Mid-Air Interactions Above Stereoscopic Interactive Tables

Daniel Mendes*
INESC-ID Lisbon / IST / University of Lisbon
Fernando Fonseca†
INESC-ID Lisbon / IST / University of Lisbon
Bruno Araùjo‡
Inria Lille / INESC-ID Lisbon
Alfredo Ferreira §
INESC-ID Lisbon / IST / University of Lisbon
Joaquim Jorge ¶
INESC-ID Lisbon / IST / University of Lisbon

ABSTRACT

Stereoscopic tabletops offer unique visualization capabilities, enabling users to perceive virtual objects as if they were lying above the surface. While allowing virtual objects to coexist with user actions in the physical world, interaction with these virtual objects above the surface presents interesting challenges. In this paper, we aim to understand which approaches to 3D virtual object manipulations are suited to this scenario. To this end, we implemented five different techniques based on the literature. Four are mid-air techniques, while the remainder relies on multi-touch gestures, which act as a baseline. Our setup combines affordable non-intrusive tracking technologies with a multi-touch stereo tabletop, providing head and hands tracking, to improve both depth perception and seamless interactions above the table. We conducted a user evaluation to find out which technique appealed most to participants. Results suggest that mid-air interactions, combining direct manipulation with six degrees of freedom for the dominant hand, are both more satisfying and efficient than the alternatives tested.

Keywords: 3D virtual object manipulation, stereoscopic environments, interactive tabletops, mid-air interactions

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles, Graphical User Interfaces, Input devices and strategies

1 INTRODUCTION

Perception of virtual environments can be enhanced by combining stereoscopic visualization with head tracking, to increase user immersion. While suitable for distinct environments, ranging from wall-sized displays to mobile devices, this combination has scarcely been explored in interactive tabletops. These scenarios are especially appealing for 3D model exploration and assembly. Recent technological advances made it easier to develop stereoscopic tabletop visualization. Currently, tracking is possible using a single, affordable and non-intrusive depth camera. This solution can be used not only to follow a user's head position, but also to track their limbs, hands or even fingers, unveiling new interaction possibilities, as depicted in Figure 1. Furthermore, using a table as the visualization surface affords extending the mid-air interaction vocabulary by providing a rigid surface for drawing and manipulating objects. In the way the expressive power of sketching and two-handed interactions can be continuously extended to use the full space above the table as proposed by Marquardt et al [20]. Finally, having a stereoscopic display allows users to naturally manipulate three dimensional entities as if they were collocated with their hands and body, extending the two-dimensional surface in very natural ways. Novel approaches to semi-immersive manipulations show promising ways in which these possibilities can be explored [1].

Interacting with stereoscopic environments has been the subject of previous research, using either multi-touch surfaces [11, 28] or wearable devices [2]. While there are robust multi-touch solutions, these are restricted to a two-dimensional space and cannot offer direct interactions above the table. Wearable devices mitigate this problem, but they are both invasive and restrictive. Other researchers [27, 29] propose solutions to interact within a three-dimensional space using a gadget-free tracking system, but do not combine them with stereoscopic systems, using them only as a more powerful, yet still indirect, cursor.

In this paper, we aim to understand which approaches to 3D virtual object manipulations, based on the literature, are better suited to interact with stereoscopic tabletops. We implemented five different techniques, four mid-air and one multi-touch, which was used as a baseline. Our setup combines affordable non-intrusive tracking technologies with a multi-touch stereoscopic tabletop. This provides both head and hand tracking, improving depth perception and enabling walk-up-and-use mid-air interactions. These allow users to approach the table and immediately interact with virtual objects, without the need to wearing intrusive and unpleasant devices or to carry out calibrations. We performed an user evaluation to compare the techniques and to understand users' preference regarding 3D virtual object manipulations in tabletop stereoscopic scenarios.

2 RELATED WORK

Many computer applications require virtual three dimensional object manipulations, such as architectural modeling, virtual model exploration, engineering component design and assembly, among others. Because of this, 3D manipulation has been the focus of intense research. Beyond the traditional WIMP-based approaches, several multi-touch solutions to manipulate 3D objects have been proposed and evaluated over the past few years.

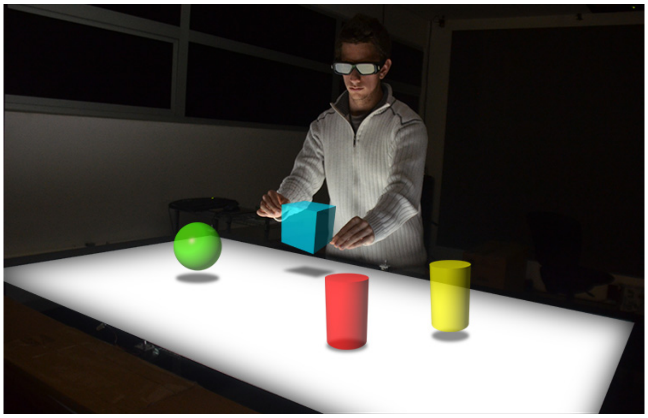

Figure 1: User manipulating virtual objects above the table.

---

*e-mail: danielmendes@ist.utl.pt
†e-mail:fernando.fonseca@ist.utl.pt
‡e-mail:brar@vimmi.inesc-id.pt
§e-mail:alfredo.ferreira@ist.utl.pt
¶e-mail:jaj@inesc-id.pt

IEEE Symposium on 3D User Interfaces 2014
29 - 30 March, Minneapolis, Minnesota, USA
978-1-4799-3624-3/14/$31.00 ©2014 IEEE

2.1 Touch Manipulations

After developing techniques to control six degrees of freedom (DOF) using one, two and three simultaneous touches, Hancock et al. [12] concluded that a 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, adding other manipulation tools [13].

Considering the *de facto* standard for 2D manipulations, the Translate-Rotate-Scale (TRS) or Two-Point Rotation and Translation with scale [14], Reisman et al. [26] 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. 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 [25] and followed in different research works. Martinet et al. [22] introduced the DS3, a 3D manipulation technique based on DOF separation, improving on the Z-technique [21] by adding rotation and scale. DS3 was compared with previous techniques [13, 26]. User evaluation [22] revealed that DOF separation led to better results.

To better understand user gestures for 3D manipulation tasks on multi-touch devices, Cohé et al. [7] 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. [6] 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). LTouchIt [23], although using direct manipulation for translations, also relies on widgets for rotations. Regarding direct versus indirect interactions, Knoedel et al. [18] 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. [3] redesigned older techniques for three-dimensional interactions [17] using multi-touch input. Through a controlled study, their techniques were compared to other approaches well-known in the literature [13, 26]. 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”.

2.2 Manipulations in Stereoscopic Environments

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 [24]. Above the table solutions have already been explored: Benko et al. [2] proposed a balloon metaphor to control a cursor, 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. [8], created a variation of this technique by adding a corkscrew metaphor. Strothoff et al. [28] 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 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 [11] 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 [2] and other widgets that provide both the remaining rotations and independent scale along three axes. Hilliges et al. [16] created a similar setup, the Holodesk, which allowed direct interaction with 3D graphics, using physical simulation and a depth camera for hand tracking. Mockup Builder [1, 9] 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 and adapted TRS to three dimensions to manipulate objects.

2.3 Mid-Air Interactions

Although not focusing on stereoscopic visualizations, Hilliges et al. [15] presented a technique to seamlessly switch between interactions on the tabletop and above it. Marquardt et al. [20] 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.

Song et al. [27] 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. The Color Glove [30], despite being an invasive wearable device, enabled precise finger and hand pose tracking. More recently, Wang et al. [29] 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.

The above mentioned research addresses 3D virtual objects in stereoscopic environments focusing on multi-touch input or wearable devices. 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 requires mapping 2D input subspaces to a 3D virtual world. Wearable devices remove these restrictions. However, they are invasive and may be unnatural and uncomfortable to users. Other authors propose non-intrusive tracking solutions for hands, based on depth cameras, but do not apply these to stereoscopic visualizations.

In our setup, we integrate depth cameras to track the users' head and hands, in a stereoscopic multi-touch tabletop environment. Our setup is similar to Araújo et al. [1, 9], but while theirs explores the continuous interaction space to create content above the surface, tracking users' fingers in 3DOF using wired rings, we explore mid-air object manipulations with 6DOF non-intrusive tracking for each hand. Based on the work presented here, we implemented several approaches to directly and indirectly manipulate 3D virtual objects using both mid-air and touch gestures.

3 SETUP

Stereoscopic displays can provide enhanced visualization. Combining the capability of sending different images to each eye with head tracking, they can create the illusion of virtual objects placed on top of the surface. Integrating this with hand and finger tracking solutions, we built a complete 3D interactive space that mimics interactions with physical objects. Our setup uses affordable and non-intrusive solutions to create such an interactive space. This setup improves on a traditional multi-touch tabletop, with a stereo

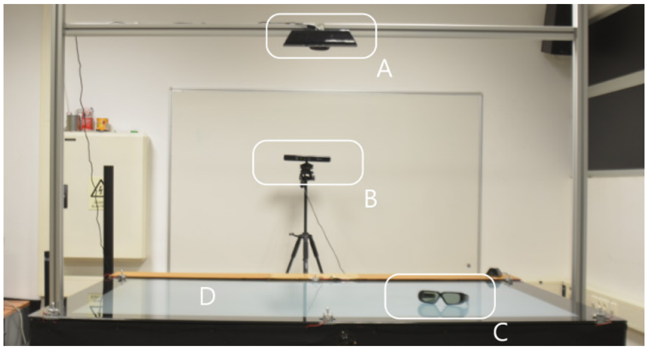

Figure 2: Our stereoscopic multi-touch tabletop (D) enhanced with depth cameras for non-intrusive tracking of head (B) and hands (A). Active shutter glasses (C) ensure the correct image for each eye.

display enabled, adding depth cameras for head and hand tracking, as illustrated in Figure 2.

Our tabletop uses the Laser Light Plane technique to detect user touches: a thin layer of infra-red light is placed slightly above the surface using lasers. An infra-red camera captures the light reflected by the users' fingers when they cross the laser plane. A NVidia Quadro K5000 equipped computer, paired with a stereo capable 120Hz HD Ready projector and NVidia 3D Vision glasses, enable the stereoscopic render and assure that the correct image reach each eye. We use a depth camera (Microsoft Kinect), placed behind the tabletop, to track the user skeleton. Then we use the tracked position of the user head to generate the corresponding visualization frustum.

To explore mid-air interactions, we use an additional depth camera (also Microsoft Kinect) placed above the surface and pointing downwards. This camera tracks the users' hands, capturing the corresponding point cloud and allowing their reconstruction. For this purpose, we use the 3Gear SDK[1]. By doing so, we get not only the hand positions, but also their orientation and pose, without forcing the user to wear specific hardware. This information allows us to explore spacial interaction techniques above the table, in a non-intrusive manner, making the setup more immediate to people.

4 SPATIAL INTERACTIONS

Using the setup described above, we implemented five different interaction techniques, based on the literature, for object manipulation. Four of these use mid-air interactions, both direct and indirect, and one is solely touch-based. All implemented techniques provide 7 DOF, three for translation, three for rotation, and a uniform scale.

In the mid-air techniques, when users place their hand inside an object, a red wired box appears around that object. The user can then perform a pinch gesture to grab the object, as shown in Figure 3. The wired box changes to solid transparent green to provide visual feedback of a successful grab. Opening their hand, the user can release the object, making the bounding box disappear.

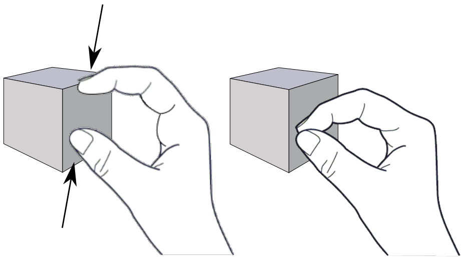

Figure 3: Pinch gesture in mid-air interaction techniques to grab an object.

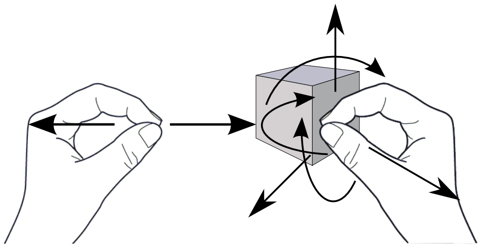

Figure 4: The 6-DOF Hand technique. The hand that grabs the object directly controls its translation and rotation. The distance between both grabbed hands scales the object.

4.1 6-DOF Hand

To mimic interactions with physical objects, as closely as possible, we use all 6 DOF information provided by the 3Gear 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 4. Dragging the object in space moves it in three dimensions, and the wrist rotation controls object rotation. Grabbing somewhere in space outside the bounding box 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.

4.2 3-DOF Hand

In this technique we divided translations and rotations by both hands (Figure 5) to prevent unwanted manipulations, as suggested in [25]. 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.

4.3 Handle-Bar

Following the work of Song et al. [27], we implemented the Handlebar metaphor in our stereoscopic tabletop. This approach mimics a physical bimanual handle-bar, commonly used, for example, to

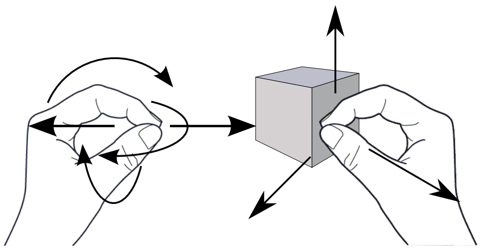

Figure 5: The 3-DOF Hand technique. The hand that grabs the object directly controls its translation. The rotations of the other hand define the object orientation. The distance between both hands scales the object.

---

[1] 3Gear Systems, http://www.threegear.com/, last visited June 10th 2013.

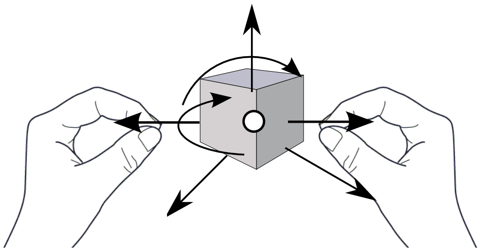

Figure 6: The Handle-Bar technique. The middle point of both hands is used to manipulate the object, reacting as if the user was holding a bar placed across the object. The distance between both hands scales the object.

roast a turkey. In this technique, we use the middle point of each hand, after the grab, to manipulate virtual objects (Figure 6). 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.

4.4 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 [14], as illustrated in Figure 7. 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.

4.5 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 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 [11], or to change the height of the object, similarly to the balloon metaphor [2].

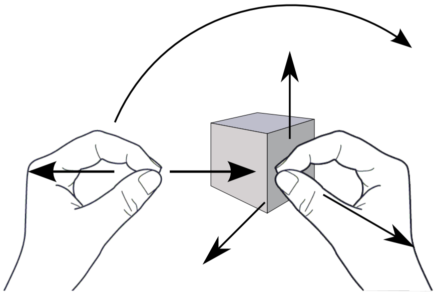

Figure 7: The Air TRS technique. The first hand grabs and moves the object. The movement of the second hand relatively to the first defines rotation and scale transformations.

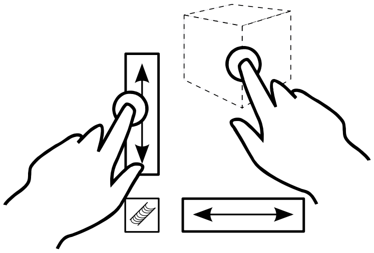

Figure 8: The Touch TRS + Widgets technique. One touch bellow the object enables widget visibility and moves the object. A second touch outside the widgets apply the TRS algorithm (translation and yaw rotation). The widgets offer height manipulation, roll and pitch rotations.

5 USER EVALUATION

To validate the techniques described above, we carried out a user evaluation. We aimed at identify which were the more natural 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 setup detailed in Section 3.

Every evaluation session for each participant followed the same protocol, starting with a short briefing explaining the experiment they were about to perform. This briefing also focused on some technical limitations participants had to be aware of, especially regarding the hand tracker. Since the camera is placed above the surface and pointing downwards, and it needs to capture the fingers used to pinch, participants should avoid wrist rotations that would finish with such fingers pointing down, being hidden by the hand. This can result in a non released object, leading to unwanted movements.

Each technique was evaluated in partial random order, to ensure that all methods were experienced in every position at least once between all participants, to avoid biased results due to participants becoming acquainted to tasks and more used to the technology. After a technique was selected, a brief demonstration video showing how to use it was presented to the participant. After the video, subjects had two minutes to explore the technique in a training scenario. We provided additional training time if a participant was not yet comfortable using the technique, and also gave further explanations as needed. We allowed limited time and a maximum of five repetitions for every task in order to prevent sessions from lasting too long. On average, each session took around sixty minutes to complete.

5.1 Tasks

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. [22], 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.

In all tasks, when a participant fulfilled the completion requirements, the movable object became white and locked in the current position. We chose to introduce this restriction to avoid tracker problems and user frustration when releasing an object that was already correctly placed. After the participant released the object, it

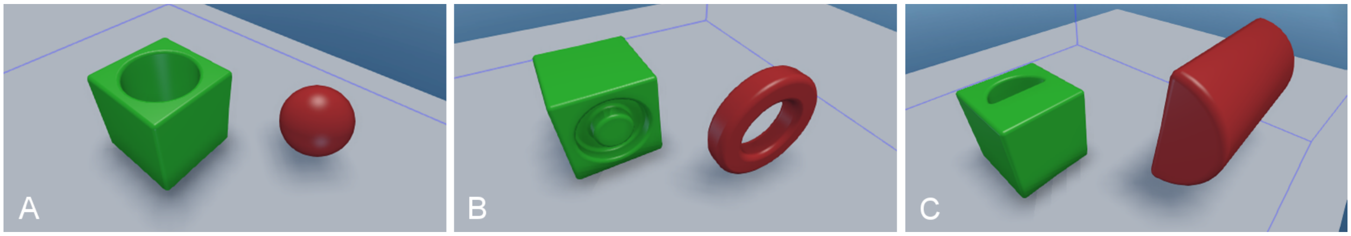

Figure 9: User evaluation Tasks. First task (A) consists in fitting a sphere inside the hole of the box. Second task (B) consists in fitting a stylized torus inside the hole on the front box face. Third task (C) consists on fitting the semi-cylinder inside the box hole.

changed to a random controlled position in order to create different completion paths. Our virtual environment supported neither gravity nor object collisions, and participants could only grab the movable object.

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. In this task we asked the participants to put a sphere inside a box with a hole as many times as they could within one minute, as depicted in Figure 9.A.

The second task required translations in all three axes and also scale transformations, but did not require any rotations. The scenario for this task consisted in a stylized torus, and a box with a torus-shaped hole in the front face (Figure 9.B). Each participant had to fit the torus inside that hole as many times as they could within two minutes.

For the third task we asked participants not only to translate and scale, but also to rotate the object. The scenario, illustrated in Figure 9.C, consisted in a semi-cylinder and a box having a hole with the same shape. As for the previous tasks, we asked participants to fit the cylinder inside the cube, keeping in mind that scale and rotation and position were equally important. It was possible to fit the object in two different orientations and both were accepted by the system. We asked subjects to try and complete this task as many times as they could within a three minute interval.

5.2 Participants

Twelve subjects performed the user evaluation, eleven males and one female. Participants' ages ranged from 19 to 35 years old (mean: 25) and all held a bachelors degree. Only two participants did not own a touch device, such as a smart-phone or tablet. However, only five had previous experience with stereoscopic visualization. Furthermore, only three had experience with 3D modeling applications. All tests were recorded in video with the written agreement of subjects. Figure 10 shows one participant interacting with the prototype during the tests.

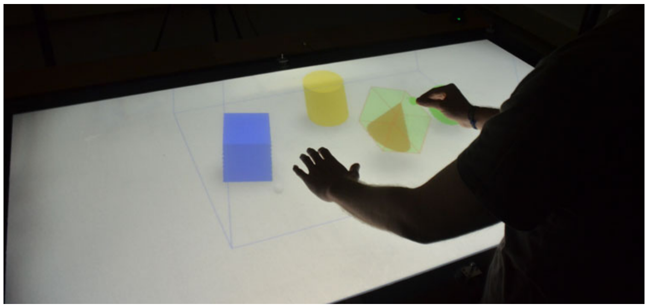

Figure 10: Participant manipulating objects in our training scenario.

6 RESULTS AND DISCUSSION

To measure performance, we monitored the time spent to complete each task for every interaction. After completing all tasks for each technique, participants answered a brief questionnaire regarding different interaction aspects. These questionnaires also included queries to profile participants. Furthermore, we registered relevant participant actions and comments both during and after the experiment.

6.1 Quantitative Analysis

Concerning quantitative analysis, we compared the implemented techniques separately. For each task we measured the time taken to complete it and registered the time participants spent using one or two hands manipulating the object. Since users repeated the task several times for each technique, we averaged their times, in order to obtain a more accurate value per user. Results were subjected to a pairwise comparison using the Wilcoxon Signed Ranks Test to assess statistically significant differences.

For the first task, the completion times are illustrated in the graph of the Figure 11 (mean times: 6-DOF Hand 2.27 sec, 3-DOF Hand 2.0 sec, Handle-Bar 1.88 sec, Air TRS 1.95 sec, Touch TRS + Widgets 2.53 sec). The Wilcoxon Signed Ranks Test suggested that statistically significant differences existed. Touch TRS + Widgets was significantly slower than the 3-DOF Hand (Z=-2.510, p=.012) and the Handle-Bar (Z=-2.589, p=.01). The absence of significant differences among the other techniques can be justified by the fact that only translation was needed to complete this task and these techniques are all similar with regard to translate. Indeed, if we consider translation with one hand, 6 DOF Hand, 3 DOF Hand and Air TRS become similar.

Regarding the second task, the times required by the participants to complete the task are plotted in Figure 12 (mean times: 6-DOF Hand 9.39 sec, 3-DOF Hand 9.73 sec, Handle-Bar 3.40 sec, Air TRS 6.99 sec, Touch TRS + Widgets 18.1 sec). Applying the same Wilcoxon Signed Ranks Test for the results obtained in this task, we found statistically significant differences. Notably, using the Handle-Bar significantly reduced the completion times (6-DOF Hand Z=-2.746, p=.006; 3-DOF Hand Z=-2.934, p=.003; Air TRS Z=-3.059, p=.002; Touch TRS + Widgets Z=-2.934, p=.003). On the other hand, the Touch TRS + Widgets was the slowest technique (6-DOF Hand Z=-2.490, p=.013; 3-DOF Hand Z=-2.395, p=.017; Air TRS Z=-2.934, p=.003). One of the possible reasons why Handle-Bar was the fastest technique is that since the task required only translation and scale, after grabbing the object, participants already have both hands in position to change the scale while moving. Also, since this technique only uses hand position, discarding wrist rotations, there were less unwanted rotations. The Touch TRS + Widget approach requires constantly changing between two touch TRS and different widgets, which led to the slowest completion times.

Analyzing the Wilcoxon Signed Ranks Test values obtained with the third task result completion times, which are depicted in Fig-

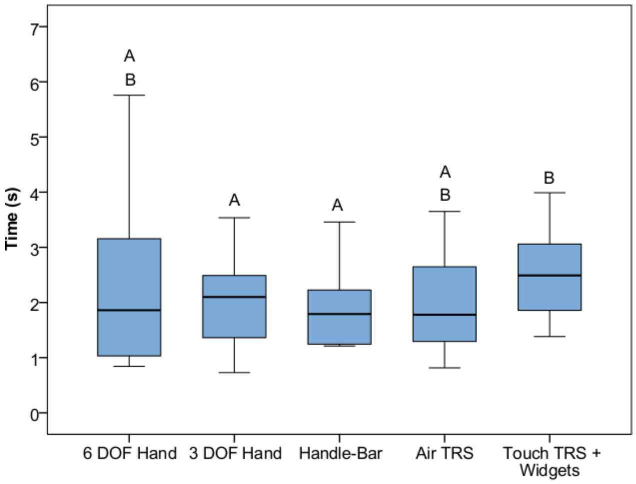

Figure 11: Time to complete the first task using the five techniques. The graphic presents the median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers). Techniques labelled by different letters are significantly different.

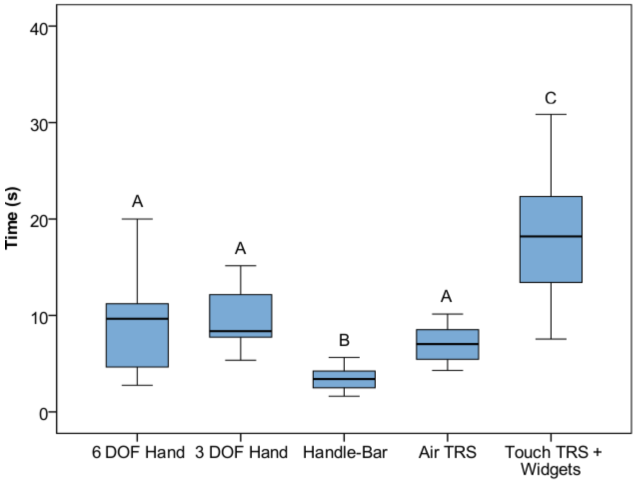

Figure 12: Time to complete the second task using the five techniques. The graphic presents the median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers). Techniques labelled by different letters are significantly different.

ure 13 (mean times: 6-DOF Hand 20.6 sec, 3-DOF Hand 46.9 sec, Handle-Bar 23.7 sec, Air TRS 63.3 sec, Touch TRS + Widgets 51.6 sec), statistically significant differences were found. We verified that both 6 DOF Hand (3-DOF Hand Z=-2.521, p=.012; Air TRS Z=-2.756, p=.006; Touch TRS + Widgets Z=-2.981, p=.003) and Handle-Bar (3-DOF Hand Z=-2.380, p=.017; Air TRS Z=-2.073, p=.038; Touch TRS + Widgets Z=-2.310, p=.021) stood out significantly from the others, with lesser completion times. We believe that the Handle-Bar was one of the fastest techniques, because besides those based on touch, it was the only method that did not lead to occlusions caused by subject's hands. This allowed a better perception of the object relative to its target position and orientation. The 6 DOF Hand mimics interactions with physical objects, which appeared more natural to participants, which is reflected in the lower times required to complete the task.

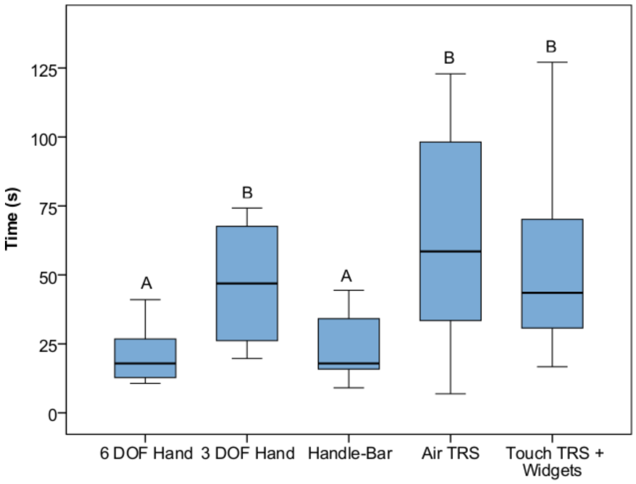

Figure 13: Time to complete the third task using the five techniques. The graphic presents the median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers). Techniques labelled by different letters are significantly different.

6.2 Qualitative Analysis

In the questionnaires, we asked participants to classify each technique regarding five different criteria using a five point Likert scale (1 - very bad, 5 - very good). The answers are depicted in Table 1. We used the Wilcoxon Signed Rank test to assess whether differences were statistically significant. In what concerns translation of objects, the five techniques presented no statistically significant differences in terms of preference.

Participants strongly agreed that, for rotation, 3-DOF Hand is more difficult to use than 6-DOF Hand, Air TRS and Touch TRS + Widgets (Z=-2.965, p=.003 and Z=-2.292, p=.022 and Z=-2.976, p=.003). Also, subjects strongly agreed that Handle-Bar offers a more difficult way to rotate objects than 6-DOF Hand and Touch TRS + Widgets (Z=-2.588, p=.010 and Z=-2.157, p=.031). The possible reason for participants to dislike 3-DOF Hand for rotations may reside in that rotating an object using the opposite hand that is grabbing it is not natural and may require some experience. The opposite can be inferred for 6-DOF Hand and that may explain why participants preferred this technique for rotations.

In terms of scaling objects, participants strongly agreed that both 6-DOF Hand and Handle-Bar were easier to use than 3-DOF Hand (Z=-2.070, p=.038 and Z=-2.913, p=.004). Because 3-DOF Hand uses the wrist of the hand that is not grabbing the object to rotate it, scaling can be more complicated to use than intended, due to unwanted rotations.

Regarding interaction fluidity, participants strongly agreed that 6-DOF Hand is superior to 3-DOF Hand, Air TRS and Handle-Bar (Z=-2.994, p=.003 and Z=-2.226, p=.026 and Z=-2.333, p=.020). They also strongly agreed that Handle-Bar is more fluid than 3-DOF Hand (Z=-2.636, p=.008). These opinions can be explained because 6-DOF Hand mimics interactions with physical objects, which makes actions seem more natural.

Finally, considering the fun factor, subjects strongly agreed that 6 DOF Hand was the better overall method (following table order: Z=-2.992, p=.003 and Z=-2.887, p=.004 and Z=-2.070, p=.038 and Z=-2.308, p=.021) and that 3 DOF Hand was the less amusing (following table order: Z=-2.877, p=.004 and Z=-2.850, p=.004 and Z=-2.854, p=.010). We think that the directness and easiness of 6 DOF Hand mimicking physical interactions explains this result.

	6 DOF Hand	3 DOF Hand	Handle-Bar	Air TRS	Touch TRS + Widgets
Translation	4,5 (1)	4 (1)	4 (2)	4 (1)	4 (2)
Rotation*	4 (2)	2 (2)	3 (2)	3 (2)	4 (2)
Scale*	4,5 (1)	3,5 (2)	5 (1)	4 (2)	4 (0)
Fluidity*	4 (1)	3 (1)	4 (1)	4 (2)	3,5 (3)
Fun*	5 (1)	2 (1)	4 (1)	4 (2)	4 (1)

* indicates statistical significance

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

6.3 Observations

During the entire experiment we observed and recorded everything relevant that participants said and did. While we implemented all techniques to support bimanual operation, we did not require any specific hand to be used first when grabbing objects. This enabled subjects to choose the most convenient hand to start the manipulation. While we assumed users to choose the dominant hand first, as suggested by Guiard asymmetric bi-manual model [10], we observed that participants changed their preference depending on which hand was closer to the object. The same was verified when each hand controlled a different degree of freedom e.g. for the 3-DOF technique. To rotate objects, some participants use the dominant hand to rotate the object rather than using it to control its position in space independently of their handedness.

For each task, we recorded the time spent performing it using either a single hand or both hands. Figure 14 presents the percentage distribution for all users in the third task which involved greater variability in transformations (translation, rotation and scale). Depending on the technique, we noticed a strong dependence on the DOF distribution. Obviously, we do not discuss the Handle-Bar method since it requires both hands. Regarding Touch TRS + Widgets, we could see that users spent almost the same time using one or two hands. This may be because participants kept the object selected with one finger while figuring out on the next transformation needed to fulfil the requirements. Both 3-DOF Hand and Air TRS require one hand to translate and two hands to rotate and scale the object, thus showing a similar time distribution between one and two hand manipulations. However, the required hand gesture for rotation is different for each technique and the 3-DOF induced participants to spend more time trying to achieve the right object rotation. The 6-DOF Hand is the approach that needed less time using two hands, because it closely mimics physical direct interactions, requiring only one hand to simultaneously execute both translation and rotation. The second hand is only needed to perform scale transformations which scarcely happens. Such behavior would free the second hand to execute other actions, such as coloring the grabbed object, which is not feasible when using the other techniques.

When rotating objects, subjects using 3-DOF Hand and 6-DOF Hand methods complained about poor tracker fidelity, which created unwanted rotations and increased the difficulty of tasks. These problems were also noticeable when releasing objects. Another interesting fact about translations regards Touch TRS + Widgets, where participants always started using the balloon widget to the wrong side of the final position they wanted to acquire. Still concerning interactions, many participants felt an initial difficulty when trying to rotate objects using either Air TRS or Handle-Bar. Indeed, the metaphor requires indirect rotation and this is easy to get wrong.

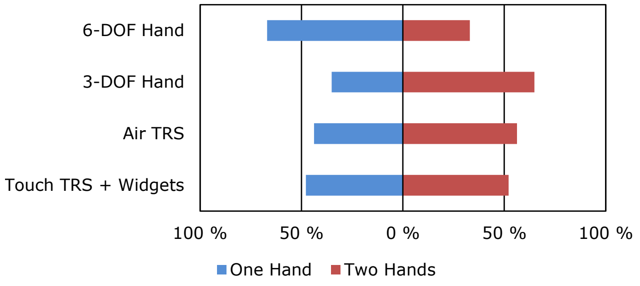

Figure 14: Time distribution of one- and two-handed object manipulations in the third task of our evaluation.

All subjects agreed that the mid-air approaches were more fun to use than touch-based ones. However, many participants complained that mid-air interactions that operate directly on objects (6-DOF Hand, 3-DOF Hand and Air TRS) suffer from having the selected object occluded by their hand. Despite finding them tedious, subjects agree that touch interactions are easy to use. Some participants also indicated that they missed a widget to perform TRS transformations on the surface.

Regarding our stereoscopic setup, while some participants reported fatigue due to active glasses, most reacted very positively to being able to co-locate their hands with objects above the table. Indeed, we noticed that most participants moved their head to take advantage of the head-tracked perspective to better understand object position and orientation. During the training sessions we also observed a strong sense of engagement in several participants as judged from the amount of time and flourish they spent on interacting with architecture constructs in our setup.

7 CONCLUSIONS AND FUTURE WORK

In this paper we compared five different techniques, both direct and indirect, to manipulate three-dimensional virtual objects using a stereoscopic tabletop, in order to assess which are best suited to this setup. Four of these techniques were mid-air air and one was multi-touch, used as a baseline. All these methods use a stereoscopic multi-touch tabletop setup with non-invasive user head and hand tracking. We track head position with affordable depth cameras, to generate images according to users' perspective, creating the illusion of three dimensional virtual objects placed above the interactive surface. Another depth camera above the tabletop provides precise tracking of each hand's position, orientation and pose.

We evaluated each implemented technique via three tasks of increasing difficulty to try and find out which method was favored in terms of object manipulation in our stereoscopic scenario. To this end, 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 our 6-DOF Hand approach was more natural to use, since it reproduces direct interactions with physical objects. Results also showed that the Handle-Bar [27] 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 [4, 5], that did not affect the Handle-Bar. We believe that overcoming such challenges will allow us to improve the technique, making it the more appropriate way to manipulate three dimensional virtual objects in stereoscopic environments, thus reducing the gap between virtual and physical interactions.

The main conclusion of this paper is that, concerning virtual objects lying above the surface, 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.

As future directions for our work, we aim at improving hand tracking precision, since the resolution offered by the Microsoft Kinect depth cameras is low and this creates difficulties. Other devices are being deployed, such as Leap Motion or the second generation Microsoft Kinect for the XBox One, with Full-HD resolution, which are worthy of attention and future explorations. We also want to explore mid-air 3D widgets that, despite being less natural, have proven to be more effective than touch based solutions for virtual object manipulations. It would be interesting to verify whether this also holds true for mid-air interactions, and whether other techniques could also mitigate object occlusion by the manipulating hand. Approaches based on indirect manipulations using a cursor with offset or simulating puppet on a string interactions might be a way to tackle this challenge. We plan to further explore bimanual input and assess user preferences for ascribing different DOFs to either dominant or non-dominant hands as suggested by prior work by Lopes et al. [19] and Araújo et al. [1]. We also plan to apply the spatial interactions presented here in real-world scenarios, such as fast prototyping of architectural models or assembling engineering parts. The ultimate goal is to abut at stereoscopic techniques which fully exploit direct manipulation affordances to provide a greater sense of control and familiarity to make interactions feel more intimate.

ACKNOWLEDGEMENTS

The work presented in this paper was partially supported by the Portuguese Foundation for Science and Technology (FCT) through projects Alberti Digital (PTDC/AUR-AQI/108274/2008), CEDAR (PTDC/EIA-EIA/116070/2009), TECTON-3D (PTDC/EEI-SII/3154/2012) and Pest-OE/EEI/LA0021/2013, and through doctoral grant SFRH/BD/91372/2012.

REFERENCES

[1] 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.

[2] H. Benko and S. K. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, page 22. IEEE, 2007.

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

[4] G. Bruder, F. Steinicke, and W. Stuerzlinger. Effects of visual conflicts on 3d selection task performance in stereoscopic display environments. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE Press, 2013.

[5] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung. Touching the void: direct-touch interaction for intangible displays. In *Proc. of CHI '10*, pages 2625–2634, NY, USA, 2010. ACM.

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

[7] 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.

[8] 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.

[9] B. R. De 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.

[10] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior*, 19:486–517, 1987.

[11] M. Hachet, B. Bossavit, A. Cohé, and J.-B. de la Rivière. Toucheo: multitouch and stereo combined in a seamless workspace. In *Proceedings of ACM UIST '11*, pages 587–592, NY, USA, 2011. ACM.

[12] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proc. of CHI '07*, pages 1147–1156, NY, USA, 2007. ACM.

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

[14] M. Hancock, F. Vernier, D. Wigdor, S. Carpendale, and C. Shen. Rotation and translation mechanisms for tabletop interaction. In *Proceedings of First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, TABLETOP '06, pages 79–88, 2006.

[15] 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 ACM UIST '09*, pages 139–148, New York, NY, USA, 2009. ACM.

[16] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. Holodesk: direct 3d interactions with a situated see-through display. In *Proceedings of ACM CHI '12*, pages 2421–2430, NY, USA, 2012. ACM.

[17] S. Houde. Iterative design of an interface for easy 3-d direct manipulation. In *Proc. of CHI '92*, pages 135–142, NY, USA, 1992. ACM.

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

[19] P. Lopes, D. Mendes, B. Araújo, and J. A. Jorge. Combining bimanual manipulation and pen-based input for 3D modelling. In *Proc. of SBIM '11*, pages 15–22, NY, USA, 2011. ACM.

[20] 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 INTERACT'11*, pages 461–476, Berlin, Heidelberg, 2011. Springer-Verlag.

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

[22] A. Martinet, G. Casiez, and L. Grisoni. The effect of dof separation in 3d manipulation tasks with multi-touch displays. In *Proceedings of ACM VRST '10*, pages 111–118, New York, NY, USA, 2010. ACM.

[23] D. Mendes, P. Lopes, and A. Ferreira. Hands-on interactive tabletop lego application. In *Proceedings of ACE '11*, pages 19:1–19:8, New York, NY, USA, 2011. ACM.

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

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

[26] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of ACM UIST '09*, pages 69–78, New York, NY, USA, 2009. ACM.

[27] 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 CHI '12*, pages 1297–1306, NY, USA, 2012. ACM.

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

[29] R. Wang, S. Paris, and J. Popović. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of ACM UIST '11*, pages 549–558, New York, NY, USA, 2011. ACM.

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