ARPen: Mid-Air Object Manipulation Techniques for a Bimanual AR System with Pen & Smartphone

Philipp Wacker, Oliver Nowak, Simon Voelker, Jan Borchers
RWTH Aachen University
52056 Aachen, Germany
[wacker,nowak,voelker,borchers]@cs.rwth-aachen.de

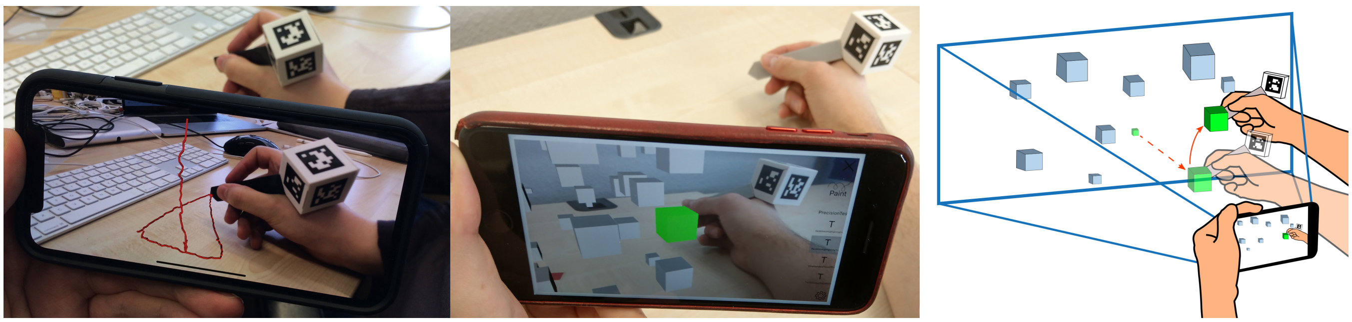

Figure 1: *Left*: Our ARPen smartphone app tracks the real-world location of the ARPen, letting the user sketch in mid-air. *Middle*: We compared the performance of different mid-air selection techniques. For *with highlight*, a box was highlighted when the pen tip was inside the box. *Right*: The preferred technique for translation tasks was *pen ray pickup*: Pressing a button on the pen with the tip in front of a target snaps it to the tip; releasing the button drops it.

ABSTRACT

Modeling in Augmented Reality (AR) lets users create and manipulate virtual objects in mid-air that are aligned to their real environment. We present ARPen, a bimanual input technique for AR modeling that combines a standard smartphone with a 3D-printed pen. Users sketch with the pen in mid-air, while holding their smartphone in the other hand to see the virtual pen traces in the live camera image. ARPen combines the pen's higher 3D input precision with the rich interactive capabilities of the smartphone touchscreen. We studied subjective preferences for this bimanual input technique, such as how people hold the smartphone while drawing, and analyzed the performance of different bimanual techniques for selecting and moving virtual objects. Users preferred a bimanual technique casting a ray through the pen tip for both selection and translation. We provide initial design guidelines for this new class of bimanual AR modeling systems.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300849

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Interaction techniques**; *Empirical studies in HCI*;

KEYWORDS

Smartphone AR, Mid-air interaction, Augmented Reality, Immersive Modeling, bimanual interaction, pen

ACM Reference Format:
Philipp Wacker, Oliver Nowak, Simon Voelker, Jan Borchers. 2019. ARPen: Mid-Air Object Manipulation Techniques for a Bimanual AR System with Pen & Smartphone. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300849

1 INTRODUCTION

With the rekindled interest in Virtual Reality (VR) and Augmented Reality (AR), several research projects [1, 2, 15, 20, 31, 41, 44] have explored using these technologies to let users create 3D models in mid-air. Using AR for this task is particularly interesting, since it allows interacting directly in the real world and on existing physical objects, without having to first digitize the environment or existing objects. AR modeling also benefits from the haptic feedback the physical objects' surfaces provide [39]. To interact with virtual objects in AR or VR requires manipulation operations, such as selecting an object to change its properties, or moving it to a new location. In this paper we evaluate different selection and translation techniques with our ARPen system.

AR systems can be classified into *wearable systems*, such as head-mounted displays, and *non-wearable systems*, such as smartphones or projector setups [30]. Most AR modeling projects use wearable systems (e.g., [1, 31, 44]). They offer a hands-free visualization that shows the digital objects overlaid over the real world. However, current systems require external tracking of the input device and precise calibration to map the real world input to the virtual coordinate system.

Among non-wearable systems, consumer smartphones have recently gained development support for AR[1], and can track their position in relation to surfaces to place 3D content into their live camera feed. While this requires holding the phone like a camera, moving the phone enables precise viewport control, and AR apps can use the phone touchscreen for interaction. Projecting a 3D scene onto the 2D screen, however, reduces depth information, making it difficult to specify a point at a specific depth in 3D by interacting with the 2D projection [21, 28, 32]. We propose to specify the position of a point in 3D bimanually, using a pen tracked by the smartphone camera.

We developed the ARPen system to prototype and evaluate this interaction. It uses a recent iPhone and a 3D-printed pen with wireless buttons near the tip and visual markers at its end. The ARPen app tracks the position of the phone in the real-world environment using Apple's ARKit, and the position of the pen tip via the pen's visual markers. This allows drawing and interacting with virtual objects that are anchored in real space. Buttons on the pen serve to start and stop drawing, and to invoke other editing commands. Only requiring smartphone & pen enables this system to be used in many situations, e.g., to design an inset for the can holder in a car to support smaller can sizes, or to combine multiple moving elements such as hinges to create a box.

The touchscreen shows the AR scene, and the user can move the phone to change the viewport, e.g., to slice through objects to look behind them. At the same time, the app can use the touchscreen to display interactive controls and options such as a library of model parts like hinges that the user can place in the scene and combine with other objects.

To better understand this new interaction technique, we first analyzed how people would hold a smartphone while drawing with the pen in the other hand, and which parts of the screen they are capable of reaching with the same hand. To evaluate the interaction with virtual content in the scene, we then performed two studies comparing different techniques for two of the central tasks when interacting with AR content [9]: selecting and translating virtual objects.

In summary, this paper makes the following contributions:

- the ARPen interaction technique that allows for bimanual mid-air modeling in Augmented Reality using a pen and consumer smartphone,
- user studies evaluating the grip and reachable touchscreen areas in this technique, and comparing different techniques to select and move virtual objects.

In the remainder of this paper, we provide an overview of related work before explaining the interaction technique and our implementation in more detail. After reporting on our grasp, selection, and translation studies, we close with a discussion and provide initial guidelines and suggestions for researchers and designers of pen-based AR systems.

2 RELATED WORK

The ARPen interaction technique and the studies in this paper draw from related work in the areas of *immersive modeling* and *manipulating objects in mid-air*. Modeling in mid-air has already been around since the early 90s [33] and interactions with different types of AR/VR have also been evaluated from an early stage [10, 36]. A recurring problem for both AR and VR are perceptual issues for detecting the correct depth of objects ([11, 12, 37]; see [21] for a survey). Interaction techniques need to address these issues.

Immersive Modeling

Immersive modeling interaction techniques focus on creating 3D models in mid-air instead of interpreting 2D drawings to estimate 3D objects as in e.g., [4, 7, 13, 18, 34]. Models can be visualized anywhere along the mixed-reality continuum [26], with VR [20] and AR [1, 31, 44] being the most frequent.

In VR, [20] lets users place 2D sketches in mid-air before 'lifting' individual strokes into the third dimension. Users can also draw freehand mid-air paths, and enhance the resulting wireframe model by defining surfaces between strokes.

AR provides a link back to the real world and its physical objects. This supports designing objects that fit existing physical shapes, a frequent task in Personal Fabrication. RoMA [31] combines AR modeling with a robotic 3D printing arm to directly manufacture the designed wireframe model. WireDraw [44] lets users draw wireframes with visual step-by-step instructions via head-mounted AR. A more artistic approach [1] combines an AR headset with a motion tracking system to draw in mid-air. Since physical surfaces improve sketching precision [2, 39], the system also includes a tablet for sketches that are projected onto mid-air drawing planes.

Similarly, in mobile AR, Napkin Sketch [41] lets users specify a mid-air drawing plane, then draw strokes via touch on a tablet that are projected onto the drawing plane. [15] maps

[1] developer.apple.com/arkit, developers.google.com/ar

touch strokes onto surfaces tracked in the world, and allows touchscreen-based manipulation of the resulting objects.

In contrast to these approaches, the ARPen does not require external tracking or additional hardware other than the phone and pen, and it offers true 3D input for positions.

Manipulating Objects in Mid-Air

Manipulations like selecting and moving virtual content are central interactions for both VR and AR [9]. In handheld AR, many interaction techniques involve the touchscreen to manipulate the virtual content. For object selection, intuitive methods are to directly touch the projection of the object, or to have a central crosshair on the screen [38]. However, touching the screen for a selection often moves the device and thus the viewport into the scene. This can cause selections errors. To address this, several approaches 'freeze' the camera feed during touch interaction (e.g., [5, 23, 38]), which improves accuracy but takes longer. Special techniques improve selection in dense environments [29].

A key limitation of handheld AR is that the touchscreen does not directly support manipulating the 6 degrees of freedom of a virtual object in space. *3DTouch* [28] addresses this by interpreting swipes on the screen in the context of the position of the smartphone in the world. Another approach is to first let the user tap to cast a ray into the scene, fix the object to be placed on that ray, then drag on the screen to move the object along the ray [32].

Device movement can also be used to transform the virtual object [14, 17, 25, 28, 32] or move a cursor around a 3D environment [3]. Selecting and holding an object attaches it to the device. Now any device movement is applied to the object, enabling compound manipulation of position and rotation. Studies show that users prefer this technique over touchscreen interactions, achieving good translation and rotation performance [25, 28].

A more natural approach tracks the user's free hand while holding her phone, using a separate depth camera [6, 8] or the device camera [5, 17]. Colored finger markers improve the tracking [17], although this project only tracked 2D movements and used device motion for the missing dimension. At the time, although users enjoyed the interaction, gesture tracking was not accurate enough for fair evaluation [5].

Unlike these approaches, the ARPen provides direct 3D positional input suitable for modeling without multiple steps or viewport adjustments, and the pen buttons support issuing commands without touching the screen.

External input devices can be tracked in various ways, such as motion capture cameras [22, 42], stereo cameras [27], regular cameras [40], or by sensing the magnetic field [43]. The determined position has, for example, been used to slice through 3D data [19], control a game by projecting gestures into the 2D gameboard [16], recreate the geometry of a traced object [27], or for character customization [35].

However, none of these projects compared different techniques for the selection and translation of virtual objects in handheld AR with a pen.

3 THE ARPEN

The ARPen explores an alternative to the above approaches, using bimanual pen+smartphone interaction. Below, we describe the interaction technique we designed, and our hardware and software implementation to enable replication, further research, and application development.

Interaction Technique. Specifying a position in mid-air is important for 3D modeling applications—a problem of handheld AR systems. A precise 3D input device—similar to the mouse in desktop settings—is needed to perform controlled strokes and manipulations in mid-air. The mobile device provides another area of interaction: the touchscreen to switch modes or adjust settings, and the device itself by setting the viewport.

Each of those areas of interaction have tasks suited to their strengths: precise 3D input capabilities for the pen, familiar haptic interaction on the touchscreen. However, the combination pen+phone has the potential to improve and simplify AR 3D model generation. We want to use the ARPen to prototype and study these combinations.

For example, the hand holding the device could select an object while the pen is used to manipulate the object's rotation or size. Or the touchscreen could be used to select a model which is placed using the pen. Similar to [24], manipulating the viewport while holding the object with the pen could scale an object. Sequencing actions such as a translation followed by rotation could also be faster by switching the action on the screen while the pen adjusts the value.

Implementation. Tracking the position of the, mostly pen shaped [1, 20], input device in relation to the world is essential for mid-air modeling. We implemented a tracking only requiring a smartphone. With marker tracking, such as with arUco[2], we are able to track the position of a marker in camera coordinates. Wu et al. [40] do this with a fixed camera. However, ARKit for iOS tracks the device's location relative to surfaces in the scene. Combining the two techniques, we can track the 3D location of a marker in the world (Fig. 2).

The ARPen has six arUco markers on its end. This ensures that at least one marker is visible for the camera even if the pen is pointing away from the camera. Knowing the physical setup of pen and markers, we can determine the pen tip from the marker location and stabilize this location by averaging if multiple markers are visible. Furthermore, the pen transmits the states of three buttons to the phone via BLE.

[2]https://docs.opencv.org/3.1.0/d5/dae/tutorial_aruco_detection.html

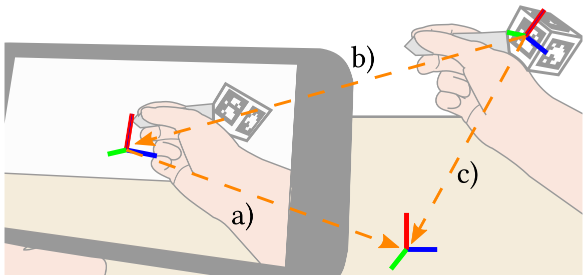

Figure 2: a) ARKit is used to determine the camera's position relative to the surfaces. b) arUco tracks the marker relative to the camera. c) The combination allows to calculate the position of the marker relative to the surfaces.

We wrote an iOS app to calculate the mid-air position of the pen. Using SceneKit[3], we can render a red ball at the tip of the pen. The implementation makes it simple to define new interactions based on the 3D position of the pen and its button states. For example, holding a button and moving the pen could draw a path mid-air (cf. Figure 1, left) or define the diagonal of a cube. To enable others to analyze their own mid-air modeling techniques, we provide an open source implementation of the ARPen.[4]

4 PHONE GRIP & INTERACTION AREA

When using the ARPen, the phone is operated with the non-dominant hand while the dominant hand is used for the pen interaction. We conducted a study on how people hold the phone and which areas they can reach on the screen while using the ARPen.

Study Setup

Participants used the ARPen on smartphones of different *size*: 4.7" iPhone 6s (*small*) and 5.5" iPhone 7 Plus (*big*). The *orientation* in which the participants should hold the phone varied between a *portrait* and two landscape orientations differing on the position of the device's camera. Holding the phone in the left hand, the hand is either grabbing the top (*camera left*) or the bottom of the screen (*camera right*). Both the order of *orientation* and *size* were counterbalanced.

No menu elements were shown on the screen to not influence the grip of the participant by the placement of buttons. We recorded the position of all touches on the screen to find out the available interaction area. This recording could be activated by the moderator to avoid accidental recordings while adjusting the grip.

Study Procedure

In the beginning of each condition, the phone was placed in front of the participant and she was told in which orientation it should be held. The participant was asked to draw freely in AR and find a suitable grip for the current orientation. After the participant said that they found a comfortable grip, we took a photo of how the phone was held and started the 'touch recording' mode. The participant had to trace and fill out the area she could reach with a finger of the hand holding the phone. No visual feedback was shown to avoid influencing the participant. We then stopped the recording, cleared the drawing area, and placed the phone back on the table. The sequence was repeated for all different *orientation*. Then, the conductor changed the phone to the other *size* and the task was repeated. Afterwards, the participant was asked about the preferred *size* and *orientation*.

Evaluation

Our initial categorization of grasps was into *valid* grasps that allowed to touch the screen and *invalid* grasps in which no interaction with the screen was possible.

For *valid* grasps, we analyzed the recorded touch points. Touch points of participants who held the phone in the right hand were mirrored to allow for a combined evaluation.

Results

18 participants (4 female, 2 n/a) took part in the study (M: 25.7 years, SD: 3.0 y). One left-handed participant and one right-handed participant held the phone in the right hand.

Size. Nine participants preferred the *big* phone because of the larger screen and because they felt that a wider camera image simplified keeping the pen's markers in view. Seven participants preferred the smaller phone and mentioned a more comfortable grasp and less weight.

Orientation. Size did not affect the subjective ratings regarding the *orientation*. *Portrait* and *camera right* both were most preferred similar times (*portrait*: 9, *camera right*: 8). However, *portrait* was placed in last place 4 times because the limited horizontal viewport would cause losing the pen's markers. *Camera left* was rated lowest by 14 participants, because most common grasps would occlude the camera.

Grasps. We categorized the *valid* landscape grasps in *pinkie*-, *thumb tray*-, *frame*-, and *front*-grasp (Figure 3, a–d):

- *Pinkie*: The phone rests on the pinkie finger, with the index finger holding the top.
- *Thumb tray*: Similar to the *pinkie*, but the phone also lies on the thumb tray.
- *Frame*: Thumb and middle finger form a frame by holding the phone from the side.
- *Front*: The phone is held with the thumb, index and middle finger laterally from the front.

Pinkie was used most for *camera right* especially (cf. Table 1). It was used less in *camera left* as it often occluded the

[3]https://developer.apple.com/documentation/scenekit
[4]https://github.com/i10/ARPen

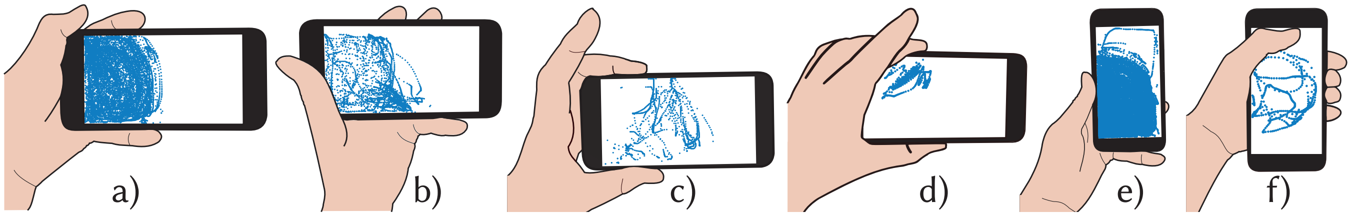

Figure 3: Grasps found in the study (with touch recordings). a) *pinkie*, b) *thumb tray*, c) *frame*, d) *front*, e) *low portrait*, f) *high portrait*). *Pinkie* and *low portrait* were used most often for their respective orientation.

		small	big
camera right	*pinkie*	50%	61.1%
	thumb tray	16.7%	11.1%
	frame	11.1%	11.1%
	front	0%	5.6%
	invalid	22.2%	11.1%
camera left	*pinkie*	27.8%	27.8%
	thumb tray	33.3%	27.8%
	frame	11.1%	16.7%
	front	11.1%	16.7%
	invalid	16.7%	11.1%
portrait	*low portrait*	100.00%	88.9%
	high portrait	0.00%	5.6%
	invalid	0.00%	5.6%

Table 1: Grasp frequencies for different orientations and sizes. *Pinkie* and *low portrait* are used most often.

phone's camera and participants adjusted the grasp towards *thumb tray*. We classified 11 grasps as *invalid*—mostly because the participant used a *frame*-grasp but with the index holding the top, leaving no finger to touch the screen.

For *portrait*, participants used two *valid* grasps: *low portrait*—the phone is held on the bottom—and *high portrait*—the phone is held around the middle (Figure 3, e&f).

Reachable Areas. The recorded touch points show that for landscape the reachable area of *pinkie* and *thumb tray* is located to the bottom left, while *frame* and *front* are located in the top center (cf. Figure 3, a–d). We grouped those grasps together and defined general touch areas by calculating average boundaries in x and y direction. For *pinkie* and *thumb tray*, the touch area has a width of 59.2 mm (*big*: 50.9 mm), a height of 52.2 mm (*big*: 56.0 mm) and is 0.8 mm away from the left edge and 2 mm from the bottom of the screen (*big*: 2.9 mm left, 5.1 mm bottom). *Frame* and *front* were closer to the screen center—20.4 mm from the left and 0.0 mm from the top of the screen—with smaller average width and height of 51.7 mm and 43.9 mm (*big*: 19.5 mm left, 2.89 mm top, 53 mm width, 41.1 mm height). Touch points for *low portrait* are 0.9 mm from the left and 2 mm from the bottom of the screen (*big*: 1.5 mm left, 1.9 mm bottom) with a width of 54.0 mm (*big*: 54.4 mm) and height of 66.7 mm (*big*: 68.3 mm). Lastly, for *high portrait*, which was used once on the *big* phone, the touches are 1.6 mm left and 20.55 mm from the bottom of the screen and the area has an average width and height of 58.9 mm and 78.3 mm.

Summary

We used a *big* iPhone and the *pinkie* grasp in the *camera right* orientation for the further studies. The *camera right* orientation was never the least preferred orientation as *portrait* had issues keeping the pen markers in view. The *big* phone gave participants the impression of seeing more of the scene also reducing issues of keeping the pen in view. Controlling this grasp avoids finding a selection technique that works well only in a screen orientation that is otherwise not working well for designing.

5 MID-AIR INTERACTION STUDY: SELECTION

To better understand the ARPen interaction, we take a look at user performance for two major mid-air interactions—*selecting* virtual objects and *translating* them [9]. We ran two studies that required participants to interact with objects within a 40cm×40cm×40cm volume on a table surface. This size enabled participants to stay seated during the study while holding phone and pen but they were encouraged to move the phone or stand up if they wanted to.

The selection of objects is required for manipulations of a specific object such as a change in color or a transformation. We compare five different techniques of selecting an object in mid-air using the ARPen (*selection technique*). We measured the success rate, selection time, deviation from the target, and the size of the object on the screen during the selection.

Selection Techniques

We sampled fundamental techniques to perform the selection from a large space of possible interactions: Pen Selection Without Highlighting (*without highlight*), Pen Selection With Highlighting (*with highlight*), One-handed Touch Selection (*one-handed*), Two-handed Touch Selection (*two-handed*), and Pen Ray Selection (*pen ray*). See Figure 4 for details.

Pen Selection Without Highlighting. The position of the pen is not visualized in the scene. Users have to match the pen

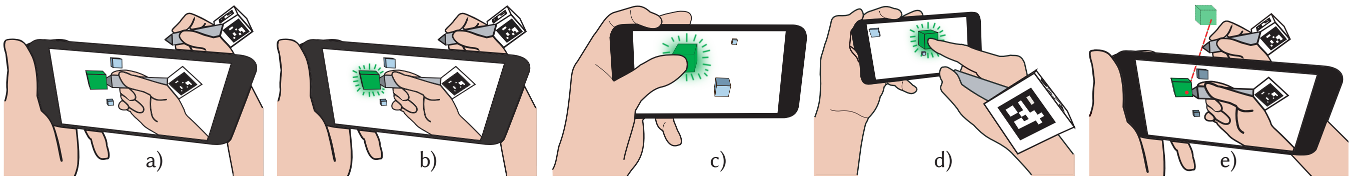

Figure 4: Selection techniques. a) *Without highlight*, b) *With highlight*, c) *One-handed*, d) *Two-handed*, e) *Pen ray*.

tip position in the real world to the position of the virtual object. Pressing a button on the pen confirms the selection at the current position.

Pen Selection With Highlighting. Here, the tip of the pen is visualized with a red sphere. The sphere disappears if the pen tip is behind or inside an object. If it is inside an object, the object changes into a visually highlighted state. While *without highlight* measures how well the selection worked just from the standard visualization, *with highlight* evaluates how basic depth cues could help improve the depth specification. A button press on the pen confirms the selection.

One-handed & Two-handed Touch Selection. From the position of a touch on the screen, a ray is cast into the scene and the first virtual object it contacts is selected. For *one-handed*, the user has to select the object using the thumb on the hand holding the phone. In the *two-handed* condition, the user performs the selection with the hand holding the pen.

Pen Ray Selection. Upon pressing a button on the pen, a ray is cast through the tip of the pen and the first target it hits is selected. This means that the user has to align the tip of the pen so that it is in front of the intended target. Pressing the button on the pen then selects this target.

Study Setup

The interaction volume was separated into 64 cubic areas with an edge length of 10 cm (*cube space*). Into each of these *cube spaces* we placed a white cube with an edge length of either 1 cm, 2 cm, 3 cm, or 4 cm. For each trial, 16 cubes of either size were assigned randomly to the *cube spaces* and their position within each *cube space* was randomized but needed to be at least 1 cm away from its side. Therefore, the distance between cubes was at least 2 cm (Figure 1, middle).

Participants had to select 64 targets using each of the five *selection technique*. At the beginning of a trial, one cube was shown as the target by changing its color to green. After a selection occurred, regardless of the success, a new cube was shown as the target. The order of targets was randomized.

Study Procedure

Participants sat in front of a table with tape markings to improve the world tracking and a visual marker to keep the position of the interaction volume consistent between participants. Each participant was asked to hold the phone in her non-dominant hand using the *pinkie* grasp. She was given time to familiarize herself with the grip and the ARPen before we introduced the task and techniques.

We demonstrated and explained each technique before letting the participant try for herself. For *two-handed*, the participant had to hold the pen in the hand tapping the screen to stay in the scenario of modeling in mid-air before selecting an object. If the participant needed a rest during the study or stopped to make a comment, we restarted the last selection. After selecting all 64 cubes, we asked the participant to rate the *ease of selection* and *confidence of selection* of the technique. For each technique, we noted qualitative observations about the selection strategy. After using all techniques, the participant was asked to rank the five techniques from best to worst. Overall, each participant selected 320 targets (64 cubes×5 *selection technique*). The order of conditions was counterbalanced between participants.

Measurements

Beside recording the success of the current selection, we measured the time from showing the target to the issued selection. For not successful selections, we calculated the deviation to the target. For *without highlight* and *with highlight* this deviation is the length of the vector between the specified 3D point and the target (in cm). To evaluate the offset for each dimension, we stored the direction of this vector in camera coordinates. For *one-handed*, *two-handed*, and *pen ray*, we measured the distance from the selection position on the screen to the convex hull of the target's projection (in cm). Since moving the phone adjusts the size of the target on the screen, we recorded the size of the bounding box of the projection at the time of the selection (in cm^2). We also collected subjective ratings for *ease of selection* and *confidence of selection* for each technique on 7-level Likert-Scales and a ranking of the five techniques.

Results

We recruited 15 participants (1 female, 1 n.a., 21–40 years, M: 28 years, SD: 5.4 years, all right handed). We discarded one participant's *without highlight* data as the selection was intentionally performed differently. Overall, we recorded 4735 selections. For every participant, we counted the successful

Technique	Success					Selection Time						Projected Size				
	Significance			Mean	SD	Significance				Mean	SD	Significance			Mean	SD
pen ray	A			87.08 %	5.74 %			C		2.6 *s*	0.4 *s*			C	0.42 cm^2	0.12 cm^2
with highlight	A			81.64 %	13.38 %	A				7.4 *s*	1.7 *s*	A	B		1.08 cm^2	0.45 cm^2
two-handed	A	B		78.54 %	14.31 %				D	1.9 *s*	0.6 *s*	A	B		1.22 cm^2	0.96 cm^2
one-handed		B		69.89 %	17.72 %			C	D	2.3 *s*	0.7 *s*	A			1.70 cm^2	1.93 cm^2
without highlight			C	2.90 %	6.85 %		B			5.8 *s*	1.7 *s*		B	C	0.69 cm^2	0.41 cm^2

Table 2: Means and standard deviations of *success*, *selection time*, and *projected size* for the main effect of *selection technique*. Rows not connected by the same letter are significantly different.

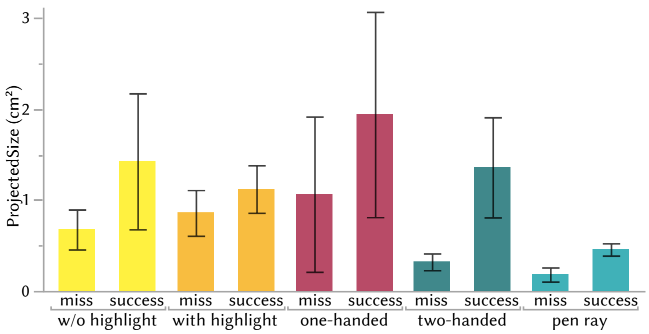

Figure 5: Interaction effect between *selection technique*×*success* on the *projectedSize*. Successfully selected targets are larger compared to missed targets. For *with highlight* this difference seems to be smaller compared to the other techniques. Whiskers denote the 95% CI.

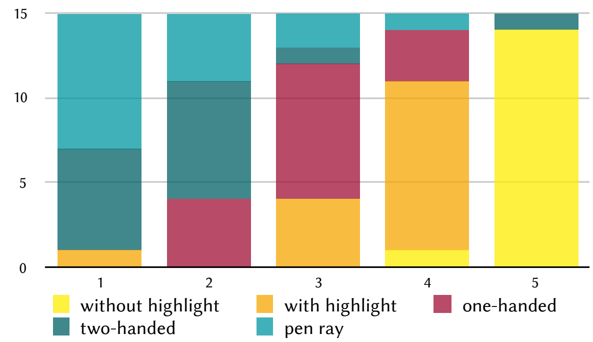

Figure 6: Subjective Ranking of *selection technique*. *Without highlight* is least preferred. *Pen ray* and *two-handed* are on the first two places. *One-handed* is mostly in third position.

selections per condition (*success*) as well as averaged the deviation for missed selections (*deviation*), time (*selection time*), and projected size (*projected size*). To analyze the effect of *selection technique*, we performed mixed-effect ANOVAs with the user as a random variable. We log-transformed *selection time* and *projected size* before the evaluation. All post-hoc pairwise comparisons were performed using Tukey HSD tests. The subjective Likert-Scale ratings were analyzed using the Kruskal-Wallis test and post-hoc comparisons using the Wilcoxon method with a Bonferroni correction.

Success, Selection Time, and Projected Size. The *selection technique* had a significant effect on *success* ($F_{4,56} = 187.09, p < .001$), *selection time* ($F_{4,55.06} = 94.016, p < .001$), and *projected size* ($F_{4,55.23} = 10.007, p < .01$). The means and results of the post-hoc tests can be seen in Table 2.

Splitting the results based on *success* shows that *projected size* for successfully selected targets is larger than for misses. This difference seems to be smaller for *with highlight*. For all other conditions, *projected size* is about twice as large or more for successful selections (Figure 5).

Deviation. The deviation in the mid-air pointing techniques was larger for *without highlight* (M: 26.9 cm, SD: 18.2 cm) compared to *with highlight* (M: 1.5 cm, SD: 3.32 cm). The average deviation vector in camera coordinates shows a large offset in z-dimension (*without highlight* x: 7.1 cm, y: 2.7 cm, z: 27.4 cm; *with highlight* x: 0.5 cm, y: 0.7 cm, z: 2.4 cm).

For the screen selection techniques, we recorded the least deviation for *pen ray* (M: 0.04 cm, SD: 0.02 cm) followed by *two-handed* (M: 0.2 cm, SD: 0.15 cm) and *one-handed* (M: 0.43 cm, SD: 0.73 cm).

Subjective Ratings. We found that *ease of selection* is rated significantly different for the techniques ($\chi^2(4) = 49.385, p < .001$). Post-hoc comparisons show that mid-air selection techniques (*without highlight* M: 2.14, SD: 1.79; *with highlight* M: 3.87, SD: 1.3) were rated as harder compared to the screen selection techniques (*pen ray* M: 6.67, SD: 0.62; *two-handed* M: 6.6, SD: 0.91; *one-handed* M: 5.8, SD: 1.32).

Confidence of selection also differed significantly ($\chi^2(4) = 42.347, p < .001$). Post-hoc tests show that *without highlight* (M: 2, SD: 1.24) is rated lower than all other techniques. Also, both *with highlight* (M: 5.06, SD: 1.67) and *one-handed* (M: 5.67, SD: 1.29) are rated lower compared to *pen ray* (M: 6.67, SD: 0.49). *Two-handed* (M: 6.33, SD: 0.62) is not rated different to other conditions except *without highlight*.

The ranking shows that participant liked both *pen ray* and *two-handed* followed by *one-handed*. *Without highlight* is ranked mostly as the least preferred technique (Figure 6).

Qualitative Remarks. A recurring qualitative remark suggests that participants understood the arrangements of the boxes better in the *with highlight* condition compared to other

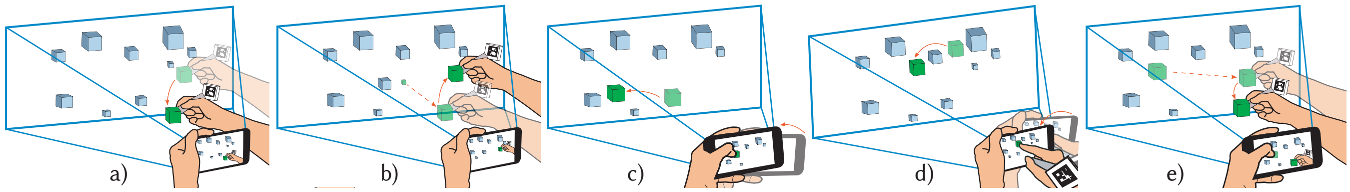

Figure 7: Translation techniques. a) *Pen drag&drop*, b) *Pen ray pickup*, c) *One-handed*, d) *Two-handed*, e) *Touch&pen*.

conditions. For *pen ray*, most participants used a strategy in which they did not vary the distance between pen and phone but moved them together. Overall, observations indicate that the device was moved more in the *one-handed* and *two-handed* conditions. As expected, participants mentioned fatigue in both their arms during the conditions.

Discussion

Pen ray seems to be the best candidate for selection tasks with the ARPen. This technique has the highest *success* rating combined with a quick *selection time*. The small *projected size* indicates that the device did not have to be moved much to select targets. Together with *two-handed*, *pen ray* is also the preferred selection technique of participants. *With highlight* has a good *success* rate but the *selection time* was the slowest likely because participants had to adjust their position to the correct depth for the selection. However, this might lead to a better understanding of the arrangement of objects in the scene. The touch conditions performed well based on the *selection time* and *two-handed* also shows a good *success* rate. The *projected size* indicates the phone was moved closer to the targets which might become more exhausting over time. *Without highlight* had the least *success* and was the least preferred technique. This shows that depth perception in handheld AR requires additional feedback. For the next study, we did not consider a condition without visual feedback.

6 MID-AIR INTERACTION STUDY: DRAG & DROP

Selection is in many cases only the starting point of a manipulation.We studied the whole interaction of selecting and moving a virtual object to a different location in this study.

Drag & Drop Techniques

We selected five techniques to drag and drop a virtual target. Four are based directly on selection techniques: *pen drag&drop*, *pen ray pickup*, *one-handed*, and *two-handed*. *Touch&pen* combines touchscreen and pen (see Fig. 7).

Pen Drag&Drop. Continuing the *with highlight* technique, holding the pen button sticks the object to the pen tip. Releasing the button, drops the target at its current location.

Pen Ray Pickup. For the *pen ray* selection there is a depth offset between pen tip and the target. As the user presses and holds the button on the pen, the center of the selected target is snapped to the tip of the pen—similar to [24], but without adjusting the scale of the object to keep its original size. Releasing the button places the target as in *pen drag&drop*.

One-handed & Two-handed Touch Translation. Both the *one-handed* and *two-handed* translation techniques differ only in the hand used to touch the screen—the hand holding the phone for *one-handed*, the hand holding the pen for *two-handed*. Holding the touch attaches the selected object to the phone in its current distance. Changes in position of the phone are directly applied to the position of the virtual object but its rotation in the real world stays the same. Lifting the touch drops the target at its current position.

Touch Pickup & Pen Drop. This two-step condition combines touchscreen and mid-air pointing device. The user first selects the virtual object by touching and holding the object on the screen. Pressing the button on the pen snaps the object to the tip of the pen as soon as the markers are in view. The touch on the screen can now be released and the object is attached to the pen movement. Releasing the button on the pen drops the target.

Study Setup

Inside the upper half of the interaction volume, we placed 32 cubes with edge length of 3 cm using rules similar to the selection study. Participants had to pick up an object and move it to a virtual drop target shown in one of four possible locations. The cube shaped drop targets were shown 5 cm above the corners of the calibration marker with a visual connection to the corner to provide a link to a real world location without being directly tied to a haptic surface. We encourage analyzing interaction techniques that interact with the physical environment.

Moving a target inside the drop target, highlighted the drop location. A correctly dropped target moved to its initial position and the next target and drop location was shown. The order of targets and drop locations was randomized.

Study Procedure

The procedure varied from the selection study only in the task and techniques evaluated. We told the participant that she was allowed to drop and re-select a target as often as

Technique	Selection Time			Translation Time			Task Time		
	Significance	Mean	SD	Significance	Mean	SD	Significance	Mean	SD
two-handed	A	2.6 *s*	0.4 *s*	A B	3.9 *s*	1 *s*	A	6.5 *s*	1.3 *s*
one-handed	A	2.8 *s*	0.5 *s*	A	4 *s*	1 *s*	A	6.8 *s*	1.3 *s*
pen ray pickup	A	2.8 *s*	0.5 *s*	B	3.1 *s*	0.7 *s*	A	5.9 *s*	1.1 *s*
touch&pen	B	4.3 *s*	0.9 *s*	C	2.6 *s*	1.3 *s*	A	6.9 *s*	2 *s*
pen drag&drop	C	5.2 *s*	1.4 *s*	A B	3.7 *s*	1.3 *s*	B	8.9 *s*	2.6 *s*

Table 3: Means and standard deviations of *selection time*, *translation time*, and *task time* for the main effect of *translation technique*. Rows not connected by the same letter are significantly different.

she wanted to. This was explicitly shown for *one-handed* and *two-handed* as pilot studies showed a 'clutching'-style of multiple drag and drop actions to get a target to a comfortable position relative to the phone. After each condition, we asked the participant to rate the technique for *ease of interaction* and *confidence of interaction*. Overall, each participant moved 160 targets (32 cubes×5 techniques). The order of conditions was counterbalanced between participants.

Measurements

We measured the time needed to select and drop each target (*task time*). This time was split into *selection time*, measuring the time until the target was picked up, and *translation time*, recording the total duration the target was moved. If the participant placed the target outside the drop location, the *selection time* increased again.

Since we had observed a clutching technique for *one-handed* and *two-handed*, we recorded how often participants picked up a target during a trial (*pickups*). However, this also records unintended drops.

Furthermore, we recorded how often the participant missed the target during the selection (*misses*).

The subjective ratings were gathered on a 7-level Likert-scale for *ease of interaction* and *confidence of interaction* and as a ranking of the five techniques.

Results

We again recruited 15 participants for this study (2 female, 1 n.a., 22–35 years, M: 27.1 years, SD: 3.6 years, one left-handed). As the phone crashed through a *pen drag&drop* condition, we did not gather data for the remaining targets. Overall, we collected 2383 drag&drop interactions. We averaged measurements for each condition among participants. Similar to the selection study, we log-transformed the time measurements and performed mixed-effect ANOVAs with the user as a random variable. All post-hoc tests were done using the Tukey HSD test. Subjective ratings were analyzed with the Kruskal-Wallis test and post-hoc comparisons using the Wilcoxon method with a Bonferroni correction.

Number of Pickups. Most users did not pick up a target more than once. The average for all conditions is close to 1 (*pen ray pickup* M: 1.05, SD: 0.05; *touch&pen* M: 1.08, SD: 0.09; *one-handed* M: 1.08, SD: 0.1; *pen drag&drop* M: 1.08, SD: 0.11; *two-handed* M: 1.14, SD: 0.17). Analyzing *pickups* shows no significant differences ($F_{4,56} = 2.453, n.s.$).

Number of Missed Selections. The *translation technique* significantly affects the number of *misses* ($F_{4,56} = 6.477, p < .001$). Post-hoc tests show that participants missed less with *pen ray pickup* (M: 0.13, SD: 0.13) compared to all other conditions except *two-handed* (M: 0.23, SD: 0.18) (*one-handed* M: 0.31, SD: 0.21; *touch&pen* M: 0.38, SD: 0.21; *pen drag&drop* M: 0.38, SD: 0.23).

Task Time, Selection Time, and Translation Time. The used *translation technique* had a significant effect on the overall *task time* ($F_{4,56} = 16.71, p < .001$), *selection time* ($F_{4,56} = 59.85, p < .001$), and *translation time* ($F_{4,56} = 14.265, p < .001$). Means and post-hoc results can be seen in Table 3.

Subjective Ratings. Ease of interaction is significantly different based on the *translation technique* ($\chi^2(4) = 18.389, p < .001$). Post-hoc comparisons only indicate differences between *pen ray pickup* (M: 6.4, SD: 1.1) and *touch&pen* (M: 5.4, SD: 1.1) as well as *pen ray pickup* and *pen drag&drop* (M: 4.2, SD: 1.8). The other conditions show no significant differences (*one-handed* M: 5.7, SD: 1; *two-handed* M: 5.3, SD: 1.3).

Similarly, for *confidence of interaction* there are also significant differences ($\chi^2(4) = 16.088, p < .005$). Post-hoc tests indicate differences only for *pen ray pickup* (M: 6.4, SD: 0.9) and *pen drag&drop* (M: 4.5, SD: 1.7). The other conditions do not differ significantly (*one-handed* M: 5.7, SD: 0.9; *two-handed* M: 5.5, SD: 1; *touch&pen* M: 5.3, SD: 1.2).

The ranking of conditions shows a preference for *pen ray pickup* before *touch&pen*. *Pen drag&drop* was mostly placed last. *One-handed* and *two-handed* are generally placed in third and fourth position (Figure 8).

Discussion

The translation techniques performed similar based on *task time* except for *pen drag&drop* which took longer compared to all other techniques because of the increased selection time. Participants also liked this technique the least.

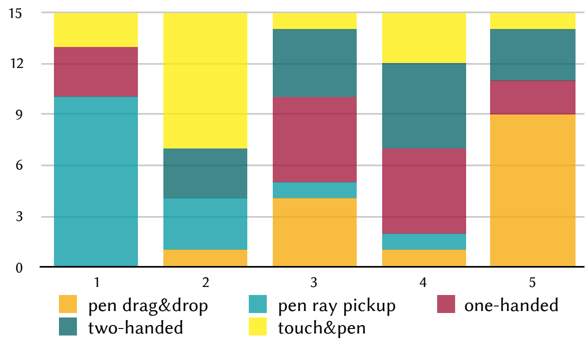

Figure 8: Subjective Ranking of translation techniques. *Pen ray pickup* is preferred before *touch&pen*. *Pen drag&drop* seems to be least preferred while *one-handed* and *two-handed* are equally distributed at third and fourth place.

Pen ray pickup was preferred by the participants and they missed also less targets with this technique indicating, similar to the selection study, that this technique should be used for mid-air selection and pickup.

Even though *touch&pen* needed more time for the selection, the translation time was significantly lower to all other conditions and participants ranked this technique only behind *pen ray pickup*. Several participants used a strategy where they left the hand holding the pen on the desk, pressing the button as the other hand touched the object. Moving the viewport to the table snapped the object to the pen which only had to complete a small translation. Participants mentioned that this enabled are more comfortable hand position. As this was only possible because the drop locations were located on the table, further investigation of this technique is necessary to judge its performance for unknown and/or mid-air drop locations.

The touch techniques had a similar performance and ranking regardless whether one or two hands were used.

The results indicate that a placement via the pen is beneficial but the pickup needs support to overcome depth issues.

7 DISCUSSION

Our studies with the ARPen show that mid-air selection of virtual objects is not optimal as it takes longer to find the correct 3D position and is not possible without highlighting or pen tip visualization. However, placement by moving the pen to the intended position ranked high on user preference.

Over both mid-air manipulation studies, users preferred the selection and pickup via a ray shot through the pen tip. The 2D representation of the scene allowed to align the pen tip with the object to select it. This interaction also achieved the highest amount of successful selections, short selection and translation times comparable to touch techniques, and least device movement based on the size of the projected targets. We recommend to employ ray-based techniques for manipulation tasks in handheld AR modeling systems.

Touch techniques, especially with two hands, achieve good results particularly for selection time. For selection, *two-handed* is comparable to *pen ray* in user preference followed by *one-handed*. However, the high projected size of the targets suggests that users have to move the device more through the scene. For translation, users preferred *pen ray pickup* and *touch&pen* to the touch techniques.

Touch&pen shows that the combination of touchscreen and mid-air pointing device has potential for the interaction. The possibility to select an object mid-air via touch and then place it near the table without having to continuously lift the pen allowed users to rest their arm providing a more ergonomic interaction. This is particularly important since users mentioned fatigue in their arms during the studies.

8 SUMMARY & FUTURE WORK

We present the ARPen, a bimanual input technique for AR modeling combining a standard smartphone and a 3D-printed pen. Studying how users hold the phone while drawing, we found that a landscape orientation leaves an area on the left side of the screen for interaction and shows the least issues. We evaluated the performance of different selection and translation techniques and saw that users preferred a technique which casts a ray through the pen tip. Selecting an object mid-air by holding the pen inside the object showed a high success rate but took significantly longer. For translation, a technique combining touchscreen and pen was rated as the second preference. We encourage further investigation of such combinations.

Further studies on the ARPen could take in account physical surfaces or different depth visualization techniques as well as new rotation and scaling techniques. Finding a consistent set of techniques for example by switching different modes via the touchscreen, is a promising direction.

Adding haptic feedback such as vibrations when contacting virtual objects could improve the pen as well as the option to draw on the touchscreen as in [1].

To simplify 3D model generation via mid-air modeling, future work could also use the ARPen to prototype new geometry generation techniques and see how users create and design objects in different contexts such as at the desk or in the car. Also, different user groups might use this tool in different ways—not just for modeling but for communicating 3D ideas in a brainstorming session.

ACKNOWLEDGMENTS

We want to thank all our participants for taking part in the studies. This work was funded by the German Federal Ministry of Education and Research as part of the Photonics Research Germany (13N14065).

REFERENCES

[1] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ. In *Proc CHI '18*. ACM, 185:1–185:15. https://doi.org/10.1145/3173574.3173759

[2] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR. In *Proc CHI '17*. ACM, 5643–5654.

[3] Teo Babic, Harald Reiterer, and Michael Haller. 2018. Pocket6: A 6DoF Controller Based On A Simple Smartphone Application. In *Proceedings of the Symposium on Spatial User Interaction (SUI '18)*. ACM, New York, NY, USA, 2–10. https://doi.org/10.1145/3267782.3267785

[4] Seok-Hyung Bae, Ravin Balakrishnan, and Karan Singh. 2008. ILoveSketch: as-natural-as-possible sketching system for creating 3d curve models. In *Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08*. ACM Press. https://doi.org/10.1145/1449715.1449740

[5] Huidong Bai, Gun A. Lee, and Mark Billinghurst. 2012. Freeze View Touch and Finger Gesture Based Interaction Methods for Handheld Augmented Reality Interfaces. In *Proceedings of the 27th Conference on Image and Vision Computing New Zealand (IVCNZ '12)*. ACM, New York, NY, USA, 126–131. https://doi.org/10.1145/2425836.2425864

[6] Huidong Bai, Gun A. Lee, Mukundan Ramakrishnan, and Mark Billinghurst. 2014. 3D Gesture Interaction for Handheld Augmented Reality. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications (SA '14)*. ACM, New York, NY, USA, Article 7, 6 pages. https://doi.org/10.1145/2669062.2669073

[7] Oriel Bergig, Nate Hagbi, Jihad El-Sana, and Mark Billinghurst. 2009. In-place 3D Sketching for Authoring and Augmenting Mechanical Systems. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09)*. IEEE Computer Society, Washington, DC, USA, 87–94. https://doi.org/10.1109/ISMAR.2009.5336490

[8] Mark Billinghurst, Tham Piumsomboon, and Huidong Bai. 2014. Hands in Space: Gesture Interaction with Augmented-Reality Interfaces. *IEEE Computer Graphics and Applications* 34, 1 (Jan 2014), 77–80. https://doi.org/10.1109/MCG.2014.8

[9] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice.* Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.

[10] Bill Buxton and George W. Fitzmaurice. 1998. HMDs, Caves & Chameleon: A Human-centric Analysis of Interaction in Virtual Space. *SIGGRAPH Comput. Graph.* 32, 4 (Nov. 1998), 69–74. https://doi.org/10.1145/307710.307732

[11] Arindam Dey, Graeme Jarvis, Christian Sandor, and Gerhard Reitmayr. 2012. Tablet versus phone: Depth perception in handheld augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 187–196. https://doi.org/10.1109/ISMAR.2012.6402556

[12] Leo Gombač, Klen Čopič Pucihar, Matjaž Kljun, Paul Coulton, and Jan Grbac. 2016. 3D Virtual Tracing and Depth Perception Problem on Mobile AR. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1849–1856. https://doi.org/10.1145/2851581.2892412

[13] Nate Hagbi, Raphaet Grasset, Oriel Bergig, Mark Billinghurst, and Jihad El-Sana. 2010. In-Place Sketching for Content Authoring in Augmented Reality Games. In *Proceedings of the 2010 IEEE Virtual Reality Conference (VR '10)*. IEEE Computer Society, Washington, DC, USA, 91–94. https://doi.org/10.1109/VR.2010.5444806

[14] Anders Henrysson, Mark Billinghurst, and Mark Ollila. 2005. Virtual Object Manipulation Using a Mobile Phone. In *Proceedings of the 2005 International Conference on Augmented Tele-existence (ICAT '05)*. ACM, New York, NY, USA, 164–171. https://doi.org/10.1145/1152399.1152430

[15] Ke Huo, Vinayak, and Karthik Ramani. 2017. Window-Shaping: 3D Design Ideation by Creating on, Borrowing from, and Looking at the Physical World. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (TEI '17)*. ACM, New York, NY, USA, 37–45. https://doi.org/10.1145/3024969.3024995

[16] Wolfgang Hürst, Ronald Poppe, and Jerry van Angeren. 2015. Drawing outside the lines: Tracking-based gesture interaction in mobile augmented entertainment. In *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*. 79–87.

[17] Wolfgang Hürst and Casper van Wezel. 2013. Gesture-based interaction via finger tracking for mobile augmented reality. *Multimedia Tools and Applications* 62, 1 (Jan 2013), 233–258. https://doi.org/10.1007/s11042-011-0983-y

[18] Takeo Igarashi, Satoshi Matsuoka, and Hidehiko Tanaka. 1999. Teddy: a sketching interface for 3D freeform design. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99*. ACM Press. https://doi.org/10.1145/311535.311602

[19] Paul Issartel, Florimond Guéniat, and Mehdi Ammi. 2014. Slicing techniques for handheld augmented reality. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 39–42. https://doi.org/10.1109/3DUI.2014.6798839

[20] Bret Jackson and Daniel F. Keefe. 2016. Lift-Off: Using Reference Imagery and Freehand Sketching to Create 3D Models in VR. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (April 2016), 1442–1451. https://doi.org/10.1109/tvcg.2016.2518099

[21] Ernst Kruijff, J. Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. 3–12. https://doi.org/10.1109/ISMAR.2010.5643530

[22] David Lakatos, Matthew Blackshaw, Alex Olwal, Zachary Barryte, Ken Perlin, and Hiroshi Ishii. 2014. T(Ether): Spatially-aware Handhelds, Gestures and Proprioception for Multi-user 3D Modeling and Animation. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction (SUI '14)*. ACM, New York, NY, USA, 90–93. https://doi.org/10.1145/2659766.2659785

[23] Gun A. Lee, Ungyeon Yang, Yongwan Kim, Dongsik Jo, Ki-Hong Kim, Jae Ha Kim, and Jin Sung Choi. 2009. Freeze-Set-Go Interaction Method for Handheld Mobile Augmented Reality Environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (VRST '09)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1643928.1643961

[24] Joon Hyub Lee, Sang-Gyun An, Yongkwan Kim, and Seok-Hyung Bae. 2018. Projective Windows: Bringing Windows in Space to the Fingertip. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 218, 8 pages. https://doi.org/10.1145/3173574.3173792

[25] Asier Marzo, Benoît Bossavit, and Martin Hachet. 2014. Combining Multi-touch Input and Device Movement for 3D Manipulations in Mobile Augmented Reality Environments. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction (SUI '14)*. ACM, New York, NY, USA, 13–16. https://doi.org/10.1145/2659766.2659775

[26] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D (1994). http://vered.rose.utoronto.ca/people/paul_dir/IEICE94/ieice.html

[27] Bojan Milosevic, Flavio Bertini, Elisabetta Farella, and Serena Morigi. 2016. A SmartPen for 3D interaction and sketch-based surface modeling. *The International Journal of Advanced Manufacturing Technology* 84, 5 (01 May 2016), 1625–1645. https://doi.org/10.1007/s00170-015-7828-1

[28] Annette Mossel, Benjamin Venditti, and Hannes Kaufmann. 2013. 3DTouch and HOMER-S: Intuitive Manipulation Techniques for One-handed Handheld Augmented Reality. In *Proceedings of the Virtual Reality International Conference: Laval Virtual (VRIC '13)*. ACM, New York, NY, USA, Article 12, 10 pages. https://doi.org/10.1145/2466816.2466829

[29] Annette Mossel, Benjamin Venditti, and Hannes Kaufmann. 2013. DrillSample: Precise Selection in Dense Handheld Augmented Reality Environments. In *Proceedings of the Virtual Reality International Conference: Laval Virtual (VRIC '13)*. ACM, New York, NY, USA, Article 10, 10 pages. https://doi.org/10.1145/2466816.2466827

[30] Jon Peddie. 2017. *Augmented Reality - Where We Will All Live*. Springer. https://doi.org/10.1007/978-3-319-54502-8

[31] Huaishu Peng, Jimmy Briggs, Cheng-Yao Wang, Kevin Guo, Joseph Kider, Stefanie Mueller, Patrick Baudisch, and François Guimbretière. 2018. RoMA: Interactive Fabrication with Augmented Reality and a Robotic 3D Printer. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 579, 12 pages. https://doi.org/10.1145/3173574.3174153

[32] Jarkko Polvi, Takafumi Taketomi, Goshiro Yamamoto, Arindam Dey, Christian Sandor, and Hirokazu Kato. 2016. SlidAR: A 3D positioning method for SLAM-based handheld augmented reality. *Computers & Graphics* 55 (2016), 33 – 43. https://doi.org/10.1016/j.cag.2015.10.013

[33] Emanuel Sachs, Andrew Roberts, and David Stoops. 1991. 3-Draw: a tool for designing 3D shapes. *IEEE Computer Graphics and Applications* 11, 6 (Nov 1991), 18–26. https://doi.org/10.1109/38.103389

[34] Hartmut Seichter. 2003. Sketchand+ a Collaborative Augmented Reality Sketching Application. In *Proceedings of the 8th International Conference on Computer Aided Architectural Design Research in Asia (CAADRIA '03)*. 209–222.

[35] Iris Seidinger and Jens Grubert. 2016. 3D Character Customization for Non-Professional Users in Handheld Augmented Reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 129–134. https://doi.org/10.1109/ISMAR-Adjunct.2016.0059

[36] Michael Tsang, George W. Fitzmaurice, Gordon Kurtenbach, Azam Khan, and Bill Buxton. 2003. Boom Chameleon: Simultaneous Capture of 3D Viewpoint, Voice and Gesture Annotations on a Spatially-aware Display. In *ACM SIGGRAPH 2003 Papers (SIGGRAPH '03)*. ACM, New York, NY, USA, 698–698. https://doi.org/10.1145/1201775.882329

[37] Klen Čopič Pucihar, Paul Coulton, and Jason Alexander. 2013. Evaluating Dual-view Perceptual Issues in Handheld Augmented Reality: Device vs. User Perspective Rendering. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI '13)*. ACM, New York, NY, USA, 381–388. https://doi.org/10.1145/2522848.2522885

[38] Thomas Vincent, Laurence Nigay, and Takeshi Kurata. 2013. Precise Pointing Techniques for Handheld Augmented Reality. In *Human-Computer Interaction – INTERACT 2013*. Springer Berlin Heidelberg, Berlin, Heidelberg, 122–139.

[39] Philipp Wacker, Adrian Wagner, Simon Voelker, and Jan Borchers. 2018. Physical Guides: An Analysis of 3D Sketching Performance on Physical Objects in Augmented Reality. In *Proceedings of the Symposium on Spatial User Interaction (SUI '18)*. ACM, New York, NY, USA, 25–35. https://doi.org/10.1145/3267782.3267788

[40] Po-Chen Wu, Robert Wang, Kenrick Kin, Christopher Twigg, Shangchen Han, Ming-Hsuan Yang, and Shao-Yi Chien. 2017. DodecaPen: Accurate 6DoF Tracking of a Passive Stylus. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 365–374. https://doi.org/10.1145/3126594.3126664

[41] Min Xin, Ehud Sharlin, and Mario Costa Sousa. 2008. Napkin sketch: handheld mixed reality 3D sketching. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology - VRST '08*. ACM Press. https://doi.org/10.1145/1450579.1450627

[42] Brandon Yee, Yuan Ning, and Hod Lipson. 2009. Augmented Reality In-situ 3D Sketching of Physical Objects. In *Intelligent UI workshop on sketch recognition (IUI '09)*.

[43] Sang Ho Yoon, Ke Huo, and Karthik Ramani. 2016. TMotion: Embedded 3D Mobile Input Using Magnetic Sensing Technique. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16)*. ACM, New York, NY, USA, 21–29. https://doi.org/10.1145/2839462.2839463

[44] Ya-Ting Yue, Xiaolong Zhang, Yongliang Yang, Gang Ren, Yi-King Choi, and Wenping Wang. 2017. WireDraw: 3D Wire Sculpturing Guided with Mixed Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press. https://doi.org/10.1145/3025453.3025792