Heatmaps, Shadows, Bubbles, Rays: Comparing Mid-Air Pen Position Visualizations in Handheld AR

Philipp Wacker **Adrian Wagner** **Simon Voelker** **Jan Borchers**
RWTH Aachen University
52056 Aachen, Germany
{wacker,wagner,voelker,borchers}@cs.rwth-aachen.de

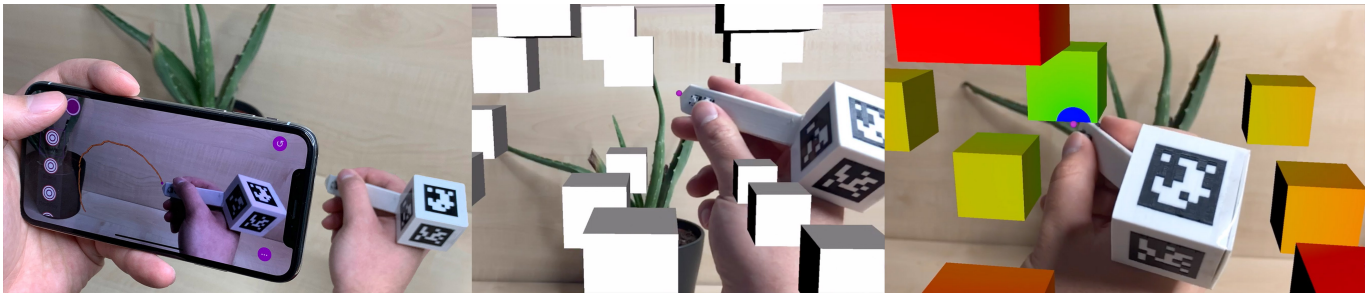

Figure 1. Handheld Augmented Reality with a mid-air pen opens up new interaction possibilities with virtual content (left). However, perceptual issues make it difficult to judge the position of the pen in relation to the virtual environment (middle). We designed and compared visualization techniques and found that a 'heatmap' visualization that colors objects based on their distance to the pen achieves good results and is preferred by users (right).

ABSTRACT
In Handheld Augmented Reality, users look at AR scenes through the smartphone held in their hand. In this setting, having a mid-air pointing device like a pen in the other hand greatly expands the interaction possibilities. For example, it lets users create 3D sketches and models while on the go. However, perceptual issues in Handheld AR make it difficult to judge the distance of a virtual object, making it hard to align a pen to it. To address this, we designed and compared different visualizations of the pen's position in its virtual environment, measuring pointing precision, task time, activation patterns, and subjective ratings of helpfulness, confidence, and comprehensibility of each visualization. While all visualizations resulted in only minor differences in precision and task time, subjective ratings of perceived helpfulness and confidence favor a 'heatmap' technique that colors the objects in the scene based on their distance to the pen.

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** *Mixed / augmented reality; Interaction techniques;*

Author Keywords
Augmented Reality; mid-air; modeling; interaction; depth perception; smartphone; 3D pen; depth cues

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376848

INTRODUCTION
Mid-air modeling in Virtual or Augmented Reality (VR/AR) has been an active research area for several decades (e.g., [2, 4, 38, 45, 51]). Especially AR offers many beneficial aspects for 3D modeling. For Personal Fabrication, for example, it enables designing and viewing 3D objects at the location where they are intended to be used later after 3D printing them. The physical properties of real world objects can be used to assist during the interaction: A surface or edge of a physical object can guide a stroke with a physical pen in 3D. This has been shown to be an important factor to improve the performance of mid-air modeling tasks in virtual environments [3, 46]. Virtual objects can also serve as a starting point for the modeling of new objects. For example, when using AR to display virtual furniture [25], mid-air modeling can be used to model a cutout in a virtual kitchen displayed in AR.

While most AR research to date has been carried out using head-mounted displays, recent advances in mobile sensors and computing power have made AR modeling on consumer smartphones possible, leading to Handheld AR applications [45]. Removing the need for a large tracking setup, Handheld AR provides interesting opportunities for quick modeling tasks, and several research projects investigate its potential (e.g., [22, 31, 40, 42, 45, 51]). As an example, we described a scenario that uses real objects for guidance while designing a cup holder in a car [45]. An example to combine virtual

models could be the goal to design a slide between two virtual marble runs.

However, perceiving the depth information of a virtual object in AR and VR is problematic [30]. This becomes more pronounced in Handheld AR, where 3D information is presented on a 2D screen and the camera angle can differ from the user's viewing angle [12, 20, 44]. This creates issues for modeling tasks in Handheld AR, because even though it is relatively simple to find a specific location in the real world and move the pen to it, it is considerably harder to find that spot if the object to target is virtual. For selection tasks, this leads to raycasting methods being preferred, which select the first object that is hit by a ray shot from a touchpoint on the smartphone's screen or through the tip of the mid-air pen [45].

At the same time, actually finding and moving the pen to a specific location is a central task for mid-air modeling systems. For example, to sketch a line from one virtual object to another, a user first has to find the correct spot on the virtual object from where she wants to draw the line. While raycasting techniques could place a remote cursor at the intersection and map pen movement onto that cursor, this approach takes away a key advantage of AR modeling, because now the position of the real world pen does not align with the position of the virtual pen anymore, and haptic properties of the physical environment can no longer be used for guidance.

To tackle this fundamental problem of mid-air interaction in Handheld AR, we defined and implemented different visualization techniques that show the position of the pen in relation to surrounding objects, and compared them to a baseline condition with no such visualization. We ran our comparisons in scenes with both solid and wireframe objects. Based on our results, we provide design considerations for visualizing the 3D position of a mid-air pointing device in Handheld AR.

Our contributions in this paper therefore are:

- the definition and comparison of different visualization techniques to show the 3D position of a mid-air pointing device in relation to surrounding objects in Handheld AR;
- design implications accounting for the advantages and disadvantages of these visualizations.

Following this introduction, we give an overview of related work in immersive modeling and depth perception in virtual environments. We then present the different visualization styles we designed, and how we implemented them followed by our study setup and procedure. After summarizing the results of our study and discussing their design implications, we conclude with an outlook on future work.

RELATED WORK

Mid-air modeling, the process of creating virtual models by moving a physical pointing device through mid-air, has been a focus of research for almost 30 years [4, 10, 38]. Placed at all locations along the mixed-reality continuum [34], many projects focus on modeling in Virtual Reality (e.g., [27]). However, in Augmented Reality, virtual objects are placed in the real world, making it possible, for example, to see virtual models interact with real objects [2] or, in the area of Personal Fabrication, to see and design objects at the same scale and location they are meant to be used after 3D printing them later [36]. Especially for modeling tasks, being able to use the physical properties of the real world can assist in modeling, and improve input precision [3, 46].

While many systems to date use head-mounted AR devices, Handheld AR offers increased portability, and commercial AR frameworks[1] have boosted the development of Handheld AR apps (see [11] for a survey on usability studies in AR and [19] for a review of object manipulation techniques in Handheld AR). Modeling and interacting with the virtual environment in Handheld AR is often done by interacting with the touchscreen of the device [22, 35, 51] or by tracking mid-air input within the 3D space the camera is capturing [23, 42, 45]. Using a tracking technique similar to the DodecaPen [50], we previously developed the ARPen, a bimanual Handheld AR system that tracks the position of a mid-air pen to provide continuous mid-air input in Handheld AR [45]. In that work, we performed basic studies comparing different techniques to select and translate virtual objects. We found raycasting techniques to perform well for selection, regardless of whether the ray is cast from a touch on the screen or from the camera through the pen tip.

However, Augmented Reality—and in particular Handheld Augmented Reality—has several perceptual issues [16, 30]. An often reported issue in both Augmented and Virtual Reality is that users generally underestimate distances (e.g., [1, 14, 48], and many studies have been carried out to investigate these perceptual issues and how to address them for Augmented Reality systems (see [11, 13, 24] for surveys).

The perception of depth in virtual environments depends on several *depth cues* [6, 9, 16]. These can be *physiological*, *kinetic*, and *pictorial*. *Physiological* cues refer to depth perception based on differences in images perceived by each eye, and many studies have evaluated depth perception in stereoscopic systems (e.g., [18, 29, 33]). However, as Handheld AR offers only one image on the screen, the camera image displayed cannot provide stereoscopic depth cues; the user instead receives cues through her own stereoscopic vision about how the entire smartphone screen itself is positioned in space.

Kinetic depth cues refer to the movement of both the viewing device and objects in the scene. One example is *motion parallax*: differences in perceived movement speeds and directions of objects depending on their distance to the viewer [9, 16].

Finally, depth perception in Handheld AR can use *pictorial* cues known from painting and photography. Cutting [8] mentions five pictorial depth cues (occlusion, relative size, relative density, height in the visual field, and aerial perspective), but also notes that their quality depends on the distance to the viewer. Cutting classifies this distance into the *near* or *personal* space, the *middle* or *action* space, and the *far* or *vista* space [8]. The near space describes the area within arm's reach of the person, and is the one of interest to our work. In this space, occlusion, relative size, and relative density are

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

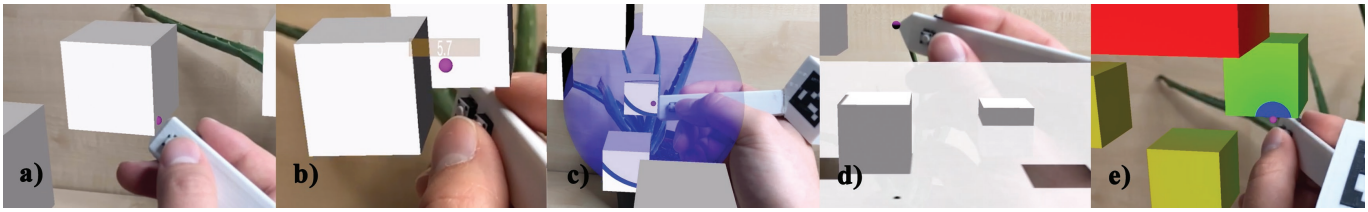

Figure 2. The visualization techniques used in our study. a) *minVis* as the baseline, b) *depth ray*, c) *bubble*, d) *shadow*, e) *heatmap*.

helpful depth cues. Current Handheld AR systems already imitate these to achieve the illusion of virtual objects at different locations in the real world. Occlusion is particularly relevant for X-ray applications that allow users to see through physical objects (e.g., [12, 13, 43]). Such applications are often used to show buildings or structures behind other buildings. These visualizations could either completely remove an area of the occluding structure to show the area behind it, or keep parts of the foreground to preserve context [13]. Studies on X-ray visualizations are mainly focussed on showing the existence and depth ordering of objects rather than the exact distances between them. For this reason, they are not directly applicable to the question addressed in this paper.

Most studies on depth cues focus on stereoscopic systems (e.g., [18, 29, 33]). Those dealing with handheld AR systems tend to look at farther distances, for applications like X-ray viewers or AR browsers (e.g., [13, 24, 49]). Wither et al. [49] compared different depth cues in a headworn AR environment for distances of more than 20 meters. They compared different visualizations in which the size of a cursor varied with the distance, shadow planes showed the layout of objects, a top-down overview of the scene was provided, or the color of objects in the scene changed based on the current cursor position. They found that other objects in the scene improve depth accuracy, but the visualizations themselves did not improve accuracy as much as expected—something they attribute in part to the visualization system used. The top-down view was preferred by users, and performed well, especially for the nearer objects in their study. The color visualization also received positive user responses. Other studies showed that one of the most effective cues to indicate the depth of a virtual object is displaying its shadow [14, 16, 21, 47].

Studies of depth perception in virtual environments differentiate between *egocentric* and *exocentric* depth perception (e.g., [12, 13, 41]. Egocentric depth perception measures the distance of an object from the viewer, while exocentric depth perception looks at the distance between different objects in the scene. Most research projects so far have focused on egocentric depth perception [12, 28].

In our study, we look at a task for which both egocentric and exocentric depth perception are important. To reach a specific point on a virtual object with her mid-air pen, a user first has to evaluate the depth of the object relative to her own position, and then continuously evaluate the position of the mid-air pen tip relative to the target position to reach the intended location in space. Our study is the first to evaluate different visualization techniques for such a task using a Handheld AR system with a mid-air pointing device.

VISUALIZATION STYLES

We define five visualization styles to indicate the 3D position of the pen tip in relation to its environment. Our baseline is a minimal visualization without any extra depth information, while four depth visualizations build on this baseline to provide more information for the depth dimension.

Minimal Visualization

In our previous selection study [45], we had found that displaying no indication of the pen tip does not allow participants to select objects in a 3D scene: only 3% of targets were successfully selected. This indicated that it would not enable users to select a point on a virtual object either. We had also found that rendering a small sphere at the tip of the tracked pen improved selection performance. Like every other virtual object in the scene, this sphere was not rendered if inside or behind another virtual object. Other than that, it did not provide additional information about the position of the pen. For our present work, we define this visualization as the minimum viable visualization (*minVis*) to estimate the position of the pen in the scene (Fig. 2, a).

We use the depth cue of occlusion only for the tip of the pen, not for the whole pen and hand holding it. One reason for this is that current smartphones cannot reliably track moving physical objects in the scene to calculate how they occlude virtual objects. While this technical limitation could be overcome with a lab study, modeling and rendering the hand would also occlude a large part of the virtual information on the smartphone's small screen, and the effects of this on scene perception and the overall user experience will need to be studied first. Rendering only the pen without the hand into the scene would be easier to achieve technically, but the perception of ownership is affected by the style of visualization [37]. Since the pen would be rendered in front of the hand holding it, "cutting off" its thumb, it would create an unnatural visualization that would likely affect the sense of *holding* the device [39]. Until perceptual issues such as these are answered and hand-tracking capabilities increase, we decided to focus on the visualization of the position of the pen tip for this work. We present potential future research directions exploring these issues at the end of this paper. Not relying on a visualization of pen or hand, however, also makes our results easier to apply to other input devices that support the fundamental interaction of specifying a mid-air position in Handheld AR.

All following techniques build up from this minimum viable visualization. For each visualization, we provided users with a way to toggle it on and off using a button on the pen. This way, the additional visualization could be brought up only

when needed, without cluttering the scene otherwise. This is important to keep in mind when interpreting our findings.

Depth Ray

Since ray casting methods have shown good results for selecting objects [45], we created a method based on the ray from the camera through the tip of the pen into the scene. We display the distance of the pen tip to the next object behind it in cm (*depth ray*; Fig. 2, b). This lets users judge the distance of the pen to the objects behind it, which, for the problematic task of moving the pen to the correct 3D position, can help to decide whether large or fine movements in the depth direction are needed to reach the intended position. However, this visualization does not give any additional information in other directions than behind the pen tip.

Bubble

To show distance information in all directions, we defined a bubble visualization. When triggered, a semi-transparent bubble linearly grows outward from the current pen tip position over one second, while remaining centered around the pen tip (*bubble*; Fig. 2, c). During the increase in size, the intersections of the bubble shell with objects in the scene are highlighted. This means that objects farther away from the center of the bubble are intersected later than objects closer to the center and the intersections indicate a larger size of the bubble. Other than that, the bubble behaves like other virtual object in the scene, and is therefore occluded by objects in front of its shell and occludes objects behind its shell.

Shadow

Previous studies have shown that shadows improve depth perception in virtual environments (e.g., [16, 14, 47]). However, the viewing area in Handheld AR is very limited, and the next surface to project shadows on, such as the table surface, is often outside the camera image, so seeing the shadow would require moving the smartphone camera. Therefore, we chose to display the shadows on an artificial horizontal plane rendered slightly lower than the current viewing height (*shadow*; Fig. 2, d). The light source is placed above the scene, so that all items cast their shadow downwards onto each other or the artificial plane. This includes the small sphere at the pen tip.

Heatmap

To show the mid-air position of the pen relative to all objects in the scene, a *heatmap* visualization can be used (Fig. 2, e). Similarly to the "Marker Color" technique by Wither and Höllerer [49], it shades all objects by replacing their original color with a color indicating their distance to the pen tip. For our naive implementation we linearly transitioned from full green to full red in RGB space over 200 mm. Far away objects are colored in red, and the shading gradually changes to green the closer the object is to the pen tip. In addition, we colored surface areas within 10 mm of the pen tip in blue to indicate the closest surfaces. We checked the visualization with a red-green color vision impaired user to confirm that the difference between the red and green color was strong enough. A linear gradient between colors in RGB space is not perceived as linear by humans [32] but it can provide us with a general impression of the effectiveness of such a visualization by showing if users are able to apply the mapping that red is farther away from the target than green.

Aside from the above five techniques, we designed and implemented several other visualizations, such as displaying grid lines, adding light sources to the pen tip or camera to light the scene, and rendering lines between the pen tip and a number of closest objects. However, pilot studies showed that these techniques were harder to understand and interpret than the visualization techniques we selected for this study.

STUDY OF VISUALIZATION STYLES

In order to test the performance of our proposed visualization styles, we compared them to the baseline condition that uses minimal visualization (*minVis*). Many 3D pointing operations require identifying the distance of the input device to its surroundings, e.g., to arrange objects or to create a connection between them. We tested the case of finding the correct location on a virtual object in mid-air. This task setup is basic enough to be applicable to many other settings since the perception issues on mobile devices stay the same. All of our techniques could be applied to a physical environment and physical objects as well if all the elements in the scene are tracked. The performance of each technique would likely increase then, because physical effects, such as peripheral vision around the device or haptic feedback from objects, could be used to support moving to the intended location.

As mid-air modeling systems can feature both solid objects and wireframe models (or sketched lines), we included both cases, *solid* and *wireframe*.

Study Setup

We implemented our visualizations for our open-source ARPen system [45][2]. Users can toggle the current visualization on or off by pressing a button on the pen—our pilot users preferred this to having to keep holding the button down to see the visualization.

The task for each trial in our study was to move the pen to a specific location on a virtual cube object and start drawing a line from that location. The interaction volume for our study was $400\times400\times400$ mm in front of the participant, so that the whole area was in arm's reach. To keep the position of the interaction volume constant between participants, we used a marker on the table to fix its position. We separated this volume into $3\times3\times3 = 27$ areas, and placed a virtual cube into each area, with an edge length of either 30 or 40 mm to prevent the depth cue of *relative size* to influence our task. Cube position in its area was randomized for each new condition, while ensuring a gap of at least 20 mm between two adjacent cubes. For each trial, we showed 8 cubes and marked a location on a target cube to indicate where to move the pen to. Since this location should be visible to the user, we first excluded cube sides that were invisible when looking from a centered position in front of the interaction volume—this included back-facing sides for all cubes and, e.g., the right-facing side of a cube on the right side of the interaction

[2] https://github.com/i10/ARPen

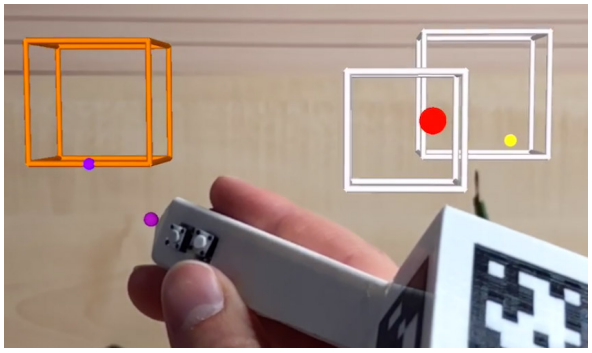

Figure 3. *Wireframe* objects in the study. Before starting the trial, participants had to move the smartphone to align the yellow and red spheres. The target location on the cube was indicated by a purple sphere.

volume. On a random remaining side, we chose a random position as the target position for this trial. In the case of *wireframe* objects, this position was limited to the edges of the surface. We indicated this position with a purple sphere clearly visible against the rest of the cube. Participants had to reach this sphere with the pen as precisely as possible, and start drawing a line from this position. This line was a placeholder for an action in a real modeling task; so its shape was not relevant for our evaluation.

To control the starting position for each trial, we added two spheres to the scene, one at the center of the near side, one at the center of the far side of the interaction volume (Fig. 3). Participants had to align these spheres on the screen and lift the pen into the camera view to start each new trial.

Study Procedure

Participants sat in front of a table with the marker. They were asked to hold the device in their non-dominant hand in the *pinkie* grip [45]: resting on the pinkie and gripped by the index finger (see Fig. 1 left). Participants could familiarize themselves with the grip and drawing with the mid-air pen before we introduced the task and the visualization techniques. Before each new technique, we explained how it works, and let them familiarize themselves with it for up to 3 minutes. Once they were confident to continue, we started recording the trials. We chose the cube with the target marking randomly from the 27 cubes, and repeated the trial 16 times, making sure that no cube was selected twice as the target cube. After completing the 16 repetitions for a condition, we asked participants to rate their confidence in the ability to find the intended spot, the helpfulness of the visualization, and how easy it was to comprehend. While they were completing the trials for a visualization technique, the moderator captured qualitative comments and observations about the interaction. After using all five different visualization techniques, we asked participants to rank them by preference from best to worst. Each participant selected 160 positions (5 visualization techniques (*minVis*, *depth ray*, *bubble*, *shadow*, *heatmap*) $\times$ 2 object styles (*solid*, *wireframe*) $\times$ 16 repetitions). The order of visualization techniques and object styles was counterbalanced using a Latin square.

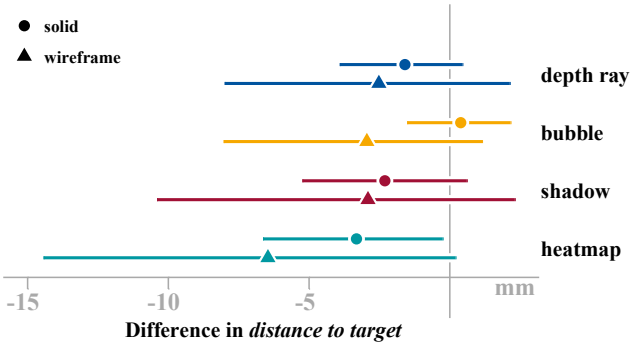

Figure 4. Difference between the visualization techniques and the baseline regarding the distance to the target (*minVis* M: 9.21 mm; CI [3.87, 15.86]). Negative values indicate smaller distances to the target compared to the baseline. In general, visualization techniques, esp. *heatmap*, seem to improve accuracy. Only *bubble solid* shows a tendency towards longer distances. Whiskers denote the 95% confidence interval (CI).

Measurements

For each trial, we recorded how far the initial drawing position deviated from the intended position indicated by the marking (*distanceToTarget*). We also measured the time from the beginning of the trial to the beginning of the first drawing operation (*timeToTarget*). To evaluate how often the visualizations were used to find the intended position, we recorded how long the visualization was active, and from this calculated relative duration (*visualizationPercentage*). The times before aligning the spheres or after completing a trial were not recorded to give the participants time for comments. If participants needed a rest or stopped during a movement task, e.g., to provide a longer comment, we repeated the last trial. This happened 49 times in our study.

For confidence, helpfulness, and comprehensibility, participants rated the techniques on a 7-point Likert-scale, and ranked them from best to worst.

Evaluation

We recruited 10 participants (4 female, 21–28 years, M: 25 years, SD: 1.8 years, all right-handed, no self-reported color-vision deficiencies). Overall, we recorded information for 1599 finding operations, as one trial was not recorded due to an issue with the device. For every participant, visualization, and object style, we averaged the deviations (*distanceToTarget*), time to move to the intended location (*timeToTarget*), and percentage of time the visualization was active (*visualizationPercentage*). We evaluated the results based on the Fair Statistical Communication Guidelines by Dragicevic [15] and the Transparent Statistics Working Group [26]. This means that we do not present p-values or dichotomous decisions on significance or non-significance, but rather use *estimation* to account for the uncertainty in the evaluation [7]. For this, we present effect sizes with 95% confidence intervals. These effect sizes show the differences between the measurements of the visualization techniques and the baseline condition (*minVis*). For each measurement, we subtracted the corresponding value of the baseline for the same object style. Consequently, in the following graphs a value to the right of 0 indicates more of the measured variable for the visualization technique

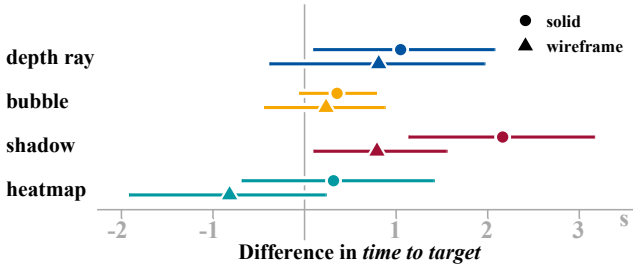

Figure 5. Difference between the visualization techniques and the baseline regarding the time to move to the target (*minVis* M: 7.16 s; CI [6.16, 8.18]). Especially *shadow* seems to take more time compared to the baseline, while *heatmap* does not show great differences and even a tendency to be faster for *wireframe* objects. Whiskers denote the 95% CI.

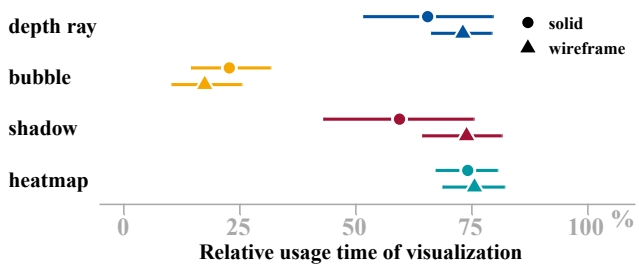

Figure 6. Percentage that the visualization technique was active. *Bubble* was active the least amount of time, but there does not seem to be a difference between the other conditions. For *depth ray* and *shadow*, it seems as if the visualization is used less in a scene with *solid* objects. Whiskers denote the 95% CI.

compared to the baseline, while a value to the left indicates less. The 95% confidence intervals for these differences were calculated using bootstrapping with 1000 repetitions and the BCa method [5, 17]. As the subjective ratings were recorded as Likert-scale ratings for which the meaning of differences is not defined, we did not compute the difference to the baseline, but report the means and bootstrapped confidence intervals. Individual results for the visualization techniques, such as measurements for the baseline, are calculated and reported the same way.

Results

We report the results of our study starting with recorded measurements before continuing with subjective ratings.

Distance to Target

Regarding the distance of the first drawing operation to the target location (Fig. 4), no clear differences of the visualizations compared to the baseline are apparent. That most of the means and the confidence intervals are to the left of 0, especially for *heatmap*, indicates smaller distances to the target compared to the baseline of *minVis* (M: 9.21 mm; CI [3.87, 15.86]). The larger spread of confidence intervals in the *wireframe* condition is linked to a larger spread particularly in the *minVis* condition. The individual distance results have narrower confidence intervals (*depth ray* M: 7.15 mm; CI [3.67, 11.33]; *bubble* M: 7.93 mm; CI [4.08, 12.34]; *shadow* M: 6.6 mm; CI [3.75, 9.9]; *heatmap* M: 4.32 mm; CI [3.44, 5.42]). The style of objects, *solid* or *wireframe*, shows the greatest effect in the *bubble* condition.

Time to Target

Evaluating *timeToTarget* (Fig. 5) shows that using the visualization techniques, except for *heatmap*, seems to increase the time required to reach the intended target compared to *minVis* (M: 7.16 s; CI [6.16, 8.18]). Especially for *shadow* + *solid*, users seem to take more time compared to the baseline. *Heatmap* performs similar to *minVis* and is perhaps marginally faster for *wireframe* objects.

Percentage that the Visualization was Active

Since there was no additional assistance in the *minVis* condition, Fig. 6 shows the individual percentages that each visualization was active while approaching the target. The diagram clearly shows that *bubble* was used considerably less than the other visualizations. All other visualizations were used a comparable percentage of the time. For *depth ray* and *shadow*, it seems as if the visualization is used less for *solid* objects.

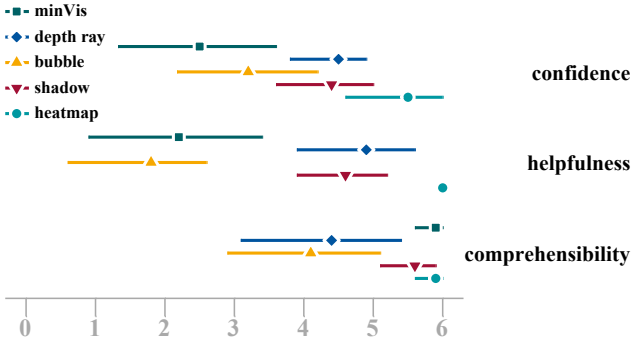

Figure 7. Subjective ratings for confidence, helpfulness, and comprehensibility. *Heatmap* achieved the high ratings overall, followed by *shadow* and *depth ray*. For comprehensibility, *minVis* is also high, with *depth ray* being rated lower. Whiskers denote the 95% CI.

In 133 trials, participants did not activate the visualization technique at all. Most of these trials belong to the *bubble* condition (87 trials) followed by *shadow* (33 trials), *depth ray* (10 trials), and *heatmap* (3 trials).

Confidence, Helpfulness, Comprehensibility

Participants' subjective ratings show that they felt most confident to be able to find the location when using *heatmap* (Fig. 7, top). *Shadow* and *depth ray* seem to be rated lower, with *bubble* and *minVis* even more so. The perceived helpfulness of the visualization follows a similar pattern (Fig. 7, middle). All participants gave *heatmap* the highest mark, while there is no difference in ratings of *depth ray* and *shadow*. *Bubble* and *minVis* clearly scored the lowest. Comprehensibility of *minVis*, *heatmap*, and *shadow* was rated very high (Fig. 7, bottom). *Depth ray* and *bubble* were rated lower.

Most of our participants ranked *heatmap* as their preferred visualization technique (Fig. 8), followed by *shadow* and *depth ray* with similar rankings. *Bubble* and *minVis* are mostly ranked on the last places.

Qualitative Comments

Participants noted that the visualization of the pen tip and its interaction with the environment provided good information during the final approach to the target object, as the tip would

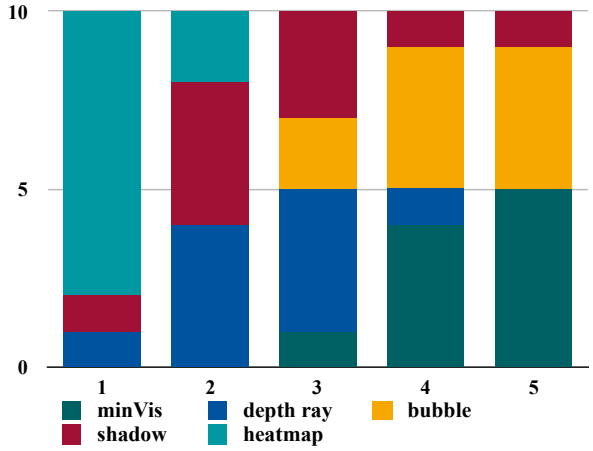

Figure 8. Subjective ranking of the techniques. *Heatmap* is ranked first most frequently, with *depth ray* and *shadow* contending for rank 2 and 3. *Bubble* and *minVis* share the last two places.

disappear once it was inside the object (P3: *"The pen tip entering the cube made it feel like I could precisely estimate the position, but [it] took long"*). Comments regarding *heatmap* were generally positive, and participants noted that they would also get information if the pen tip was inside a *solid* object. For *depth ray*, *wireframe* created problems: Participants mentioned that it was hard to differentiate between lines that were rendered close to each other but varied in distance (P6: *"The full cubes helped a lot more, because you exactly know what surface the measurement is based on"*). Participants also stated that it was difficult to interpret the numerical value, and that therefore it was often used just as a trend of movement. In the *shadow* visualization, participants mentioned that they preferred to use the technique for *wireframe* objects, as it also provided information *inside* the object, whereas shadows of *solid* objects often overlapped each other (P5: *"Filled cubes made finding the shadow of the pen tip more difficult"*). Participants also said that tilting the device downward to see the shadow plane limited their viewing options with the device. *Bubble* received mostly critical comments, stating that the visualization was hard to understand (P5: *"The bubble is not helpful at all, I do not get it"*).

DISCUSSION & DESIGN IMPLICATIONS

The results of our study indicate that *bubble* is not a good visualization technique to show the position of a mid-air pen, as participants rated the helpfulness of the visualization low and mentioned that they were not able to comprehend the depth cue. This also becomes apparent in the high number of trials in which participants did not even activate the visualization.

Heatmap achieved the highest ratings and had good performance results, as it showed the greatest possibility for improving accuracy and movement time. While *depth ray* only showed the distance in one specific direction, and *shadow* required the user to tilt their device to see the visualization, *heatmap* and *bubble* showed the position of the pen tip in all directions. However, *bubble* did this in a non-permanent, time-based way, whereas *heatmap* was visible all the time and on all objects in the scene while active. Coloring the whole scene worked well in our setup. If, however, the texture of an object provides the target that the user wants to move the pen to, then this might require the user to switch off the visualization for the final approach. If such a case occurs regularly, the application designer might want to switch to other visualization techniques such as *shadow* or *depth ray*. We provide additional solutions in the Future Work section.

The additional time required for *shadow* and *depth ray* could be due to the need to interpret the visualization: the numeric value for *depth ray* and the shadow on the plane beneath the smartphone for *shadow*. *Heatmap* does not seem to be affected by this added need for interpretation, or the visualization helps the user enough to outweigh the added task.

Depth ray and *shadow* performed quite similar, and also received comparable ratings. While *depth ray* seems to be slightly faster than *shadow*, *shadow* scored higher in comprehensibility. However, the qualitative comments show that the usefulness of these techniques was seen differently based on the style of objects in the scene. While *shadow* was considered more useful in scenes with *wireframe* objects, because the shadows do not overlap or occlude the shadow of the pen tip, *depth ray* was said to work better in scenes with *solid* objects. A reason for this is that it is more difficult to understand which element in the scene the current distance is referring to in scenes with *wireframe* objects.

The result that no visualization technique clearly produces more accurate results in finding the starting position can be explained by the impact that occlusion of the pen tip already has on precision. All visualizations could use the pen tip to see when they would penetrate the target object's surface.

In other projects investigating depth cues, using shadows achieved the best results for egocentric depth estimation with stereoscopic systems (e.g., [14, 21]). In our study, *heatmap* achieved better results than *shadow* particularly for subjective ratings and rankings of our participants. This is closer to the results by Wither et al. [49], since they state that in their study the color condition was more preferred than their shadow planes but the performance was largely the same. While not significant, participants in their study needed the most time for the shadow condition. This is similar to the trend in our findings that it took participants longer to reach the target when using the *shadow* visualization for *solid* objects.

In conclusion, *heatmap* seems to be the best visualization used in this study, followed by *depth ray* and *shadow*.

SUMMARY & FUTURE WORK

Moving a mid-air pen to a specific location in Handheld Augmented Reality is hard, as perceptual issues make it difficult to estimate the correct distance to a virtual object. However, aligning the pen with a virtual object is an essential mid-air interaction when trying to connect virtual objects to their real environment, such as in mid-air modeling applications. In addition to a baseline condition with minimal visualization, we designed four visualization techniques to show the 3D position of the pen in relation to its virtual environment, and compared their performance. A heatmap visualization, which shades

every object in the scene based on its distance to the pen tip, achieved good results and was most preferred by participants.

For this study, we chose the basic task of aligning the pointing device with a virtual object in the scene to gather findings that are applicable to many other settings. Further studies could look at more specific scenarios to see how more specialized visualizations compare to the performance in the basic setup. For example, while our environment consisted of virtual objects, the inclusion of real world surfaces and objects is a promising direction to look at. Our study with 10 participants provides first insights into the effects of the different visualization techniques but studies with more participants might uncover additional findings. Therefore, we provide our study software and data with this publication[3] to encourage replication and use of our results for further exploration.

The heatmap visualization adjusts the appearance of the whole scene instead of providing more localized feedback such as with the depth ray. Future studies could investigate this relationship closer to see whether different textures of objects and targets affect the targeting action and usage pattern of the visualization technique. Heatmap visualizations could also take textures into account and present the distance using different brightness or color adjustments that preserve texture structure. Our heatmap implementation used a computational linear gradient in RGB space which is not perceived as linear by humans. Future studies could use more perceptually precise gradients (e.g., [32]) to see whether this more accurate encoding of the distance between pen and environment improves the interaction even more. Furthermore, users with color-vision impairments might require specific color gradients to counterbalance their impairment. While the heatmap technique in this work encoded the Euclidean distance of the pen tip to objects regardless of the position of the camera, the color gradient could also be used to only encode the depth information from the current point of view. Since this is the most problematic dimension to perceive, this could focus the visualization on this dimension. However, movement of the camera would then also impact the color of objects since it changes the depth axis, making it potentially more difficult to interpret the information.

Our work presented here looked at position visualizations for a single point in space. It would be interesting to see how rendering the real-time occlusion of virtual objects by the real pen and hand holding it affects the task of moving towards a specific location. While occluding objects behind the whole hand would certainly improve depth estimation overall, it would also hide more information in an already small window into the virtual environment. It seems like a promising research direction to investigate how visualizations from X-ray applications perform if applied to the *near* space and a more dynamic environment: In Handheld AR with a mid-air pen, the state between being occluded and not being occluded switches regularly for objects in the scene.

In our study, the task was to move to a specific location. However, for other interactions it may also be necessary to understand the overall layout of objects in a scene. While local visualizations like the depth ray are probably not the best techniques for such a task, global visualizations like the heatmap or shadows are promising candidates for further research. Unlike in our task, users would then have to comprehend and interpret different visualizations on objects to understand their relative positions to each other.

Finally, scene composition also likely influences performance. In our study, we only showed eight cube objects for all visualizations to compare performance. Future studies could evaluate environments with varying object density, to understand how this affects performance of these visualization techniques.

[3] https://hci.rwth-aachen.de/heatmaps

REFERENCES

[1] Claudia Armbrüster, Marc Wolter, Torsten Kuhlen, Wilhelmus Spijkers, and Bruno Fimm. 2008. Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *CyberPsychology & Behavior* 11, 1 (2008), 9–15. DOI: http://dx.doi.org/10.1089/cpb.2007.9935 PMID: 18275307.

[2] 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 *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 185, 15 pages. DOI: http://dx.doi.org/10.1145/3173574.3173759

[3] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5643–5654. DOI: http://dx.doi.org/10.1145/3025453.3025474

[4] Jeff Butterworth, Andrew Davidson, Stephen Hench, and Marc. T. Olano. 1992. 3DM: A Three Dimensional Modeler Using a Head-mounted Display. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics (I3D '92)*. ACM, New York, NY, USA, 135–138. DOI: http://dx.doi.org/10.1145/147156.147182

[5] James Carpenter and John Bithell. 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 19, 9 (2000), 1141–1164. DOI: http://dx.doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

[6] Zeynep Cipiloglu, Abdullah Bulbul, and Tolga Capin. 2010. A Framework for Enhancing Depth Perception in Computer Graphics. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization (APGV '10)*. ACM, New York, NY, USA, 141–148. DOI: http://dx.doi.org/10.1145/1836248.1836276

[7] Geoff Cumming. 2014. The New Statistics: Why and How. *Psychological Science* 25, 1 (2014), 7–29. DOI: http://dx.doi.org/10.1177/0956797613504966 PMID: 24220629.

[8] James E. Cutting. 2003. Reconceiving Perceptual Space. In *Looking Into Pictures: An interdisciplinary approach to pictorial space*, Heiko Hecht, Robert Schwartz, and Margaret Atherton (Eds.). MIT Press, 215–238.

[9] James E. Cutting and Peter M. Vishton. 1995. Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth. In *Perception of Space and Motion*, William Epstein and Sheena Rogers (Eds.). Academic Press, San Diego, 69–117. DOI: http://dx.doi.org/https://doi.org/10.1016/B978-012240530-3/50005-5

[10] Michael F. Deering. 1995. HoloSketch: A Virtual Reality Sketching/Animation Tool. *ACM Trans. Comput.-Hum. Interact.* 2, 3 (Sept. 1995), 220–238. DOI:http://dx.doi.org/10.1145/210079.210087

[11] Arindam Dey, Mark Billinghurst, Robert W. Lindeman, and J. Edward Swan. 2018. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI* 5 (2018), 37. DOI: http://dx.doi.org/10.3389/frobt.2018.00037

[12] 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. DOI: http://dx.doi.org/10.1109/ISMAR.2012.6402556

[13] Arindam Dey and Christian Sandor. 2014. Lessons Learned: Evaluating Visualizations for Occluded Objects in Handheld Augmented Reality. *Int. J. Hum.-Comput. Stud.* 72, 10-11 (Oct. 2014), 704–716. DOI:http://dx.doi.org/10.1016/j.ijhcs.2014.04.001

[14] Catherine Diaz, Michael Walker, Danielle A. Szafir, and Daniel Szafir. 2017. Designing for Depth Perceptions in Augmented Reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 111–122. DOI: http://dx.doi.org/10.1109/ISMAR.2017.28

[15] Pierre Dragicevic. 2016. *Fair Statistical Communication in HCI*. Springer International Publishing, 291–330. DOI:http://dx.doi.org/10.1007/978-3-319-26633-6_13

[16] David Drascic and Paul Milgram. 1996. Perceptual issues in augmented reality. In *Proc. SPIE Vol. 2653: Stereoscopic Displays and Virtual Reality Systems III*. 123–134. DOI:http://dx.doi.org/10.1117/12.237425

[17] Bradley Efron. 1987. Better Bootstrap Confidence Intervals. *J. Amer. Statist. Assoc.* 82, 397 (1987), 171–185. DOI: http://dx.doi.org/10.1080/01621459.1987.10478410

[18] Stephen R. Ellis and Brian M. Menges. 1998. Localization of Virtual Objects in the Near Visual Field. *Human Factors* 40, 3 (1998), 415–431. DOI: http://dx.doi.org/10.1518/001872098779591278 PMID: 11536894.

[19] Eg Su Goh, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 2019. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access* 7 (2019), 40581–40601. DOI: http://dx.doi.org/10.1109/ACCESS.2019.2906394

[20] 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. DOI:http://dx.doi.org/10.1145/2851581.2892412

[21] Geoffrey S. Hubona, Philip N. Wheeler, Gregory W. Shirah, and Matthew Brandt. 1999. The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3D Depth Visualization. *ACM Trans. Comput.-Hum. Interact.* 6, 3 (Sept. 1999), 214–242. DOI: http://dx.doi.org/10.1145/329693.329695

[22] 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. DOI: http://dx.doi.org/10.1145/3024969.3024995

[23] 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. DOI: http://dx.doi.org/10.1007/s11042-011-0983-y

[24] J. Edward Swan II, Liisa Kuparinen, Scott Rapson, and Christian Sandor. 2017. Visually Perceived Distance Judgments: Tablet-Based Augmented Reality Versus the Real World. *International Journal of Human-Computer Interaction* 33, 7 (2017), 576–591. DOI: http://dx.doi.org/10.1080/10447318.2016.1265783

[25] IKEA. 2017. A better reality. Blog. (2017). Retrieved September 17, 2019 from http://highlights.ikea.com/2017/ikea-place/.

[26] Transparent Statistics in Human-Computer Interaction Working Group. 2019. Transparent Statistics Guidelines. (Jun 2019). DOI: http://dx.doi.org/10.5281/zenodo.1186169 (Available at https://transparentstats.github.io/guidelines).

[27] 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. DOI:http://dx.doi.org/10.1109/tvcg.2016.2518099

[28] J. Adam Jones, J. Edward Swan, II, Gurjot Singh, Eric Kolstad, and Stephen R. Ellis. 2008. The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV '08)*. ACM, New York, NY, USA, 9–14. DOI:http://dx.doi.org/10.1145/1394281.1394283

[29] Volkert Jurgens, Andy Cockburn, and Mark Billinghurst. 2006. Depth Cues for Augmented Reality Stakeout. In *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI (CHINZ '06)*. ACM, New York, NY, USA, 117–124. DOI:http://dx.doi.org/10.1145/1152760.1152775

[30] 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. DOI: http://dx.doi.org/10.1109/ISMAR.2010.5643530

[31] 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. DOI: http://dx.doi.org/10.1145/2659766.2659785

[32] Ingmar Lissner and Philipp Urban. 2012. Toward a Unified Color Space for Perception-Based Image Processing. *IEEE Transactions on Image Processing* 21, 3 (March 2012), 1153–1168. DOI: http://dx.doi.org/10.1109/TIP.2011.2163522

[33] Mark A. Livingston, Zhuming Ai, J. Edward Swan, and Harvey S. Smallman. 2009. Indoor vs. Outdoor Depth Perception for Mobile Augmented Reality. In *2009 IEEE Virtual Reality Conference*. 55–62. DOI: http://dx.doi.org/10.1109/VR.2009.4810999

[34] 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

[35] 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. DOI:http://dx.doi.org/10.1145/2466816.2466829

[36] 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. DOI: http://dx.doi.org/10.1145/3173574.3174153

[37] Nina Rosa, Remco C. Veltkamp, Wolfgang Hürst, Tanja Nijboer, Carolien Gilbers, and Peter Werkhoven. 2019. The Supernumerary Hand Illusion in Augmented Reality. *ACM Trans. Appl. Percept.* 16, 2, Article 12 (Aug. 2019), 20 pages. DOI:http://dx.doi.org/10.1145/3341225

[38] 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. DOI:http://dx.doi.org/10.1109/38.103389

[39] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, New York, NY, USA, 507–515. DOI: http://dx.doi.org/10.1145/3116595.3116596

[40] 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. DOI: http://dx.doi.org/10.1109/ISMAR-Adjunct.2016.0059

[41] J. Edward Swan, Mark A. Livingston, Harvey S. Smallman, Dennis Brown, Yohan Baillot, Joseph L. Gabbard, and Deborah Hix. 2006. A Perceptual Matching Technique for Depth Judgments in Optical, See-Through Augmented Reality. In *IEEE Virtual Reality Conference (VR 2006)*. 19–26. DOI: http://dx.doi.org/10.1109/VR.2006.13

[42] Jeff K. T. Tang, Tin-Young A. Duong, Yui-Wang Ng, and Hoi-Kit Luk. 2015. Learning to create 3D models via an augmented reality smartphone interface. In *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. 236–241. DOI: http://dx.doi.org/10.1109/TALE.2015.7386050

[43] Takahiro Tsuda, Haruyoshi Yamamoto, Yoshinari Kameda, and Yuichi Ohta. 2005. Visualization Methods for Outdoor See-through Vision. In *Proceedings of the 2005 International Conference on Augmented Tele-existence (ICAT '05)*. ACM, New York, NY, USA, 62–69. DOI:http://dx.doi.org/10.1145/1152399.1152412

[44] 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. DOI: http://dx.doi.org/10.1145/2522848.2522885

[45] Philipp Wacker, Oliver Nowak, Simon Voelker, and Jan Borchers. 2019. ARPen: Mid-Air Object Manipulation Techniques for a Bimanual AR System with Pen & Smartphone. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 619, 12 pages. DOI:http://dx.doi.org/10.1145/3290605.3300849

[46] 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. DOI: http://dx.doi.org/10.1145/3267782.3267788

[47] Leonard R. Wanger, James A. Ferwerda, and Donald P. Greenberg. 1992. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications* 12, 3 (May 1992), 44–58. DOI: http://dx.doi.org/10.1109/38.135913

[48] John P. Wann, Simon Rushton, and Mark Mon-Williams. 1995. Natural problems for stereoscopic depth perception in virtual environments. *Vision Research* 35, 19 (1995), 2731–2736. DOI:http://dx.doi.org/https://doi.org/10.1016/0042-6989(95)00018-U

[49] Jason Wither and Tobias Hollerer. 2005. Pictorial depth cues for outdoor augmented reality. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*. 92–99. DOI: http://dx.doi.org/10.1109/ISWC.2005.41

[50] 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. DOI: http://dx.doi.org/10.1145/3126594.3126664

[51] 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. DOI:http://dx.doi.org/10.1145/1450579.1450627