Among others, notable names appear in their team: Bill Buxton, Andy Wilson and Hrvoje Benko.

HCI group at the University of Hamburg This group[2] explores perceptually-inspired and (super-)natural forms of interaction to seamlessly couple the space where the flat 2D digital world meets the three dimensions we live in. Their research is driven by understanding the human perceptual, cognitive and motor skills and limitations in order to reform the interaction as well as the experience in computer-mediated realities. The group is led by Frank Steinicke, who is also responsible for taking part in the organization of relevant conferences in the field, namely ACM SUI and IEEE 3DUI. Working also in this group, Gerd Bruder, a postdoctoral researcher, and Paul Lubos, a PhD candidate, have been working in 3D user interfaces for immersive environments.

Potioc at Inria Bordeaux The overall objective of Potioc[3] is to design, to develop, and to evaluate new approaches that provide rich interaction experiences between users and the digital world. They are interested in popular interaction, mainly targeted at the general public. They explore input and output modalities that go beyond standard interaction approaches which are based on mice/keyboards and (touch)screens. Similarly, they are interested in 3D content that offer new opportunities compared to traditional 2D contexts. More precisely, Potioc explores interaction approaches that rely notably on interactive 3D graphics, augmented and virtual reality (AR/VR), tangible interaction, brain-computer interfaces (BCI) and physiological interfaces. Martin Hachet leads this group, whose main area of interest is 3D User Interaction. One of the

---

[1] Natural Interaction Research group: `research.microsoft.com/en-us/groups/natural/`

[2] HCI group at the University of Hamburg: `www.inf.uni-hamburg.de/en/inst/ab/hci`

[3] Potioc: `team.inria.fr/potioc`

former members of this group, Aurélie Cohé, explored widget based manipulations for 3D object in her PhD.

MINT The MINT team[4] focuses on gestural interaction, i.e. the use of gesture for human-computer interaction. In the particular context of HCI, they are more specifically interested in users' movements that a computing system can sense and respond to. Their goal is to foster the emergence of new interactions, to further broaden the use of gesture by supporting more complex operations. They are developing the scientific and technical foundations required to facilitate the design, implementation and evaluation of these interactions. This team is led by Laurent Grisoni, and one of their alumni is Anthony Martinet, whose PhD focused on techniques for 3D object manipulation for multi-touch displays.

InnoVis at the Unniversity of Calgary InnoVis[5] was founded and is directed by Sheelagh Carpendale. They investigate innovations in the area of Information Visualization and Human Computer Interaction. Active research topics in Information Visualization include the exploration of effective use of display space, and navigation, exploration, and manipulation of information spaces. In the context of Human Computer Interaction members of the lab study how to best support collaborative work on large displays. One of their former PhD students, Mark Hancock, made great contributions to the field of interaction with 3D virtual objects using touch enabled surfaces.

2.2 Relevant events and journals

The research in the field of interaction with three dimensional content has been presented in several venues for the past years, being the most relevant described below. The first two are more focused than the remaining, and are more relevant for our work, while others have a wider scope and, consequently, a greater impact factor.

IEEE 3DUI The IEEE Symposium on 3D User Interfaces is the main symposium focused on the topic of 3D User Interfaces, currently on its tenth edition. The work presented in this forum range from traditional desktop interfaces to virtual and augmented reality setups, and have been addressing challenges such as selection, manipulation and navigation within virtual environments.

ACM SUI The ACM Symposium on Spatial User Interaction is focused on the user interface challenges that appear when users interact in the space where the flat, two-dimensional, digital world meets the volumetric, physical, three-dimensional space we live in. This considers both spatial input and 3D output, with an emphasis on the issues around interaction between humans and systems. This is a very recent symposium, happening annually since 2013.

ACM ITS The ACM International Conference on Interactive Tabletops and Surfaces has been established as a premier venue for research in the design, development and use of tabletop and interactive surface technologies. In the

---

[4] MINT team: `www.lifl.fr/mint/`

[5] InnoVis: `innovis.cpsc.ucalgary.ca`

past few years the conference and its community has grown out of its current name. In its ten editions, several works about mid-air interaction above surfaces, interactive spaces and spatial interaction, among others, have been presented here. As a result, there has been ample support for changing the name towards something that is more inclusive of the technologies that are being studied and developed in ITS.

IEEE VR IEEE Virtual Reality is an international conference and exhibition on virtual reality. This conference covers all areas related to virtual reality, including augmented reality, mixed reality, and 3D user interfaces.

ACM VRST The ACM Symposium on Virtual Reality Software and Technology is an international forum for the exchange of experience and knowledge among researchers and developers concerned with virtual reality software and technology.

GI Graphics Interface is an annual international conference devoted to computer graphics and human-computer interaction. With a graphics track and an HCI track having equal weights in the conference, GI offers a unique venue for a meeting of minds working on computer graphics and interactive techniques. GI is the longest running conference in the field (the first conference was held in 1969), consistently attracting high-quality submissions from graphics, HCI, as well as visualization. Since 2004, accepted papers have been archived in the ACM Digital Library.

IFIP INTERACT Starting with the first INTERACT conference in 1990, this conference series has been organized under the aegis of the Technical Committee 13 on Human-Computer Interaction of the UNESCO International Federation for Information Processing. This committee aims at developing the science and technology of the interaction between humans and computing devices.

ACM UIST The ACM Symposium on User Interface Software and Technology is an international forum for innovations in human-computer interfaces. UIST brings together people from diverse areas including graphical and web user interfaces, tangible and ubiquitous computing, virtual and augmented reality, multimedia, new input and output devices, and CSCW.

ACM CHI For over 30 years, the CHI conference has attracted the world's leading researchers and practitioners in the field of Human Computer Interaction from businesses and universities to share ground-breaking research and innovations related to how humans interact with digital technologies. The ACM CHI conference is the world's premiere conference on Human Factors in Computing Systems, presenting a highly selective showcase of the very best advances across the disciplines of computer science, cognitive psychology, design, social science, human factors, artificial intelligence, graphics, visualization, multimedia design and other disciplines.

ACM TOCHI This ACM Transaction seeks to be the premier archival journal in the multidisciplinary field of human-computer interaction. Since its first issue in March 1994, it has presented work of the highest scientific quality that contributes to the practice in the present and future. The primary emphasis is on results of broad application, but the journal considers original work focused on specific domains, on special requirements, on ethical issues - the full range of design, development, and use of interactive systems.

2.3 Virtual Environments Overview

Virtual environments have been around for some time, and are used for a plethora of purposes. Ranging from bioengeneering and geology [60], automotive engineering [44], manufacturing [46], architectural mockup [3] and CAD [28], to even creative painting [29], animation movies [42] and entertainment with building blocks [40], virtual environments are something we take for granted nowadays.

Perception of virtual environments can be enhanced by combining stereoscopic visualization and head tracking, to increase user immersion. By knowing the user's head position, it is possible to generate a visualization frustum to each eye to create the illusion of virtual objects being part of the physical world. This illusion is even stronger when users are allowed to freely move their heads and see different sides of a virtual object in their own perspective, without the need to manipulate cameras or widgets.

Although HMDs and CAVEs[6], which allow a fully immersive viewing experience, have existed for a while, in the last few years interest in them increased considerably. One of the main issues with older HMDs was the nausea they caused, commonly referred to as virtual reality sickness or cybersickness. New technological developments, namely in gyroscopes and in small displays, led to reduced latency of the digital imagery, thus reducing nausea and increasing presence, as stated in Virginia Heffernan's article, *Virtual Reality Fails Its Way to Success* (New York Times, 2015).

Other recent technological advances also made it easier to develop immersive visualization scenarios. Not so long ago, user tracking required rooms equipped with expensive infra-red cameras and markers attached to people or invasive wire-based systems. Currently, tracking is possible using affordable and non-intrusive depth cameras. This tracking solution can be used to not only find the user perspective to render the virtual scene, but also to track user limbs and hands, unveiling new interaction possibilities. Also, this combination of stereoscopic displays and user tracking allows users to naturally manipulate three dimensional entities as if they were collocated with their hands and body, extending traditional two-dimensional interactions in very natural ways.

A virtual environment that can be explored through immersive displays is often called an immersive virtual environment (IVE), or that it is an immersive virtual reality (IVR). Although a fully immersive environment should explore other human senses besides vision, as studied by Azevedo et al. [4], the IVE classification is often used when using only an immersive display. According to Bowman et al. [9], these kinds of displays can be divided into two categories: fully immersive displays, such as HMDs, totally occlude the real world; semi-

---

[6] CAVE: Cave Automatic Virtual Environment.

immersive displays, such as stereo tabletops, allow users to see both the physical and virtual world. Benefits of higher levels of immersion in virtual reality setups have already been presented [7].

To classify some aspects of virtual environments, we used the taxonomy depicted in Figure 1. This taxonomy is based on the one proposed by Grossman and Wigdor [19], which was initially conceived for tabletops, but some concepts can easily be easily applied to virtual environments in general.

The first area we considered is the display properties. This area is divided into perceived space, actual space and viewpoint correlation. The perceived space is defined by the possible locations of the virtual objects. This location can be: constrained to the display, which is the case of traditional screens that even when displaying a perspective projection the image displayed is 2D; or volumetric, when stereoscopic technologies are used, such as shutter glasses or HMDs, providing the illusion of objects placed in 3D. The actual space relates to where the rendered image is presented, which although not changing the user's perception of the virtual environment, can influence user's depth perception and performance in 3D tasks. This space can be constrained to the 2D screen, which is the case even when perceiving the image in 3D with shutter glasses, and issues like hands occlusions may arise. To overcome this, there are heads-up surfaces, such HMDs or see-through screens placed between the user's eyes and hands. The viewpoint correlation concerns the relation between the user's point of view and the viewpoint of the virtual scene. In systems where the user moves around the display and the viewpoint remains constant, there is none relation. For systems that change the viewpoint of the render accordingly to user's head position, we say that there is a high or total correlation. High refers to setups comprised of a screen, either vertical or horizontal, that when the user moves his head behind it will see the back of the screen instead of the virtual environment from a different perspective. When using a HMD to create a virtual reality experience, total correlation between the user's viewpoint and the displayed imagery can be achieved.

Regarding the input properties, that focus on how the user interacts with the

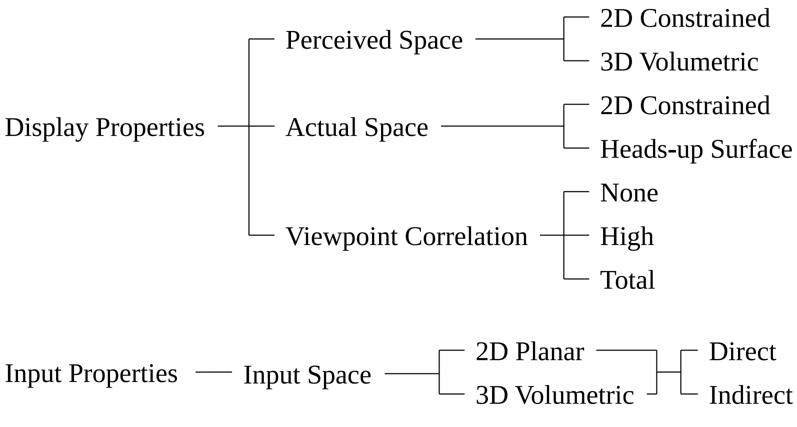


Figure 1: Taxonomy of virtual environments (adapted from [19]).

displayed image, we can categorize the input space. It can be planar, having a 2D input data, like touch surfaces, or volumetric, capturing user's actions in a 3D space. On the other hand, the input can also be direct, when the user's directly selects and manipulates virtual objects, or indirect, when the user interacts with a cursor or widgets to manipulate the object.

In addition, interactive computer applications that deal with 3D objects often must enable manipulation of one or more of those objects. Manipulation has three parts: selection, spatial transformation, and release. We will focus on the spatial transformations, which can be divided into translation, rotation and scale. Each one of this transformation can be applied to three different axis (x, y, z). A single transformation on one of these axes is commonly referred to as a degree-of-freedom (DOF). So, for a system that allows all transformations in all these axes, it is said that it allows transformations in 9DOF. For systems that only offer translation and rotation in 3D, they are referred to support 6DOF, and for for those who add to this uniform scaling it is said they support 7DOF. DOF is also used to specify devices' tracking capabilities. For example, a mouse can track position changes in a plane (2D), so it is a 2DOF device. A spatial wand, whose position in space (3D), pitch, roll and yaw are tracked, is a 6 DOF device.

In the following sections we present the most relevant research work regarding 3D virtual object manipulation. We will cover: traditional desktop interaction (with 2D constrained displays and mouse based 2D indirect input); touch manipulation (on similar 2D constrained displays, but with 2D direct input); interaction with stereoscopic tabletops (also resorting to touch, but offering 3D volumetric perceived space and high viewpoint correlation); mid-air manipulations (using 2D displays, either with 2D or 3D perceived space, and a 3D input space); and lastly interactions within immersive virtual environments (3D volumetric perceived space, total viewpoint correlation and 3D input space). We then discuss the presented works, which will motivate our research proposal.

2.4 Mouse and Keyboard based 3D Interaction

Many computer applications require virtual three dimensional object manipulations, such as architectural modeling, virtual model exploration, engineering component design and assembly, among others. To work with virtual environments for this purpose, several interaction techniques for traditional desktop setups have been explored, resorting to mouse and keyboard devices.

In order to overcome the mapping of 2D mouse input to 3D transformations, Stephanie Houde developed an approach based on a handle box [27]. It consisted of a bounding box surrounding the object, and had a lifting handle attached to it, to move the object up and down, and four rotation handles, to rotate the object about its central axis, as illustrated in Figure 2. No handle was provided for sliding in the object's resting plane, on the assumption that the most direct way to slide an object would be to click and drag on the object inside the box itself. Conner et al. [13] also resorted to virtual handles to develop 3D widgets for performing transformations on virtual objects. They allow full 9 DOF control and even twist deformations. The handles have a small sphere in their end, and are used to constrain geometric transformations to a single plane or axis (Figure 3). Dragging one of the spheres can translate, rotate or scale the object depending on which mouse button is pressed. For rotations, the direction

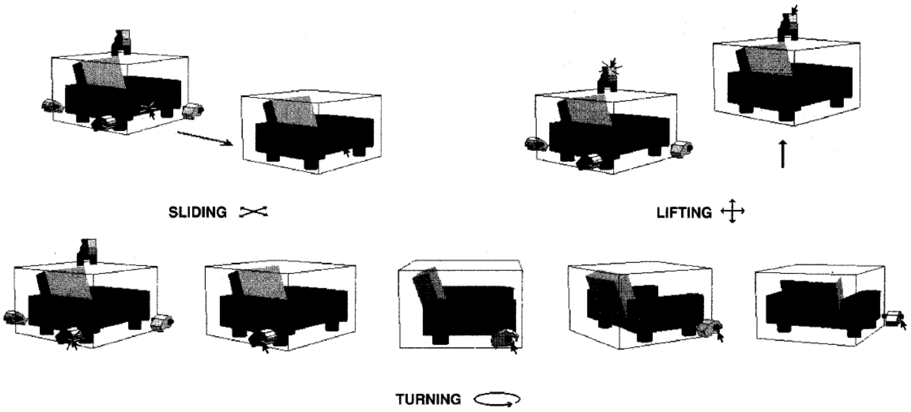


Figure 2: Sliding, lifting and turning a virtual object using the handle box approach (extracted from [27]).

of the user's initial gesture determines which of the two axes perpendicular to the handle is used as rotation's axis.

Focusing only in rotations, Ken Shoemake proposed Arcball [56], an input technique that uses a mouse to adjust the spatial orientation of an object. To change the object's orientation, the user draws an arc on a screen projection of a sphere. For axis constrained rotations, Arcball include the view coordinate axes, the selected object's model space coordinate axes, world space coordinate axes, normals and edges of surfaces, and joint axes of articulated models (such as robot arms). Mouse, menu, or keyboard combinations can be used to select among axis sets. As an example, for body coordinate axes, three mutually perpendicular arcs would be drawn, tilted with the object. When the mouse is clicked down to initiate a rotation, the constraint axis selected will be that of the nearest arc.

More than 20 years have passed since these techniques have been proposed, and they are still being used today in several solutions, even commercial ones. Indeed, some applications that require object manipulation, like Unity3D[7] or SketchUp[8], resort to widgets both for mapping between input devices and cor-

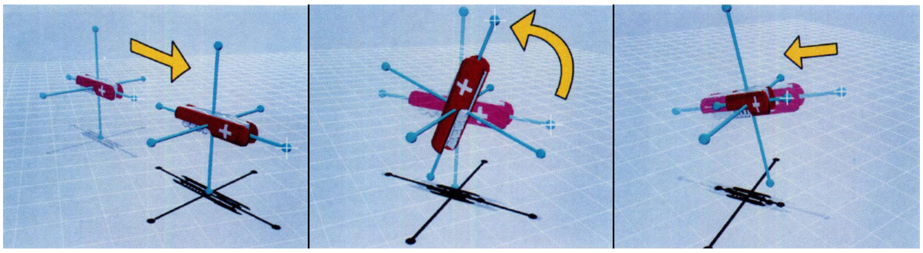


Figure 3: Virtual handles for object manipulation: translation (left), rotation (middle) and scale (right) along a single axis (extracted from [13]).

---

[7] Unity3D: `unity3d.com`

[8] SketchUp: `www.sketchup.com`

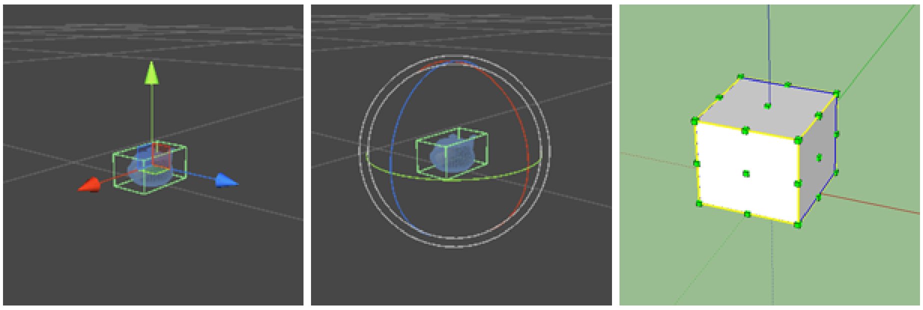


Figure 4: Widgets used in current commercial applications: virtual handles (left) and Arcball (middle) in Unity3D; handle box (right) in SketchUp.

responding 3D transformations and for restricting DOF manipulation. For interactively translate and scale virtual objects, Unity3D, a commonly used game engine, allows users to do so through virtual handles, as depicted in Figure 4, similarly to Conner et al. [13]. For rotations it uses a direct implementation of the Arcball [56]. SketchUp, a 3D modelling application, resorts to a handle box for object scaling, also shown in Figure 4. It provides quick and accurate modeling, aided by dynamic snapping, input of exact values for distances, angles and radius. All these solutions allows users to perform a transformation in a single axis at a time.

Other commercial applications, namely for 3D modelling, often present a different option. Instead of using widgets to restrict DOF manipulation, they allow the 3D virtual environment to be presented through three orthogonal views. Examples of this are 3D Studio Max[9] or Blender[10] (Figure 5). This way, each view allows simple 2D manipulations, along different axes, overcoming mapping issues. However, they require users to have greater spacial perception, rendering them suitable only for expert users. AutoCAD[11], which is more focused in architectural and engineering projects, also features these orthogonal viewports and allow for extremely precise manipulation of the elements within the virtual environment.

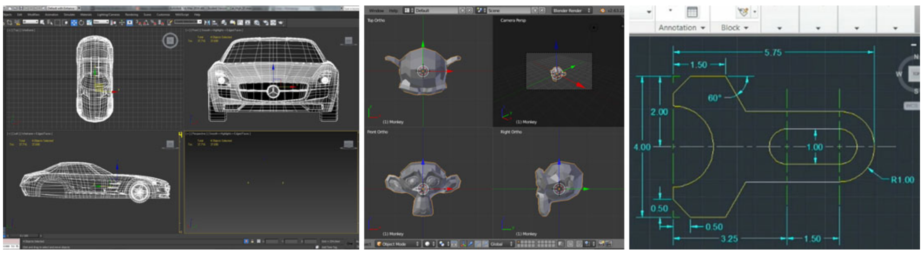


Figure 5: Orthogonal viewports in 3D Studio Max (left), Blender (middle) and AutoCAD (right).

---

[9] Autodesk 3D Studio Max: `www.autodesk.com/products/3ds-max/overview`

[10] Blender: `www.blender.org`

[11] Autodesk AutoCAD: `www.autodesk.com/products/autocad/overview`

2.5 3D Manipulation on Interactive Surfaces

Beyond the traditional WIMP-based[12] approaches, several multi-touch solutions to manipulate 3D objects have been proposed and evaluated over the past few years. In fact, touch enable displays have been available for long, but their booming interest came after Jeff Han's work [21], and his acclaimed speech in TED. With these interactive surfaces, new interaction possibilities came, allowing researchers to explore more Natural User Interfaces (NUI) [64]. Efforts have been done trying to create more direct interactions with virtual content, closer to the ones with physical objects, which successfully can surpass mouse based interactions [31]. Touch enabled surfaces are now present in our everyday life, through smartphones and tablets. Interactive tabletops are also becoming more and more popular. This kind of surfaces have been used for a variety of purposes, including interaction with 3D virtual content.

Hancock et al. [22] developed techniques to control 6DOF using one, two and three touches. The authors started by extending the RNT algorithm [34] to the third dimension. Touching an object, that object will follow the finger, rotating along all three axis and translating in two dimensions, as depicted in Figure 6. Using two touches, the original two-dimensional RNT is used with the first touch, while the second touch rotates the object in the remaining axes. The distance between the two touches changes the object depth. The three touches approach uses the first contact point for translations in a two-dimensional plane, the second to yaw and manipulate depth, and the third to pitch and roll. After evaluating this techniques, the authors concluded that an higher number of touches provides both better performance and higher user satisfaction. These results suggest that a close mapping of input and output DOFs is desirable. Authors also defined a set of requirements for multi-touch interfaces, such as creating a visual and physical link with objects and providing suitable 3D visual feedback. Later, they improved the proposed techniques with Sticky Fingers and the Opposable Thumb [23]. This solution is very similar to the three touches technique, but in this solution the third touch is used to rotate the object around the axis defined by the first two touches (Figure 7).

Considering the *de facto* standard for 2D manipulations, the Translate-Rotate-Scale (TRS) or Two-Point Rotation and Translation with scale [24], Reisman et al. [53] proposed a method to use several points of contact in a multi-touch device to manipulate 3D objects in 6 DOF. Their solution keeps the contact points fixed throughout the interaction, using a constraint solver to move and rotate objects at the same time. This solution is similar to the Opposable Thumb, but if the movement of the third finger is not perpendicular to the defined axis, that axis is no longer used and the object will rotate in or-

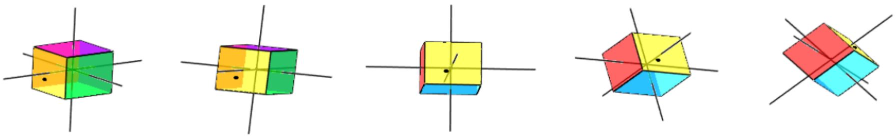


Figure 6: Shallow-depth interaction using one touch (extracted from [22]).

---

[12] WIMP: windows, icons, menus and pointing devices.

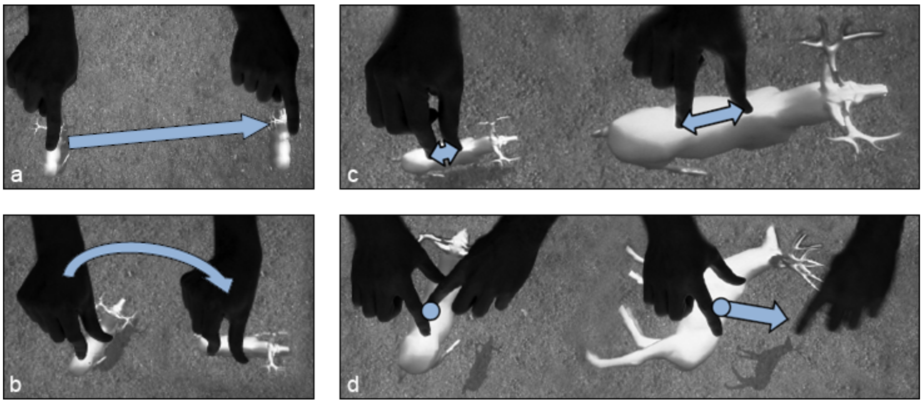


Figure 7: Sticky Fingers technique (a, b, c) and the Opposable Thumb (d) (extracted from [23]).

der to follow the finger, as illustrated in Figure 8. The main issue of providing an integrated solution to manipulate different transformations simultaneously is that unwanted operations arise frequently. To remedy this, the separation of DOF manipulation has been suggested [47] and followed in different research works.

Martinet et al. [36] proposed two techniques to translate 3D objects, shown in Figure 9. The first extends the viewport concept found in many CAD applications (four viewports, each displaying a different view of the model). Touching and dragging the object within one of the viewports translates the object in a plane parallel to that view. Manipulating the object with a second touch in a different viewport modifies depth relatively to the first touch. For the second method, denoted as Z-technique, only one view of the scene is employed. In this technique, the first touch moves the object in the plane parallel to the view, while the backward-forward motion of a second touch controls the depth relatively to the camera position. The authors preliminary evaluation suggests that

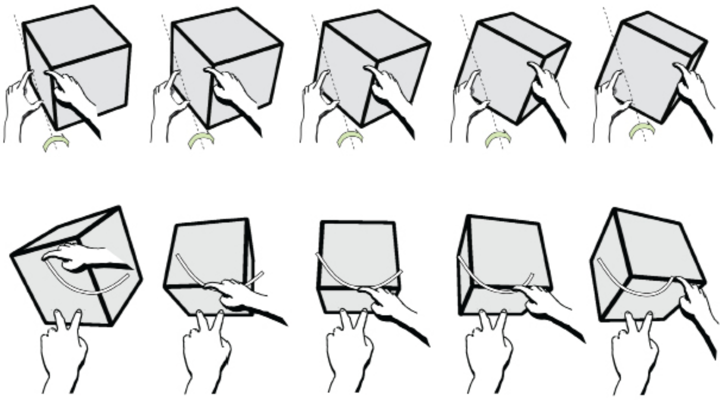


Figure 8: Screen-Space formulation - two different rotations with three touches (extracted from [53]).

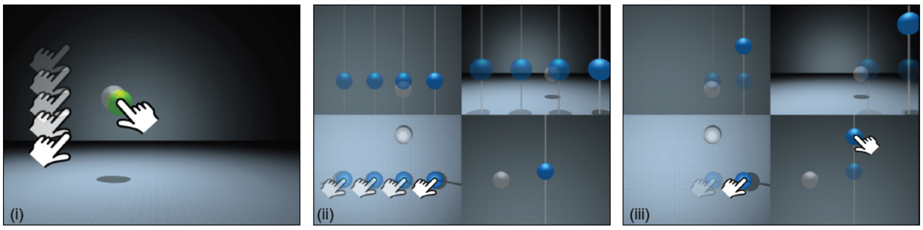


Figure 9: Z-Technique (i) and the orthogonal viewports approach (ii and iii). Gray lines indicate possible motions for the second touch (extracted from [36]).

users prefer the Z-technique.

Improving upon the Z-Technique, Martinet et al. introduced the DS3 [37], a 3D manipulation technique based on DOF separation. TRS is applied using two touches in the object and the Opposable Thumb is used for pitch and roll. To manipulate object depth, authors resorted to their previous approach, the Z-Technique [36], which uses the vertical motion of a touch outside the object. The authors compared DS3 with previous works [23, 53] and a user evaluation revealed that DOF separation led to better results. However, using a transformation plane parallel to the view plane can sometimes result in awkward transformations, when the view plane is not orthogonal to one of the scene axis [38].

To better understand user gestures for 3D manipulation tasks on multi-touch devices, Cohé et al. [12] conducted a user study and concluded that physically plausible interactions are favored and there are different strategies to develop an application focusing in a broad usage or ease of use. Based on observations of users interacting with widgets for 3D manipulations, Cohé et al. [11] designed a 3D transformation widget, the tBox. This widget allows the direct and independent control of 9 DOF (translation, rotation and scale along each axis). The tBox consists in a wire-frame cube, visible in Figure 10. Users can drag an edge of the cube to move the object in an axis containing the edge, and rotations are achieved by dragging one of the cube's faces.

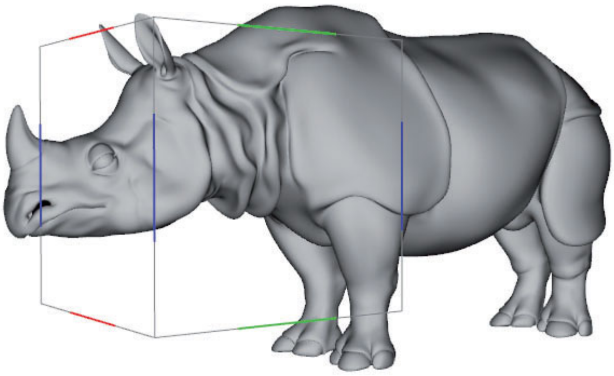


Figure 10: The tBox widget (extracted from [11]).

Figure 11: Placing (left) and rotating (middle) objects in Eden (extracted from [36]), and LTouchIt's Rotation-Handles (right) (extracted from [40]).

To create virtual environments for computer-animated films, Kin et al. [32] designed and developed Eden, a fully functional multi touch set construction application (Figure 11). Virtual objects can be translated in a horizontal plane using the usual direct drag approach, and up and down with a second finger, similar to the Z-technique [36]. Rotations are performed similar to the Arcball [56]. It also supports both uniform and one dimensional scaling transformations.

Our previous work, LTouchIt [40], although using direct manipulation for translations, also relies on widgets for rotations. Following the DOF separation, we developed a set of interaction techniques that provide direct control of the object's position in no more than two simultaneous dimensions and rotations around one axis at a time, using Rotation-Handles. The translation plane is perpendicular to one of the scene axes and is defined by the camera orientation. Using the Rotation-Handles, the user can select a handle to define a rotation axis and, with another touch, specify the rotation angle, as exemplified in Figure 11.

Regarding direct versus indirect interactions, Knoedel et al. [33] investigated the impact of the directness in TRS manipulation techniques. Their experiments indicated that a direct approach is better for completion time, but indirect interaction can improve both efficiency and precision.

More recently, Bollensdroff et al. [6] redesigned older techniques for three-dimensional interactions [27] using multi-touch input. They developed a cube shaped widget, the Gimbal Box, which uses a touch in one of its faces to translate in the plane defined by that face (Figure 12.a). To rotate the object the widget has two variations. One uses the TRS applied to a cube's face (Figure 12.b) or, alternatively, touching an edge of the box induces a rotation around an axis parallel to the edge (Figure 12.c). The other variation is based oh the Arcball [56]

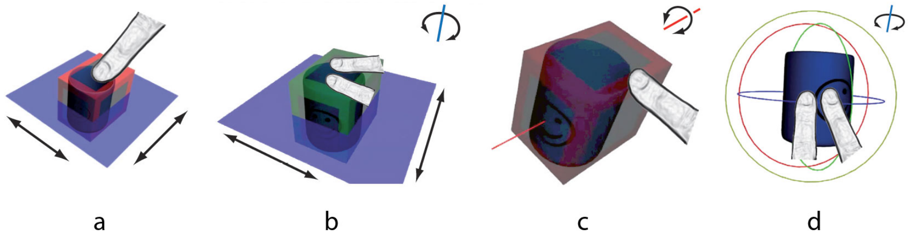


Figure 12: GimbalBox - translation (a) and different approaches for rotation (b, c, d). (extracted from [6]).

(Figure 12.d). Through a controlled study, their techniques were compared to other approaches well-known in the literature [23, 53]. They concluded that adapted widgets are superior to other approaches to multi-touch interactions, supporting DOF separation through the reduction of simultaneous control to 4 DOF in a defined visible 2D subspace. Moreover, the authors suggest that "multi-touch is not the final answer", since "the projection of an object as input space for interaction can never reproduce precise motions of the object in 3D space".

2.6 Touching Stereoscopic Tabletops

To improve both three dimensional visualization and spacial perception, several researchers explored interactions using stereoscopic environments. Considering the placement of virtual objects inside the tabletop in a fish-tank approach, touch solutions suffer from parallax issues [45]. Above the table solutions have already been explored. Using the Responsive Workbench, one of the first stereoscopic tabletop VR devices, Cutler et al. [14] built a system that allows users to manipulate virtual 3D models with both hands. The authors explored a variety of two-handed 3D tools and interactive techniques for model manipulation, constrained transformations and transitions between one- and two-handed interactions. However, they resorted to toolboxes to allow the user to transition between different operations.

Benko et al. [5] proposed a balloon metaphor to control a cursor (Figure 13), which is then used to manipulate three-dimensional virtual objects on a stereoscopic tabletop. Moving two fingers closer, the user allows the object to move up and, likewise, if the user moves the fingers away, the object will translate downwards. Later, Daiber et al. [15], created a variation of this technique by adding a corkscrew metaphor, that can be used with either both hands or single-handed. With this approach, the user can use a circular motion in a widget to manipulate object's height, instead of the distance between fingers. The authors compared their technique with the previous in both positive and negative parallax scenarios. Although none of the techniques was clearly identified as better, the negative parallax space was shown to be more difficult to interact with.

Strothoff et al. [59] proposed another approach to select and manipulate a cursor in stereoscopic tabletops. Using two fingers to define the base of a triangle, the height of the cursor, placed in the third vertex, is defined by the

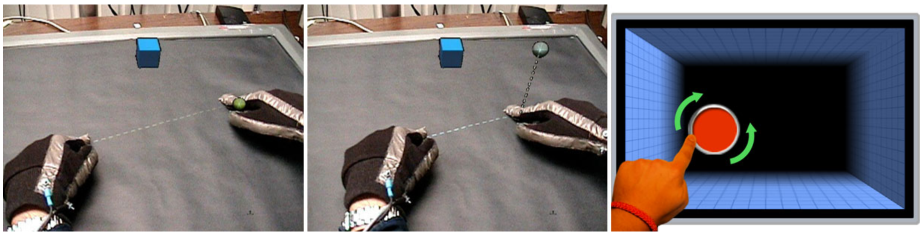


Figure 13: The balloon metaphor (left and middle): moving two fingers closer translates the cursor upwards (extracted from [5]). Corkscrew variation (right): circular motions replace the distance between touches (extracted from [15]).

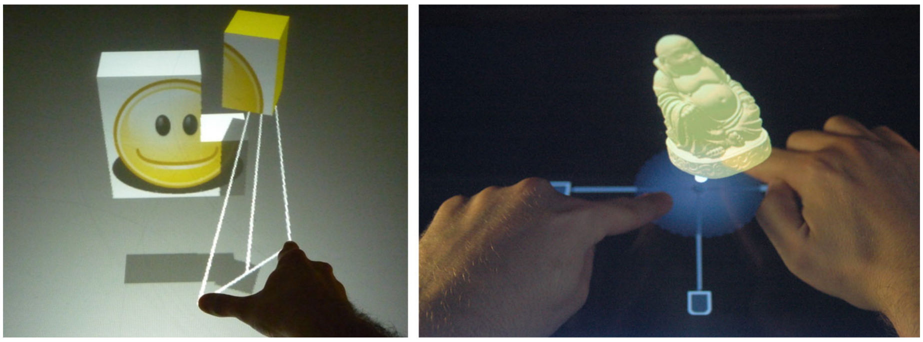


Figure 14: Left: triangle cursor (extracted from [59]). Right: Toucheo interaction (extracted from [20]).

distance of the two touches. Using this triangle cursor, users can manipulate selected objects in 4 DOF: translation in three dimensions and rotations around a vertical axis.

To manipulate virtual objects in full 9 DOF, Toucheo [20] proposes a setup with co-located 3D stereoscopic visualization, allowing people to use widgets on a multi-touch surface, while avoiding occlusions caused by the hands. The authors combined a two-dimensional TRS interaction on the surface with the balloon metaphor [5] and other widgets that provide both the remaining rotations and independent scale along three axes.

2.7 Mid-Air Interactions

Mid-air interaction has the potential to manipulate objects in 3D with more natural input mappings. Hilliges et al. [25] presented a technique to seamlessly switch between interactions on the tabletop and above it. The main goal of the authors was to create a solution that resembles physical manipulations, enabling depth based interactions. Using computer vision, the user's hand is tracked in 4 DOF (3 for translation and 1 for rotation) and the grab gesture can be detected. Shadows of user's hands are projected into the scene, which are used to interact with virtual objects in three dimensions. After an object being grabbed by the user's shadow, the modifications in the corresponding hand are applied to the object, as exemplified in Figure 15. Marquardt et al. [35] also combined the multi-touch surface and the space above it, in a continuous interaction space. Taking advantage of this space, they leveraged the user's hands movements to allow full 6 DOF interaction with digital content.

Hilliges et al. [26] created a similar setup to Toucheo [20], the Holodesk, which allowed direct interaction with 3D graphics, a shown in Figure 16, using physical simulation and a depth camera for hand tracking. Mockup Builder [2, 3] offers a semi-immersive modeling environment, in which the user can freely manipulate three dimensional virtual objects. The authors used Gametrack devices to follow users' fingers position in 3DOF, which acted as cursors, and adapted TRS to three dimensions to manipulate objects with 7DOF (we will refer to this technique as Air-TRS).

Instead of using a cursor approach for the handheld device, Kim and Park [30] proposed a Virtual Handle with a Grabbing Metaphor (VHGM). When the user

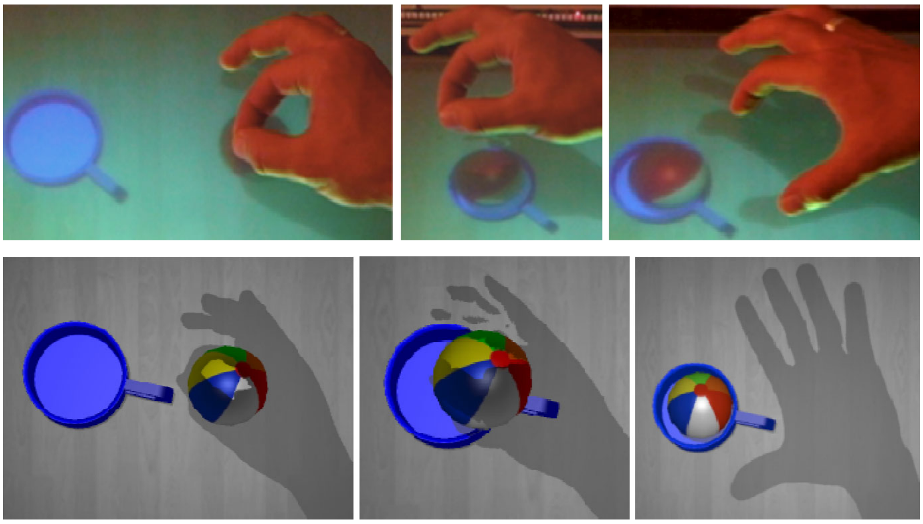


Figure 15: Virtual shadows used to manipulate an object (extracted from [25]).

selects an object, the system generates a bounding sphere around it. From the sphere's center, a ray with the direction opposite that of the virtual handle is projected to find the intersecting point on the sphere. This point serve as the reference frame for the following transformations (translation and rotation). User evaluation results suggest that VHGM can lead to better rotation efficiency than a standard 3D cursor.

Song et al. [57] explored spatial interactions proposing a handle bar metaphor as an effective way to transform 3D objects in mid-air. This technique allows users to manipulate single objects or pack multiple objects along the line described by both hands. Users can translate and rotate objects in space by moving their hands, as depicted in Figure 17, and scale it by changing the distance between them.

Schultheis et al. [55] made a comparison between mouse, wand and a two-

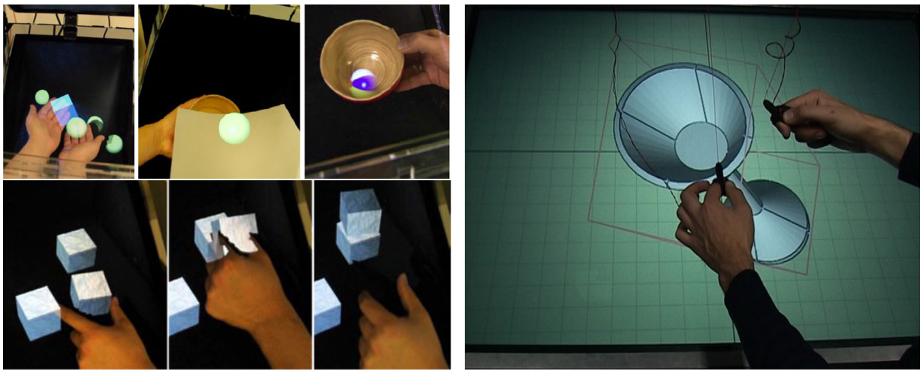


Figure 16: Left: users interacting with Holodesk (extracted from [26]). Right: User scaling an object in Mockup Builder (extracted from [3]).

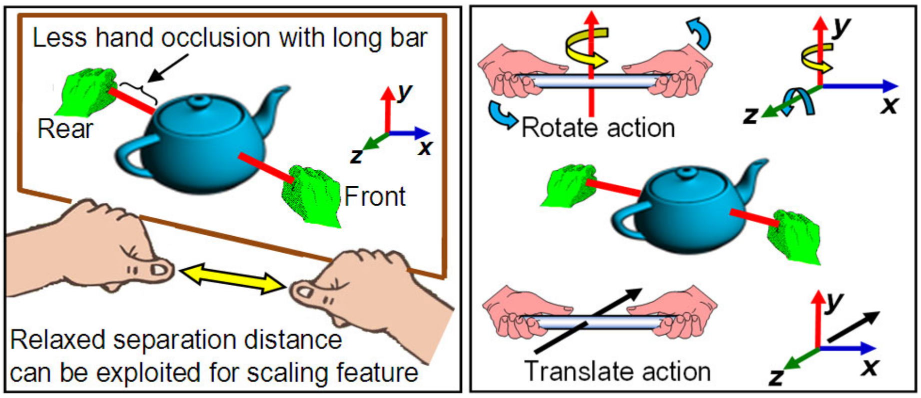


Figure 17: The handle bar metaphor used to translate, rotate and scale a virtual object (extracted from [57]).

handed interface for 3D virtual object and world manipulation through user evaluation, using both monoscopic and stereoscopic displays (although nothing is said about viewpoint correlation or co-location of users' hands and virtual imagery). The mouse interface resorted to manipulators (or widgets) for controlling translation and rotation angles for each axis. The wand behaved like a regular 6 DOF tracked device, allowing direct manipulation of the selected object. The two-handed approach is extremely similar to the Handle-Bar [57]. The two-handed interface out-performed the mouse and wand, and the wand out-performed the mouse, albeit requiring appropriate training. The authors state that these results suggest that well designed many-DOF interfaces have an inherent advantage over 2DOF input for fundamental 3D tasks.

The Color Glove [63], despite being an invasive wearable device, enabled precise finger and hand pose tracking. The system uses a simple RGB camera to capture the coloured areas of the gloves, being able to reconstruct the whole user's hand, thus attaining full 6 DOF tracking in real time. More recently, Wang et al. [62] introduced a new way to track hands and fingers using affordable depth cameras. Their approach, besides pose detection, tracks each hand in 6 DOF in a non-invasive manner. These tracking solutions allow hand reconstruction, which can be used to closely mimic physical interactions.

Vuibert and et al. [61] compared the performance of three mid-air interaction options, using either a physical replica of the virtual object, a wand-like device or the user's fingers. For this, they carried out a user evaluation with a docking task with 6DOF. As baseline they resorted to a mechanically constrained input device, the Phantom Omni. Authors found that the Phantom was the most accurate device for position and orientation, whereas the tangible mid-air interactions (wand and object's replica) were the fastest. Even though the fingers did not outperform the Phantom in accuracy or speed, the difference between these two conditions was small. Moreover, subjects preferred the wand and fingers, while interaction with the replica was the least favored.

We performed a comparative study between different interactions for 3D object manipulations using a setup that combines spatial 6DOF hand tracking and a multi-touch stereo tabletop [39]. We compared a touch approach similar to Toucheo [20], and four mid-air techniques: 6DOF Hand, which uses the dominant hand to grab, move and rotate objects, and the distance between both hands for scale; 3DOF Hand, in which the dominant hand only moves the object, while rotation and scale are given by the non-dominant hand; Air-TRS, as used in Mockup Builder [3]; and the Handle-bar [57]. User evaluation results suggest that mid-air interactions are better than touch based, and 6DOF Hand and Handle-bar are both faster and preferred by participants.

Feng et al. [16] conducted an evaluation similar to ours [39], but with held devices. Their setup differed from ours, where instead of co-locating users' hands with stereoscopic imagery, they used a fish tank stereoscopic visualization with offset manipulation techniques. Similarities with our results lead to a tentative guideline: if satisfying each individual user's preference is of high importance to the interface designer, give the user the option of Spindle+Wheel (Handle-Bar) or Grab-and-Scale (6DOF Hand) derived methods; otherwise use Grab-and-Scale (6DOF Hand).

2.8 Within Immersive Virtual Environments

Using HMDs and a tracker for hands' position and orientation, interactions such as grab, move and rotate objects can be done in virtual environments, similarly to how they are performed with physical objects [54].

One of the first challenges tackled concerning objects' manipulation in immersive virtual environments was how to extend users' capabilities by allowing interaction with objects out of reach of users' hands. The Go-Go immersive interaction technique [52] uses the metaphor of interactively growing the user's arm and nonlinear mapping for reaching and manipulating distant objects (Figure 18). When the user's hand move above a certain distance, the arm grows accordingly to a predefined coefficient. Below that distance, a 1:1 mapping is used. This technique allows for seamless direct manipulation of both nearby objects and those at a distance. However, when comparing the Go-Go technique with other approaches, such as infinite arm-extension and ray-casting (Figure 18), there is no clear winner [8]. User evaluation's results shows significant drawbacks in all techniques.

A different approach for interacting with out-of-reach objects in large virtual environments is the Worlds in Miniature technique [58], also visible in Figure 18. Users can interact with a miniature of the virtual world to promptly move around and change their point-of-view or to manipulate virtual objects.

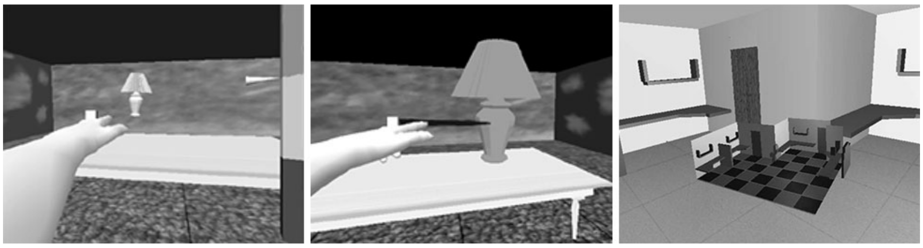


Figure 18: The Go-Go (left), ray-casting (middle) and Worlds in Miniature techniques (right) (extracted from [8] and [58]).

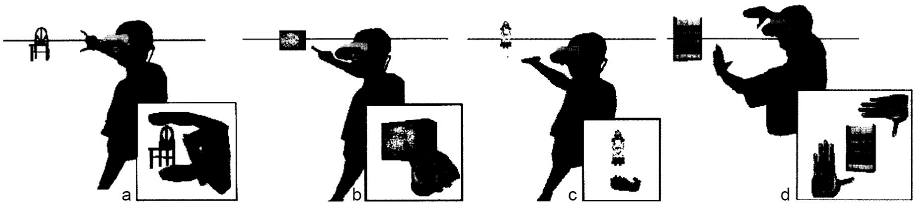
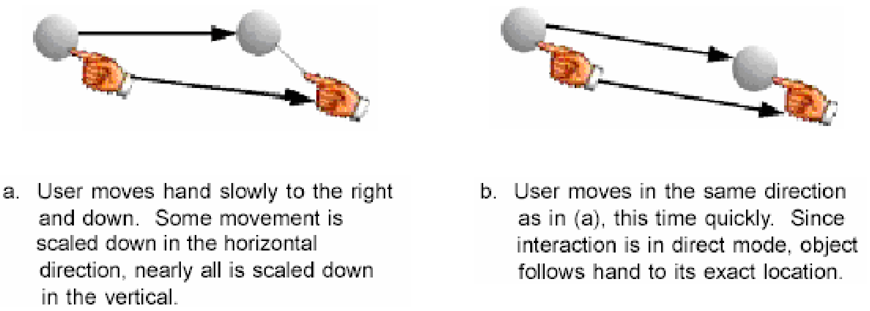


Figure 19: Head Crusher (a), Sticky Finger (b), Lifting Palm (c) and Framing Hands techniques (extracted from [51]).

Pierce et al. [51] presented a set of four selection techniques for immersive virtual environments with head tracking, depicted in Figure 19. With these techniques the user interacts with the 2D projections that 3D objects in the scene make on his image plane. The techniques consist in: positioning the thumb and forefinger around the desired object (named Head Crusher); placing the tip of the index above the desired object (Sticky Finger); flattening the hand and positioning the palm so that it appears to lie below the desired object on the image plane (Lifting Palm); or positioning both hands to form the two corners of a frame in the 2D image surrounding the object to be selected (Framing Hands). This way the user can interact with virtual objects placed at any distance.

To overcome the lack of precision with object positioning techniques in immersive virtual environments, Frees et al. [17] proposed PRISM (Precise and Rapid Interaction through Scaled Manipulation) technique. In contrast to techniques like Go-Go, which scale up hand movement to allow long distance manipulation, PRISM 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, as exemplified in Figure 20. 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



a. User moves hand slowly to the right and down. Some movement is scaled down in the horizontal direction, nearly all is scaled down in the vertical.

b. User moves in the same direction as in (a), this time quickly. Since interaction is in direct mode, object follows hand to its exact location.

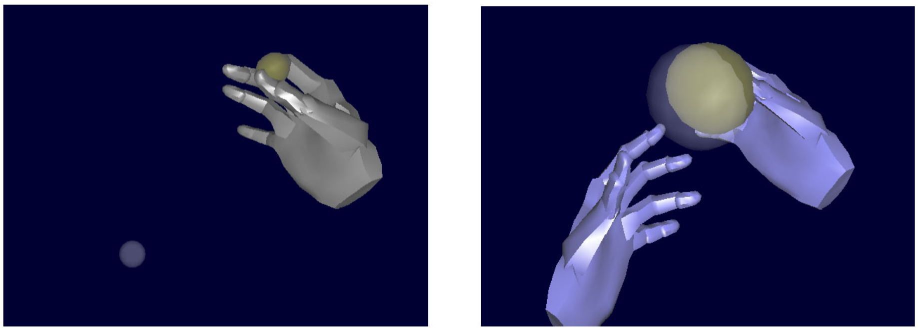
Figure 20: Scaled down movement with the PRISM technique (extracted from [17]).

object and the hand. Using PRISM, the user is always in complete control of the position of the object being manipulated (in contrast to gravity and snapping techniques). User evaluation's results show faster performance and higher user preference for PRISM over a traditional direct approach.

The authors later extended the previous work, by adding support in PRISM for object rotation, which uses the angular speed of the hand [18]. They also present how their approach can be useful for faster object selection using a 3D cursor, either for out-of-reach objects using a smoothed ray-cast approach or for cluttered environments, such as the Worlds in Miniature approach [58].

Moehring et al. [44] presented a study that compares finger-based interaction to controller-based interaction in a CAVE as well as in HMD, for car models' exploration. The authors focused on interaction tasks within reach of the users' arms and hands and explored several feedback methods including visual, pressure based tactile and vibrotactile feedback. Results suggest that controller-based interaction is often faster and more robust, since the button-based selection provides very clear feedback on interaction start, stop and status. However, finger based interaction is preferred over controller-based interaction for the assessments of various functionalities in a car interior, as the abstract character of indirect metaphors leads to a loss of realism and therefore impairs the judgment of the car interior. Grasping feedback is a requirement to judge grasp status. It is not sufficient to just have an object follow the user's hand motion once it is grasped. While visual feedback alone is mostly sufficient for HMD-applications, tactile feedback significantly improves interaction independent of the display system. Vibrational feedback is considerably stronger than pressure based sensations but can quickly become annoying.

One- and two-handed control techniques for precise positioning of 3D virtual objects in immersive virtual environments were proposed by Noritaka Osawa [50]. For this, the authors propose a position adjustment that consists in a scale factor for slowing hand movement, similar to PRISM [17], and viewpoint adjustment, that automatically approaches the viewpoint to the grabbed point so that the object being manipulated appears larger (Figure 21). To control the adjustments, two techniques are presented. The first uses only one hand and is based on its speed, on the assumption that the user moves the hand slowly when


(a) Normal stage

(b) Precise stage with viewpoint adjustment

Figure 21: Viewpoint adjustment for increased precision, which causes the manipulated object to appear larger (extracted from [50]).

wants to precisely manipulate an object. The other uses the distance between both hands. When the distance between them is small, the adjustments are activated. Through a user evaluation, the position and viewpoint adjustment methods showed improvements for small targets over a base scenario where this adjustments were disabled. Also, their results also showed that the two handed control technique performed better than the one handed.

Aguerreche et al. [1] introduced a 3D interaction technique called 3-Hand Manipulation, for multi-user collaborative manipulation of 3D objects. The 3-Hand Manipulation relies on the use of three manipulation points that can be used simultaneously by three different hands of two or three users. The three translation motions of the manipulation points can fully determine the resulting 6 DOF motion of the manipulated object. When a hand is close enough to the object to manipulate, ray-casting from the hand gives an intersection point with the object. This point is called a manipulation point. A rubber band is drawn between a hand and its manipulation point to avoid ambiguity concerning its owner and to display the distance between the hand and the manipulation point. It is elastic and its color varies according to the distance between the hand and the manipulation point. The authors point out that a possible solution for implementation their technique is to use three point-to-point constraints of a physics engine.

Inspired by the previous work, Nguyen et al. [48] proposed a widget consisting of four manipulation points attached to objects, called 3-Point++ tool, which including three handle points, forming a triangle, and their barycenter. With it, users can control and adjust the position of objects. By moving the manipulation points, the position and the orientation of the object are controlled. The barycenter can be used for approximate positioning to control the object directly without constraint, while the three handle points are used for precise positioning. For this, the barycenter has 6-DOF, while the three handle points have only 3-DOF. If one handle point is manipulated, the object is rotated around an axis created by the two other handle points, as shown in Figure 22. If two handle points are manipulated at the same time, the object is rotated around the third handle point. An evaluation was carried out comparing the

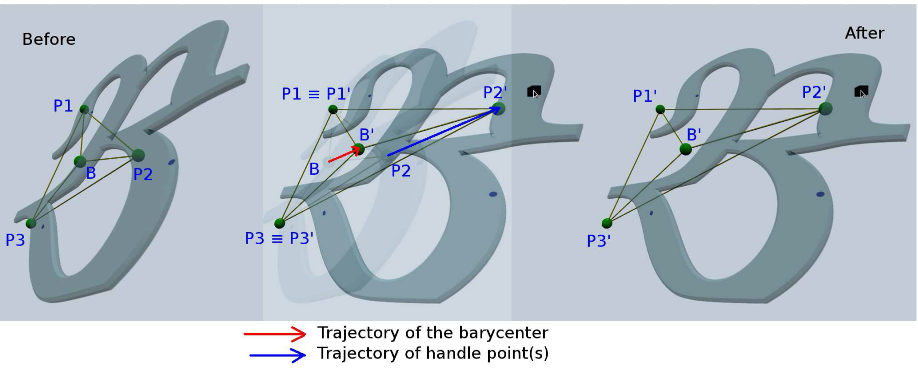


Figure 22: 3-Point++ tool: moving the handle point P2 causes the object to rotate around an axis created by the other two handle points P1 and P3 (extracted from [48]).

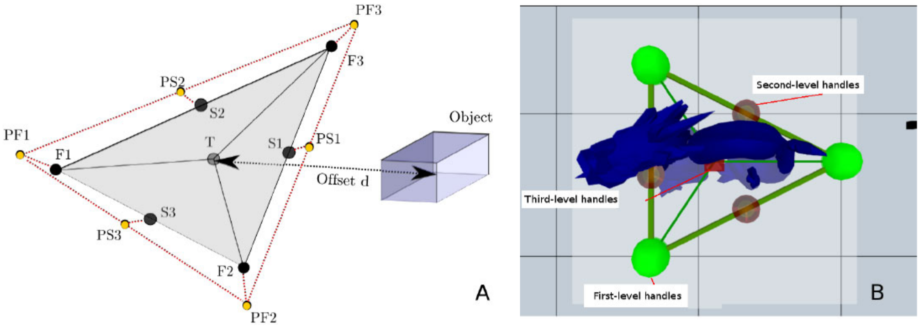


Figure 23: A: Set of seven points of the 7-Handle tool. B: Implementation of the 7-Handle tool (extracted from [49]).

3-Point++ tool with a well-known technique using a 3D cursor to control an object directly with 6-DOF. The 3-Point++ technique had worst results due to its complexity.

Extending their previous work, Nguyen et al. [49] presented the 7-Handle manipulation technique. This technique consists of triangle shaped widget with seven points, depicted in Figure 23. Three points called first-level handles, are the three vertices of the triangle, which act similarly to the 3-Point++ tool. The second-level handles are positioned at the midpoints of the three sides of the triangle and are used to control its two adjacent first-level handles. The last point, the third-level handle is positioned at the centroid of the three first-level handles and can be used as a direct manipulation tool with 6 DOF. Results of a user evaluation showed that the 7-Handle technique is only better suited than the traditional direct 6-DOF approach for manipulating large objects (side bigger than 1.5 meters).

Mine et al. converted the desktop application SketchUp into a virtual reality application: VR SketchUp [41, 42]. Their objective was to develop interaction techniques that can run across a spectrum of displays, ranging from the desktop, to head-mounted displays to large CAVE environments, minimizing energy while maximizing comfort. For this, they built a hybrid controller that collocates a touch display and physical buttons, through a 6 DOF a tracked smartphone attached to a handheld controller. 3D spatial input was used to achieve a coarse

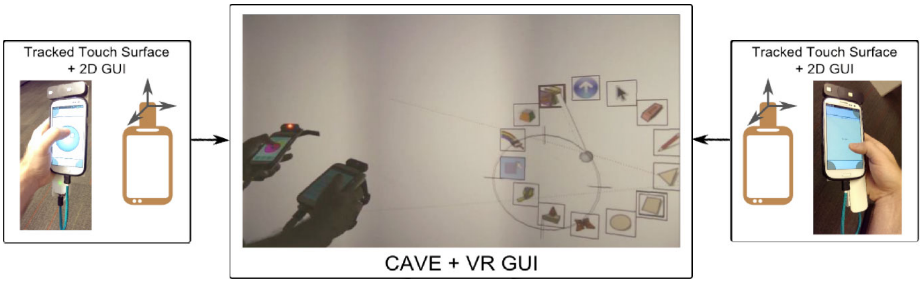


Figure 24: VR SketchUp Interface. Middle: floating VR GUI, left: non-dominant hand controller, right: dominant hand controller (extracted from [41]).

starting step. On the other hand, 2D touch was used for precision input, such as controlling widgets, define constraints and specific values for transformations, and giving numeric or textual input. To manipulate objects, the authors present three alternatives: direct 6-DOF manipulation, where scaling of the object can be achieved using bi-manual interaction and DOF constraints, rotational axes, and special behavior such as position-only manipulation are specified using the touchscreen interface; image plane interactions, where movement of the user's hand within their field of view is mapped to screen space interactions; and track-pad interaction, where the user manipulates objects via a touchpad widget on the touch screen to emulate mouse interactions within the user's screen space. Although the authors focused in several kinds of displays, resorting to imagery on the smartphone screen may not work well in conjunction with HMDs. However, some interactions on the touch surface were designed to not require the user to have to look down at them, such as menu navigation, which is represented by floating graphical elements in the VE.

2.9 Discussion

The main focus of all presented research works is 3D virtual object manipulations, however each one has its own features. We applied the taxonomy presented in Section 2.3 to the main techniques mentioned above, which is shown in Table 2.9. The techniques are listed in the order they were mentioned. The minimum number of simultaneously controlled DOFs are best scenarios. For instance, techniques such as the Handle Box [27] allow movements along a vertical axis to be independently controlled, while translations along the other two axes are simultaneous. In the enhanced precision column we refer to whether it has been a concern for the authors to improve precision over direct mapping between input and output variations.

Mouse based techniques [27, 13, 56], since the user controls a cursor on the screen, are all indirect with a 2D input space. Although it is possible to manipulate virtual objects without widgets using the mouse, the great majority of these kind of interfaces rely on them. Indeed, common commercial applications for creating or editing 3D virtual models still resort to the same techniques presented more than twenty years ago. These widgets allow users to select both a transformation they wish to apply, and the axis they want to apply it along. Since the input has only 2 DOF, existing techniques focus on reducing the simultaneous DOF being controlled.

While 2D interaction has found easy-to-use *de facto* standards for multi-touch devices, adapting these interaction techniques to manipulate 3D objects is not trivial in that it, likewise mouse based interactions, requires mapping 2D input subspaces to a 3D virtual world. However, and distinctly from mouse techniques, these devices allow users to directly touch objects displayed. Trying to create more natural interactions, researches initially proposed techniques for controlling several DOFs at the same time [22, 23, 53]. Nonetheless, reduction of simultaneous DOFs controlled have been suggested and followed by several authors [36, 37]. Thus, techniques that allow manipulations with high DOFs, but with no more than one controlled at each moment, have been later proposed. To clearly and undoubtedly select the transformation and axis, researches turned once again to virtual widgets [11, 40, 6], which evaluations' results suggest to improve users' performance. Even when interacting with stereoscopic imagery

Technique	**Display Properties**			**Input Properties**	**Degrees-of-Freedom**				**Widgets**	**Enhanced Precision**
	Perceived Space	Actual Space	Viewpoint Correlation	Input Space	Total	Trans.	Rot.	Scale		
Handle Box [27]	2D	2D	None	2D Indirect	4 / 1	3 / 1	1 / 1	-	Yes	-
Virtual Handles [13]	2D	2D	None	2D Indirect	9 / 1	3 / 1	3 / 1	3 / 1	Yes	-
Arcball [56]	2D	2D	None	2D Indirect	3 / 1	-	3 / 1	-	Yes	-
Shallow-depth [22]	2D	2D	None	2D Direct	6 / 4	3 / 2	3 / 1	-	-	-
Sticky Fingers [23]	2D	2D	None	2D Direct	6 / 2	3 / 2	3 / 1	-	-	-
Screen-Space [53]	2D	2D	None	2D Direct	6 / 2	3 / 2	3 / 1	-	-	-
Z-Technique [36]	2D	2D	None	2D Direct	3 / 2	3 / 2	-	-	-	-
DS3 [37]	2D	2D	None	2D Direct	6 / 2	3 / 2	3 / 1	-	-	-
tBox [11]	2D	2D	None	2D Indirect	9 / 1	3 / 1	3 / 1	3 / 1	Yes	-
LTouchIt (translation) [40]	2D	2D	None	2D Direct	3 / 2	3 / 2	-	-	-	-
LTouchIt (rotation) [40]	2D	2D	None	2D Indirect	3 / 1	-	3 / 1	-	Yes	-
GimbalBox [6]	2D	2D	None	2D Indirect	6 / 1	3 / 2	3 / 1	-	Yes	-
Balloon Selection [5]	3D	2D	High	2D Indirect	4 / 2	3 / 2	1 / 1	-	-	-
Triangle Cursor [59]	3D	2D	High	2D Indirect	4 / 4	3 / 3	1 / 1	-	-	-
Toucheo [20]	3D	Heads-up	High	2D Indirect	9 / 1	3 / 1	3 / 1	3 / 1	Yes	-
In the Air [25]	2D	2D	None	3D Indirect	4 / 4	3 / 3	1 / 1	-	-	-
Holodesk [26]	3D	Heads-up	High	3D Direct	6 / 6	3 / 3	3 / 3	-	-	-
Air-TRS [3]	3D	2D	High	3D Direct	7 / 3	3 / 3	3 / 3	1 / 1	-	-
Handle bar [57]	2D	2D	None	3D Indirect	7 / 7	3 / 3	3 / 3	1 / 1	-	-
6D Hands [62]	2D	2D	None	3D Indirect	6 / 6	3 / 3	3 / 3	-	-	-
PRISM [17]	3D	Heads-up	Total	3D Direct	3 / 3	3 / 3	-	-	-	Yes
PRISM w/ Rotations [18]	3D	Heads-up	Total	3D Direct	6 / 6	3 / 3	-	-	-	Yes
Viewpoint Adjustment [50]	3D	Heads-up	Total	3D Direct	3 / 3	3 / 3	-	-	-	Yes
3-Point+++ [48]	3D	Heads-up	Total	3D Indirect	6 / 1	3 / 3	3 / 1	-	Yes	-
7 Handle [49]	3D	Heads-up	Total	3D Indirect	6 / 1	3 / 3	3 / 1	-	Yes	-

Table 1: Classification of the most relevant techniques for manipulating 3D virtual objects.
(Degrees-of-Freedom: Total DOFs available / Minimum number of explicitly constrained DOFs controlled simultaneously.)

above tabletops, the only technique that allow full 9 DOF manipulations [20] resorts to widgets.

Having an input with higher DOFs, most current mid-air approaches for 3D virtual object manipulation try to mimic physical world interactions [25, 26, 3, 57, 62]. Having been realized that human accuracy is limited, sometimes aggravated by input devices' resolution, efforts have been carried out to alleviate this. While for mouse and keyboard interfaces it is possible to directly input values, this can be challenging in IVE and could hinder mid-air interactions fluidity. To improve manipulations' accuracy, authors already tried to either scale down hand motions [17, 18] or move the viewpoint closer to object being manipulated [50], but without regard to DOF separation. On the other hand, approaches based on virtual widgets have been proposed [48, 49], to limit simultaneous transformations. However, these techniques do not have promising results: 3-Point++ [48] performs worst than direct manipulation with 6 DOF and 7 Handle [49] is only suited for very large objects. One explanation for these poor results might be that the widgets of these approaches 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.

As for display properties, there are differences between HMDs and stereoscopic tabletops. A number of interface issues arise with semi-immersive displays, as stated by Bowman et al. [9]. For instance, because users can see their own hands in front of the display, they can inadvertently block out virtual objects that should appear to be closer than their hands. Within IVE, since users do not see the position and orientation of their bodies and limbs, solutions must be explored to increase users' proprioception [43].

In conclusion, there is no mid-air technique, as far as we know, that offers 9 DOF manipulations, let alone allow changes in only 1 DOF at a time. To provide DOF separation, widgets became common in mouse and touch based solutions, and are now appearing in mid-air. Nevertheless, they have yet to be improved in order to be a viable alternative to traditional WIMP interfaces. The lack of precision in mid-air interactions needs also to be taken into account when considering 3D object manipulation in these kind of environments. While techniques that move away from direct 6 DOF manipulations are less natural, they can avoid unwanted side effects of replicating the physical world exactly, and can provide users with enhanced abilities that may improve performance and usability [10].

3 Research Proposal

Immersive virtual environments are very appealing not only for entertainment purposes, but also for engineering and architecture, among others. In this context, the capability to precisely move, rotate and scale 3D virtual elements assumes a great role.

3.1 Problem

The most popular approach for 3D object manipulations within immersive virtual environments consists in simulating those with physical objects: grab and directly move and rotate with simultaneous 6 DOF. As it has been previously identified, this approach lacks precision, since human dexterity in mid-air is quite limited. While with traditional WIMP-based interfaces it is possible to specify exact values to objects' transformations, mid-air approaches are only suitable for coarse actions. However, traditional displays can hinder perception of 3D content, whereas this perception can be greatly enhanced with immersive displays, which offer a natural stereoscopic view upon the virtual world.

Due to the aforementioned limitations, precise manipulations of virtual objects are mainly performed in desktop setups, with techniques still very similar to those proposed more than twenty years ago. Despite the many advances in virtual reality we have been witnessing in that period, IVE are almost solely used for either entertainment or pure visualization purposes. For example, when engaged in engineering projects, users resort to immersive setups to have a better understanding of the virtual content, take notes of what needs to be modified, and then perform the identified modifications back in the desktop computer with a CAD tool.

With the presented motivation, our research question can be put as follows:

How can users' accuracy with spacial manipulations be improved, when immersed in a virtual reality environment, in order to achieve millimetric precision?

3.2 Hypotheses

A path that can be followed to attain more precise and controlled 3D object manipulations is to follow DOF separation. DOF separation first appeared to overcome the mapping difficulties between 2D input and desired 3D output, and it has been shown that controlling only 1 DOF at a time can be beneficial. So, if mid-air input has even more DOF, we reckon that it can also benefit from highly controlled manipulations. Although we will lose the naturalness of direct 6 DOF interactions, with an hyper-natural approach we might extend human capabilities in ways that are not possible in the physical world, which have been shown to be helpful in specific scenarios.

A common way to achieve is DOF separation is through virtual widgets, which allow users to select specific transformations and axes. Indeed, as discussed in Section 2.9, virtual widgets for 3D manipulation became a *de facto* standard in mouse-based 3D user interfaces, are becoming ever more common in multi-touch devices, and are even been proposed for mid-air interactions. While in the first two results suggest that this approach is successful, in mid-air there are still challenges to be tackled. Although widgets themselves are not natural, interaction with them can be. Grabbing and moving an handle, for example, can be performed with a direct approach, while restricting its movement using the metaphor of being on rails.

While for pure performance purposes close mapping of input and output DOFs is desirable, as stated by several researchers, this is not true when more accurate positioning is in order. As such, our first hypothesis is:

Hypothesis 1 *Mid-air manipulations with DOF separation through virtual widgets increases users performance in docking tasks that have low error tolerance.*

Albeit widgets can offer single DOF manipulations, they do not solve the problem of limited human dexterity in mid-air. This challenge has been subject of previous research, but a definitive solution is yet to be found. It has been suggested that techniques that enhance precision, such as scaled hand motions, can be an improvement over direct mappings. We believe that virtual widgets will benefit from such techniques, which leads to our second hypothesis:

Hypothesis 2 *Virtual widgets in mid-air can be combined with precision enhancing techniques to augment user accuracy in 3D manipulation tasks.*

When restricting manipulated DOFs, widget based approaches allow users to select an axis to apply the transformation. This axis is usually from the word or object frame. Some techniques that follow DOF separation without widgets allow users to specify custom transformation axis. We think that this can contribute for faster and more straightforward manipulations, since users can promptly specify a custom direction for translation, for example, instead of performing several transformations along different axes. As such, our third hypothesis can be put as follows:

Hypothesis 3 *The possibility to specify custom transformation axes leads to faster mid-air manipulations.*

3.3 Objectives

The main objective of our research is to increase users precision in IVE, in order to be a viable alternative to traditional CAD solutions for moving, rotating and scaling virtual objects. Considering the our hypothesis, we can define a set of sub-objectives:

1. Identify which existing mid-air techniques for 3D virtual object manipulation perform better and appeal most to users;
2. Study whereas mid-air techniques identified in #1 are suited both for SIVEs and IVEs, since not all were conceived for the same environments;
3. Develop a 3D manipulation technique based on virtual widgets that offers constrained manipulation with 1 DOF at a time;
4. Improve upon the technique developed in #3 by adding precision enhancing features, such as scaled hand motions, and adequate feedback, to allow exact manipulations;

5. Create an approach to specify custom transformation axis, in addition to the usual object and/or world frames;
6. Validate the complete mid-air solution against a commercial WIMP-based software, for example Sketch Up, through user evaluation with precision demanding docking tasks.

4 Preliminary Study of Mid-Air Manipulations

Several approaches for manipulating 3D virtual objects in mid-air have been proposed in the last few years. To assess with which users can perform better and be satisfied with, we implemented and evaluated five different techniques. This study was publish in the IEEE Symposium on 3D User Interfaces (3DUI) 2014. Please refer to this article [39] for a more detailed description of the evaluation and its results.

4.1 Implemented Techniques

We implemented five different interaction techniques for object manipulation, based on the literature, using a stereoscopic tabletop. Four of these use mid-air interactions, both direct and indirect, and one is solely touch-based, which acted as baseline. All implemented techniques provide 7 DOF, three for translation, three for rotation, and a uniform scale.

6-DOF Hand To mimic interactions with physical objects, as closely as possible, we use all 6 DOF information provided by the tracker (3 DOF for position and 3 DOF for orientation). With this technique, a person grabs the object directly with one hand, typically the dominant hand. All the hand movements are directly applied to the object, as depicted in Figure 25.a. Dragging the object in space moves it in three dimensions, and the wrist rotation controls object rotation. Grabbing somewhere in space outside the object with the non-dominant hand and varying the distance to dominant hand, uniformly scales the object. The grabbed point in the object will remain the center of all transformations, during the entire manipulation, until the object is released.

3-DOF Hand In this technique we divided translations and rotations by both hands (Figure 25.b) to prevent unwanted manipulations, as suggested in [47]. After grabbing the object with one hand, the user can translate it by moving that hand. The rotation is achieved by rotating the wrist corresponding to the other hand, by grabbing somewhere in space, while keeping the object selected with the first hand. Similarly to the 6-DOF technique, varying the distance between hands will uniformly scale the object, while the grabbed point in the object will remain as the center of all transformations.

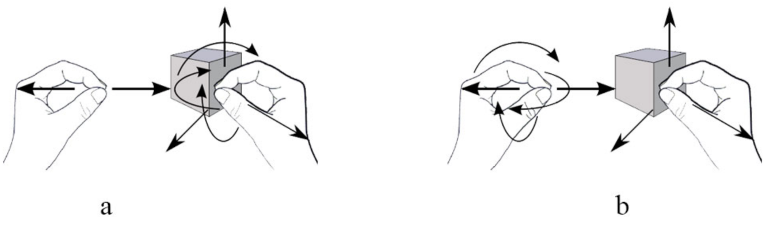


Figure 25: Implemented techniques: 6-DOF Hand (a) and 3-DOF Hand (b).

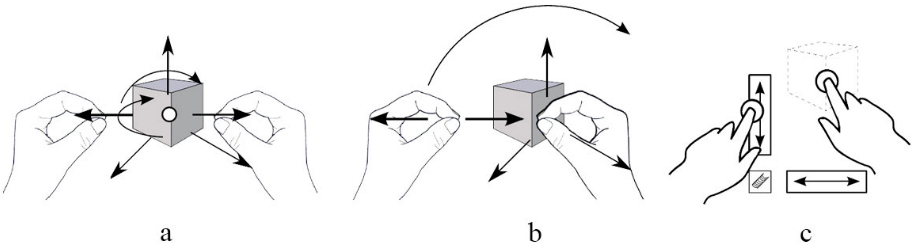


Figure 26: Implemented techniques: Handle-bar (a), Air TRS (b) and Touch TRS + Widgets (c).

Handle-Bar Following the work of Song et al. [57], we implemented the Handle-bar metaphor. This approach mimics a physical bimanual handle-bar, commonly used, for example, to roast a turkey. In this technique, we use the middle point of each hand, after the grab, to manipulate virtual objects (Figure 26.a). The user can translate the object by moving both hands in the same direction and rotate it by moving the hands in different directions. Changing the distance between hands evenly scales the object.

Air TRS Since the coordinates of each hand in space are known, the two-point Translate-Rotate-Scale (TRS) can be extended to the third dimension. We consider user hands as two points and use them in a similar fashion to the Two-Point Rotation and Translation with scale [24], as illustrated in Figure 26.b. The hand that grabs the object moves it. The other hand, after pinching somewhere in space, allows the user to manipulate the object rotation and scale. These two transformations are centered in the object pinched point. The rotation angle is defined by the variation in the position of one hand relatively to the other. For scaling, the distance between both hands is used.

Touch TRS + Widgets Although only allowing indirect manipulations of virtual objects in the three-dimensional space above the surface, multi-touch is, nowadays, a commonly used input method, present in our everyday life. Our touch technique uses the TRS algorithm combined with three widgets, depicted in Figure 26.c, to achieve 7 DOF manipulations. This implementation, provides DOF separation, allowing the user to translate virtual objects in a plane parallel to the surface, by touching directly below it with one finger and dragging. While this touch is active, three widgets appear to the left or to the right of the touch, depending on which hand the finger corresponds to. By using a second finger outside of any widget, the user can either rotate around a vertical axis or scale the object. If the second touch is on one of the three widgets, the user will be able to rotate around one of the two axis parallel to the surface, following a rod metaphor [20], or to change the height of the object, similarly to the balloon metaphor [5].

4.2 User Evaluation

To validate the techniques described above, we carried out a user evaluation with twelve participants. We aimed at identify which were the more natural

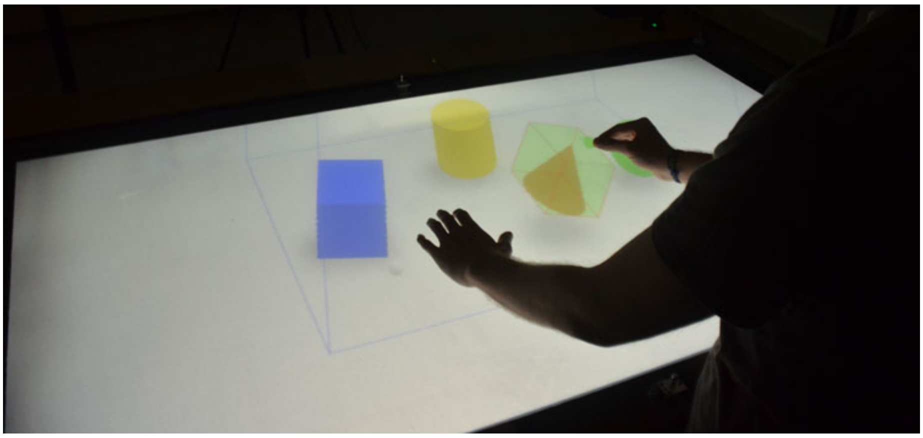


Figure 27: Participant manipulating objects in our semi-immersive setup.

and easier to use and also which were preferred by participants. To accomplish that, we tested our techniques in a practical task scenario. We developed a set of three different tasks with increasing difficulty. The virtual environment of all tasks had no gravity and it was collision-free. This experiment was performed in our laboratory with a controlled environment, using the semi-immersive setup depicted in Figure 27.

We devised three tasks for user evaluation. These were easy to understand and followed a wooden toy metaphor, as the peg-in-hole task in Martinet et al. [37], which requires subjects to fit an object inside a hole in other object. To provide incremental difficulty between tasks, we started with an easy task, followed by an intermediate one and ended up with a more complex effort.

The first task consisted only in translations on a two-dimensional plane parallel to the surface, requiring neither height translation, nor rotation or scale, in order to be fairly accessible to all participants. The second task required translations in all three axes and also scale transformations, but did not require any rotations. For the third task we asked participants not only to translate and scale, but also to rotate the object.

4.3 Results

We registered participant preferences, meaningful actions, and comments throughout every session. Finally, we performed a quantitative and qualitative analysis, using statistical methods to assess which results were significant.

Participants agreed that the 6-DOF Hand approach was more natural to use, since it reproduces direct interactions with physical objects. Results also showed that the Handle-Bar [57] solution was as fast as the 6-DOF Hand. Additionally, we observed that our approach to directly controlling 6-DOF with the dominant hand created unwanted occlusions, a consequence of stereoscopic displays already identified in the literature [9], that did not affect the Handle-Bar.

The main conclusion of this study is that, concerning virtual objects lying above a stereoscopic tabletop, mid-air manipulations that have a greater resemblance to interactions in the physical world appeal more to users. These techniques also allow non-experienced users to readily achieve good performance.

5 Work Plan

In this section, the work plan for our research is presented. Figure 28 shows the research work already done and the plan for the remaining period.

5.1 Past: 2013 / 2014

Since March 2013, the start date of my PhD, I completed the curricular component of the PhD, composed of five courses. During this time we made a first review of the state-of-the-art regarding mid-air manipulations and conducted a study of which are more suited for SIVE, using stereoscopic tabletops with non-intrusive solutions for head and hands tracking that we built.

In the same period, I've been involved in several research projects that did not were directly related to this PhD thesis, but contributed for several experiences with virtual environments that comprised different tracking solutions and immersive displays. For user tracking, I explored Optitrack solution, with IR cameras and markers, and other non-invasive solutions. In the later kind, I looked into: Nimble VR[13], which is now owned by Oculus and no is longer provided; both Microsoft Kinect v1 and v2; Leap Motion; PNI Space Mouse; and accelerometers and gyroscopes from Android smartphones and gaming controllers.

Regarding immersive displays we started with custom made semi-immserive tabletop based setups, as already mentioned. We built two of this kind: one with back projection and laser plane technology, and one that uses off-the-shelf components such as a 3D TV and a IR frame. In the meantime Oculus Rift appeared and the interest on HMD escalated very quickly. We now own both development kits 1 and 2, which I experienced with, both with wired and custom made wireless configurations.

I have also been using and exploring several frameworks for developing software for virtual environments. I started with OpenSG[14], which is now deprecated but supported stereoscopy with custom projection frustum that can be used in tabletops. Then tried G3D[15], that could generate appealing visual effects but required low level programming. Finally, I've been using Unity3D for developing and deploying VR applications to Windows and Android devices. Unity3D is a game engine that offers very quick prototype development, even for research purposes in virtual environments, and it is now the main tool used in our research group and the platform we're going to use in the following development stages of this thesis.

In late 2014, and after focusing more on 3D spatial manipulations, I started a more extensive bibliographic review and the writing of this thesis proposal.

5.2 Present: 2015

In 2015, I continued with the bibliographic review and finished writing this proposal. Simultaneously, we've been studying how the best techniques from our previous evaluation perform in IVE, as opposed to SIVE, focusing on the relation between task completion time and precision.

---

[13] Nimble VR: `nimblevr.com`

[14] OpenSG: `www.opensg.org`

[15] G3D: `g3d.sourceforge.net`

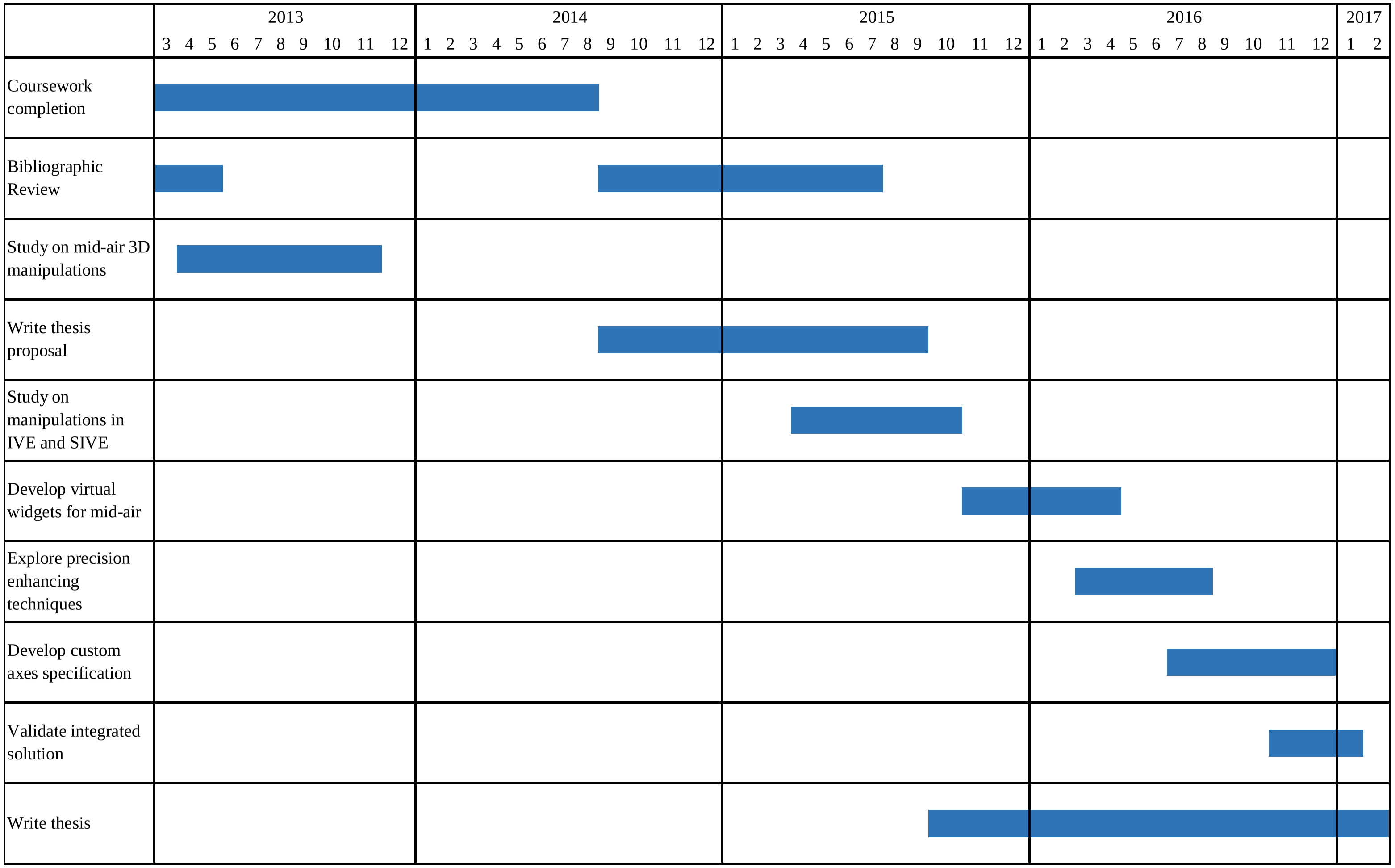


Figure 28: Gantt chart representing the proposed thesis work plan.

Still this year, we intend to start the development of mid-air interaction techniques based on virtual handles. Further user evaluation will suggest if this explicit DOF separation is suited for faster and precise object manipulations, helping us to validate our first hypothesis.

Due to limitations in existing tracking solutions, we are now developing a custom sensing hardware to be combined with the data provided from Microsoft Kinect v2, to track the whole user skeleton and precise hand orientation. The device is wireless, will be shaped like a hand clip, and it will contain a pressure area to detect grab intentions, and an IMU sensor, which combines accelerometers and gyroscope. This will be the main source of user input in our following prototypes.

5.3 Future: 2016 / 2017

Starting in 2016, we will focus in validating our two remaining hypotheses. We will implement precision enhancing techniques in the virtual handles solution, and assess if it indeed increases user accuracy. We will also conceive an approach to allow users to specify custom transformation axis, believing that it will lead to faster, yet still constrained, manipulations.

After testing all our hypotheses, we will compare an integrated solution for mid-air manipulations, which covers DOF separation, user defined transformation axes and precision enhancing techniques, against a commercial WIMP-based software.

During development and evaluation stages of this research, I will be writing the thesis document. After, I will prepare the presentation and defend the work done in this PhD thesis. We also plan to submit the results attained in every stage to international scientific conferences and/or journals.

References

[1] L. Aguerreche, T. Duval, and A. Lécuyer. 3-hand manipulation of virtual objects. In *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pages 153–156. Eurographics Association, 2009.

[2] B. R. D. Araùjo, G. Casiez, and J. A. Jorge. Mockup builder: direct 3d modeling on and above the surface in a continuous interaction space. In *Proceedings of GI '12*, pages 173–180. Canadian Information Processing Society, 2012.

[3] B. R. D. Araújo, G. Casiez, J. A. Jorge, and M. Hachet. Mockup builder: 3d modeling on and above the surface. *Computers & Graphics*, 37(3):165 – 178, 2013.

[4] A. S. Azevedo, J. Jorge, and P. Campos. Combining eeg data with place and plausibility responses as an approach to measuring presence in outdoor virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 23(4):354–368, 2014.

[5] H. Benko and S. K. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *3DUI*, page 22, 2007.

[6] B. Bollensdorff, U. Hahne, and M. Alexa. The effect of perspective projection in multi-touch 3d interaction. In *Proceedings of Graphics Interface 2012*, GI '12, pages 165–172, Toronto, Ont., Canada, Canada, 2012. Canadian Information Processing Society.

[7] D. Bowman, R. P. McMahan, et al. Virtual reality: how much immersion is enough? *Computer*, 40(7):36–43, 2007.

[8] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 35–ff. ACM, 1997.

[9] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. Poupyrev. An introduction to 3-d user interface design. *Presence: Teleoperators and virtual environments*, 10(1):96–108, 2001.

[10] D. A. Bowman, R. P. McMahan, and E. D. Ragan. Questioning naturalism in 3d user interfaces. *Communications of the ACM*, 55(9):78–88, 2012.

[11] A. Cohé, F. Dècle, and M. Hachet. tbox: a 3d transformation widget designed for touch-screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3005–3008, New York, NY, USA, 2011. ACM.

[12] A. Cohé and M. Hachet. Understanding user gestures for manipulating 3D objects from touchscreen inputs. In *Graphics Interface*, pages 157–164, Toronto, Canada, May 2012. ACM.

[13] B. D. Conner, S. S. Snibbe, K. P. Herndon, D. C. Robbins, R. C. Zeleznik, and A. Van Dam. Three-dimensional widgets. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 183–188. ACM, 1992.

[14] L. D. Cutler, B. Fröhlich, and P. Hanrahan. Two-handed direct manipulation on the responsive workbench. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 107–114. ACM, 1997.

[15] F. Daiber, E. Falk, and A. Krüger. Balloon selection revisited: multi-touch selection techniques for stereoscopic data. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 441–444, New York, NY, USA, 2012. ACM.

[16] J. Feng, I. Cho, and Z. Wartell. Comparison of device-based, one and two-handed 7dof manipulation techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, pages 2–9. ACM, 2015.

[17] S. Frees and G. D. Kessler. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, pages 99–106. IEEE, 2005.

[18] S. Frees, G. D. Kessler, and E. Kay. Prism interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(1):2, 2007.

[19] T. Grossman and D. Wigdor. Going deeper: a taxonomy of 3d on the tabletop. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on*, pages 137–144. IEEE, 2007.

[20] M. Hachet, B. Bossavit, A. Cohé, and J.-B. de la Rivière. Toucheo: multitouch and stereo combined in a seamless workspace. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 587–592, New York, NY, USA, 2011. ACM.

[21] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118. ACM, 2005.

[22] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1147–1156, New York, NY, USA, 2007. ACM.

[23] M. Hancock, T. ten Cate, and S. Carpendale. Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 133–140, New York, NY, USA, 2009. ACM.

[24] M. Hancock, F. Vernier, D. Wigdor, S. Carpendale, and C. Shen. Rotation and translation mechanisms for tabletop interaction. In *Horizontal Interactive Human-Computer Systems, 2006. TableTop 2006. First IEEE International Workshop on*, pages 8 pp.–, 2006.

[25] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 139–148, New York, NY, USA, 2009. ACM.

[26] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. Holodesk: direct 3d interactions with a situated see-through display. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2421–2430, New York, NY, USA, 2012. ACM.

[27] S. Houde. Iterative design of an interface for easy 3-d direct manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 135–142, New York, NY, USA, 1992. ACM.

[28] C. E. Hughes, L. Zhang, J. P. Schulze, E. Edelstein, and E. Macagno. Cavecad: Architectural design in the cave. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 193–194. IEEE, 2013.

[29] D. F. Keefe, D. A. Feliz, T. Moscovich, D. H. Laidlaw, and J. J. LaViola Jr. Cavepainting: a fully immersive 3d artistic medium and interactive experience. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, pages 85–93. ACM, 2001.

[30] T. Kim and J. Park. 3d object manipulation using virtual handles with a grabbing metaphor. *IEEE Computer Graphics and Applications*, (3):30–38, 2014.

[31] K. Kin, M. Agrawala, and T. DeRose. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *Proceedings of Graphics interface 2009*, pages 119–124. Canadian Information Processing Society, 2009.

[32] K. Kin, T. Miller, B. Bollensdorff, T. DeRose, B. Hartmann, and M. Agrawala. Eden: a professional multitouch tool for constructing virtual organic environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1343–1352. ACM, 2011.

[33] S. Knoedel and M. Hachet. Multi-touch rst in 2d and 3d spaces: Studying the impact of directness on user performance. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces*, 3DUI '11, pages 75–78, Washington, DC, USA, 2011. IEEE Computer Society.

[34] R. Kruger, S. Carpendale, S. D. Scott, and A. Tang. Fluid integration of rotation and translation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 601–610. ACM, 2005.

[35] N. Marquardt, R. Jota, S. Greenberg, and J. A. Jorge. The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part III*, INTERACT'11, pages 461–476, Berlin, Heidelberg, 2011. Springer-Verlag.

[36] A. Martinet, G. Casiez, and L. Grisoni. The design and evaluation of 3d positioning techniques for multi-touch displays. In *3D User Interfaces (3DUI), 2010 IEEE Symposium on*, pages 115–118, 2010.

[37] A. Martinet, G. Casiez, and L. Grisoni. 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, pages 111–118, New York, NY, USA, 2010. ACM.

[38] D. Mendes and A. Ferreira. Evaluation of 3d object manipulation on multi-touch surfaces using unconstrained viewing angles. In *Human-Computer Interaction–INTERACT 2011*, pages 523–526. Springer, 2011.

[39] D. Mendes, F. Fonseca, B. Araujo, A. Ferreira, and J. Jorge. Mid-air interactions above stereoscopic interactive tables. In *3D User Interfaces (3DUI), 2014 IEEE Symposium on*, pages 3–10. IEEE, 2014.

[40] D. Mendes, P. Lopes, and A. Ferreira. Hands-on interactive tabletop lego application. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE '11, pages 19:1–19:8, New York, NY, USA, 2011. ACM.

[41] M. Mine, A. Yoganandan, and D. Coffey. Making vr work: building a real-world immersive modeling application in the virtual world. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pages 80–89. ACM, 2014.

[42] M. Mine, A. Yoganandan, and D. Coffey. Principles, interactions and devices for real-world immersive modeling. *Computers & Graphics*, 48:84–98, 2015.

[43] M. R. Mine, F. P. Brooks Jr, and C. H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 19–26. ACM Press/Addison-Wesley Publishing Co., 1997.

[44] M. Moehring and B. Froehlich. Effective manipulation of virtual objects within arm's reach. In *Virtual Reality Conference (VR), 2011 IEEE*, pages 131–138. IEEE, 2011.

[45] M. Möllers, P. Zimmer, and J. Borchers. Direct manipulation and the third dimension: co-planar dragging on 3d displays. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, ITS '12, pages 11–20, New York, NY, USA, 2012. ACM.

[46] T. S. Mujber, T. Szecsi, and M. S. Hashmi. Virtual reality applications in manufacturing process simulation. *Journal of materials processing technology*, 155:1834–1838, 2004.

[47] M. A. Nacenta, P. Baudisch, H. Benko, and A. Wilson. Separability of spatial manipulations in multi-touch interfaces. In *Proceedings of Graphics Interface 2009*, GI '09, pages 175–182, Toronto, Ont., Canada, Canada, 2009. Canadian Information Processing Society.

[48] T. T. H. Nguyen and T. Duval. Poster: 3-point++: A new technique for 3d manipulation of virtual objects. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 165–166. IEEE, 2013.

[49] T. T. H. Nguyen, T. Duval, and C. Pontonnier. A new direct manipulation technique for immersive 3d virtual environments. In *ICAT-EGVE 2014: the 24th International Conference on Artificial Reality and Telexistence and the 19th Eurographics Symposium on Virtual Environments*, page 8, 2014.

[50] N. Osawa. Two-handed and one-handed techniques for precise and efficient manipulation in immersive virtual environments. In *Advances in Visual Computing*, pages 987–997. Springer, 2008.

[51] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 39–ff. ACM, 1997.

[52] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 79–80. ACM, 1996.

[53] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 69–78, New York, NY, USA, 2009. ACM.

[54] W. Robinett and R. Holloway. Implementation of flying, scaling and grabbing in virtual worlds. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 189–192. ACM, 1992.

[55] U. Schultheis, J. Jerald, F. Toledo, A. Yoganandan, and P. Mlyniec. Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3d tasks. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*, pages 117–124. IEEE, 2012.

[56] K. Shoemake. Arcball: a user interface for specifying three-dimensional orientation using a mouse. In *Graphics Interface*, volume 92, pages 151–156, 1992.

[57] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1297–1306, New York, NY, USA, 2012. ACM.

[58] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 265–272. ACM Press/Addison-Wesley Publishing Co., 1995.

[59] S. Strothoff, D. Valkov, and K. Hinrichs. Triangle cursor: interactions with objects above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 111–119, New York, NY, USA, 2011. ACM.

[60] A. Van Dam, D. H. Laidlaw, and R. M. Simpson. Experiments in immersive virtual reality for scientific visualization. *Computers & Graphics*, 26(4):535–555, 2002.

[61] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock. Evaluation of docking task performance using mid-air interaction techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, pages 44–52. ACM, 2015.

[62] R. Wang, S. Paris, and J. Popović. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 549–558, New York, NY, USA, 2011. ACM.

[63] R. Y. Wang and J. Popović. Real-time hand-tracking with a color glove. In *ACM SIGGRAPH 2009 papers*, SIGGRAPH '09, pages 63:1–63:8, New York, NY, USA, 2009. ACM.

[64] D. Wigdor and D. Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture.* Elsevier, 2011.