Investigating Smartphone-based Pan and Zoom in 3D Data Spaces in Augmented Reality

Wolfgang Büschel
bueschel@acm.org
Interactive Media Lab, Technische Universität Dresden
Dresden, Germany

Annett Mitschick
annett.mitschick@tu-dresden.de
Interactive Media Lab, Technische Universität Dresden
Dresden, Germany

Thomas Meyer
thomas.meyer@mailbox.tu-dresden.de
Interactive Media Lab, Technische Universität Dresden
Dresden, Germany

Raimund Dachselt
dachselt@acm.org
Interactive Media Lab, Technische Universität Dresden
Dresden, Germany

Figure 1: Left: 3D data spaces can be explored by 3D panning (a) and zooming (b) relative to their fixed presentation space. Right: A user wearing a HoloLens explores such a 3D data space with smartphone-based interaction techniques using our prototype. The devices are equipped with tracking markers as used in our study for improved tracking precision.

ABSTRACT

In this paper, we investigate mobile devices as interactive controllers to support the exploration of 3D data spaces in head-mounted Augmented Reality (AR). In future mobile contexts, applications such as immersive analysis or ubiquitous information retrieval will involve large 3D data sets, which must be visualized in limited physical space. This necessitates efficient interaction techniques for 3D panning and zooming. Smartphones as additional input devices are promising because they are familiar and widely available in mobile usage contexts. They also allow more casual and discreet interaction compared to free-hand gestures or voice input. We introduce smartphone-based pan & zoom techniques for 3D data spaces and present a user study comparing five techniques. Our results show that spatial device gestures can outperform both touch-based techniques and hand gestures in terms of task completion times and user preference. We discuss our findings in detail and suggest suitable techniques for specific AR navigation tasks.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Interaction techniques**.

KEYWORDS

Augmented Reality; 3D Navigation; 3D Data Exploration; Pan & Zoom; Interaction Techniques; Immersive Visualization

ACM Reference Format:
Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. Investigating Smartphone-based Pan and Zoom in 3D Data Spaces in Augmented Reality. In *21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19), October 1–4, 2019, Taipei, Taiwan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3338286.3340113

MobileHCI '19, October 1–4, 2019, Taipei, Taiwan

ACM ISBN 978-1-4503-6825-4/19/10…$15.00
https://doi.org/10.1145/3338286.3340113

1 INTRODUCTION

Head-mounted Augmented Reality (AR) is currently one of the most promising directions for future human-computer interfaces. Due to available hardware (e. g., the Microsoft HoloLens or Magic Leap), the application of head-mounted AR in real-world use cases becomes feasible and attractive for both business and research. People will increasingly use three-dimensional data visualizations right in their real everyday contexts for their professional and private data analysis needs, which has been described as Immersive or Situated Analytics [13, 20]. While early days AR solutions were mostly concerned with correctly registered rendering of 3D objects within their real context, AR in our days also encompasses augmenting reality with other media or even complex 2D or 3D information spaces. Reality-based Information Retrieval [10] is just one novel research direction exploiting the full potential of mixed reality visualizations beyond 3D models.

A recent study [40] showed that AR consistently outperformed Virtual Reality (VR) regarding task completion time for 3D object selection and transformation. This and the assumption that AR glasses might become the successor of today's digital workplaces yields the research challenge of how to effectively navigate these augmented virtual information spaces. Take for example the visualization of a 3D model of planetary systems, star clusters, and galaxies, presented in room-sized scale in an educational context. The presentable volume shows a selected part of the model and users can adjust the selection to look into details through directed panning and zooming, similar to a classic 2D zoomable information space. Another exemplary use case is the virtual presentation of a larger number of 3D objects or photos arranged in non-instrumented, physical space as a result of query or search activities triggered by real objects [10]. Here again, the augmented information space needs to be moved (translation) and zoomed (scaling) for further inspection and refinement (Figure 1, left). In general, such virtual information spaces might be larger than the available physical space (e. g., in indoor scenarios) or need to be scaled down to be manageable. We therefore believe that navigating 3D information spaces through 3D panning and zooming will be an essential task in future AR user interfaces and necessitates efficient interaction strategies. We chose to use the terms of pan and zoom throughout this paper—instead of the more technical terms of translation and scaling—to emphasize the relationship with classic 2D data space navigation rather than free 6+ DoF object manipulation.

Input modalities for head-mounted AR range from traditional devices, such as keyboard, mouse, and touch screens to advanced and natural 3D interaction techniques like free-hand gestures, voice input, body motion, or gaze [8]. Purpose-built wearables for interaction in AR have also been proposed [19, 21, 37, 38, 52]. Datcu et al. [18] showed that tangible interaction allows users to perform better when navigating and manipulating a menu in AR. Regarding availability, familiarity, and social acceptance, smartphones have a great potential to serve as default control devices in head-mounted-AR settings. We believe that smartphones, even if their form factor changes, will be a staple of personal, mobile devices for the foreseeable future and thus lend themselves to be used in combination with AR glasses.

In this work, we therefore contribute an exploration of smartphone-based interaction techniques for 3D pan & zoom interaction as essential navigational tasks in head-mounted AR. Based on a comprehensive set of design goals, we carefully designed four unimanual, eyes-free interaction techniques for panning and zooming augmented information spaces. They differ with regard to the mapping of smartphone-based *spatial input* or *touch input* to the degrees of freedom required. We specifically focus on 3D pan & zoom and do not consider rotation of the data space. For our use case and in contrast to general object manipulation, rotation is often not necessary (the user can physically move around the data) and could even be problematic (if the data is intentionally aligned to the physical world). In order to investigate the suitability of our four phone-based pan & zoom techniques for head-mounted AR, we conducted a user study with 25 participants comparing them to a free-hand interaction technique based on the default Microsoft HoloLens air-tap gesture. We measured task completion time and efficiency of several navigation tasks to evaluate the techniques' performance and also investigated how users assess their usability. Both the quantitative results and the participants' ratings strongly indicate the superiority of those smartphone-based techniques employing spatial input and clear, intuitive mappings. From that we can conclude that spatial-aware smartphones are a viable and promising solution for serving as controllers for head-mounted AR navigation and interaction.

In the remainder of this paper we first review related work regarding distant control of 3D manipulation tasks as well as panning and zooming of large information spaces. We then discuss the design decisions for the development of our smartphone-based 3D pan & zoom techniques. Subsequently, our user study is presented in detail with all results. Finally, we discuss our findings and present recommendations for using smartphones in head-mounted AR as controllers for 3D navigation tasks and beyond.

2 RELATED WORK

To our knowledge, there are very few previous publications with close resemblance to our work, i. e., quantitative user studies on smartphone-assisted pan & zoom in 3D data sets, in particular in head-mounted AR. A vision of a collaborative 3D data analysis tool using tablets and head-mounted AR was recently proposed in [57] but has not been fully implemented yet. Recently Lee & Chu [42] combined a HoloLens with a smartphone to provide touch and spatial input. A preliminary study showed a clear advantage of their technique compared to the gestures provided by the HoloLens. Other work on head-mounted AR was focused on selection techniques using head- and eye-direction [41] and multitouch input for direct manipulation of virtual interfaces [68].

A lot of work has been done regarding interaction techniques for pan & zoom in 2D information spaces on distant displays [16, 51, 53, 56], hand-held devices [62], and in VR [14]. Nevertheless, 2D and 3D interaction are quite different in terms of quantifying pan & zoom coordination. According to [9], zooming in 2D changes both target width and target distance proportionally and thus does not interfere with panning regarding the index of difficulty of Fitts' Law. In contrast, 3D interaction is far more complex, as can be argued from 3D pointing studies which observed that movement time was significantly affected by the direction due to motor skills [50], and errors along the visual axis are dominant due to visual perception [45]. Occlusion also affects user performance [54]. Thus, additional studies specifically for 3D pan & zoom, as presented in this paper, are necessary.

In this section we concentrate on strongly related work in the field of 3D interaction which involves *distant 3D object manipulation techniques* on the one hand and strategies for *distant pan & zoom in large data spaces*, both in 2D (e. g., on large displays) and 3D environments (stereoscopic displays, AR and VR), on the other hand. Not in the direct focus of our investigation are the broad field of mobile, hand-held AR and existing techniques to interact with such applications like [2, 11, 27, 30, 31, 46, 66], as well as novel and innovative input modalities for navigation in information spaces, such as gaze-supported interaction [29, 63], around-the-device interaction [4, 36, 39, 58], directional walking [69], or specialized controllers and wearables such as [21, 32, 37].

Distant 3D manipulation techniques

3D interaction and 3D object manipulation have been the focus of much research in the last decades [26, 28, 35], with increasing importance through the emergence of affordable, feasible hardware for head-mounted AR and VR. Lots of techniques for direct and widget-based manipulation of 3D objects have been proposed, e. g., [1, 64]. However, these techniques are typically not suitable for the use with immersive AR headsets or to navigate large, multi-scale datasets. Also, head-mounted AR has specific challenges like cluttered real-world- backgrounds, limited FOV, and own body visibility that may influence user performance and distinguish it from interaction with VR systems and distant displays.

Mendes et al. wrote a survey on 3D object manipulation [47], including sections on touch and mid-air interaction. A study of distant 3D manipulation techniques was presented by Daiber et al. [17]. They compared mid-air hand gestures, touch on a mobile device and a combination of touch and device tilt. Bimanual hand gestures outperformed the other techniques. In contrast to our work, they did not test scaling (zooming) and used a stereoscopic display instead of immersive AR headsets. They also did not consider spatially-aware mobile devices, which have now become available with the introduction of sophisticated AR frameworks by Apple and Google. Bimanual input was also investigated by Feng et al. [22], presenting a comparison of three techniques for 7-DoF manipulation (with and without scaling) in a FishTank VR system with a pair of special input devices.

Vuibert et al. [65] compared 6-DoF docking task performance of different mid-air interaction techniques, including a physical prop and the user's hand. In contrast to our work, they did not investigate scaling (zooming) and used desktop VR and 3D shutter glasses for stereoscopic view. Also Speicher et al. [60] presented a 6-DoF 3D manipulation technique for stereoscopic displays, but using a mobile device. The authors studied this technique in a monoscopic vs. stereoscopic view but did not compare it to other interaction techniques. Similarly, Millete & McGuffin [48] proposed a set of 3D interaction techniques using a tracked smartphone and touch input. However, in their paper, they only presented informal user feedback. Besançon et al. [7] compared mouse, touch, and spatial input for a 6-DoF 3D docking task, but also without scaling and only with a regular non-stereoscopic display. They found that the three input modalities are similar regarding precision but different in completion times: spatial interaction through the tangible was faster than touch, and touch was faster than mouse interaction. Spatial input for the manipulation of 3D objects with tracked displays above a tabletop has been examined by Spindler et al. [61], however they only collected initial user feedback. Rodríguez and León [55] used smartphones as remote 3D interaction devices for basic 3D manipulation tasks in a digital heritage application but did not formally study the techniques. Finally, Grandi et al. [24] presented a collaborative user interface making use of smartphones for the manipulation of 3D objects. Their design combined touch and accelerometer based input, but only used touch input for translation and scaling, reserving device orientation for rotation and camera orientation.

Distant Pan & Zoom Navigation in Large Data Spaces

Bergé et al. [5] examined smartphone interaction for large 3D displays (e. g., public displays), comparing touch, mid-air movement, and hand movement around device. Similarly, Song et al. [59] use touch and device gestures with a mobile device to move and rotate a 3D slicing plane on a display wall. Neither paper considered scaling/zooming with the mobile device and, in contrast to us, they did not employ immersive AR headsets but additional, large displays.

In contrast to these papers and most 3D object manipulation techniques described above, pan & zoom navigation involves the predefinition (e. g., the center of the visualization) or selection of a zoom focus point (focus of expansion) in the scene. Olwal and Feiner [53], for instance, used spatially-aware handhelds to determine the zoom focus on large displays. Nancel et al. [51] studied unimanual and bimanual mid-air techniques for pan & zoom on large distant displays, using the dominant hand for pointing to the focus of expansion. Scrolling and zooming in a World in Miniature (WIM) view has been examined by Wingrave et al. [67].

Besançon et al. [6] combined touch and spatial input for the exploration of 3D visualizations with a mobile device, therefore used a conventional display and the phone display, but did not provide a comparison study. López et al. [44] studied interaction for 3D data exploration combining a stereoscopic display and a monoscopic mobile device. Two existing touch interfaces [15, 70] were used to manipulate the data space and tested in a study.

Our use case differs from these distant display interactions in that we can assume neither a purely virtual scene nor a mostly stationary user in front of a planar frame of reference.

3 DESIGNING SMARTPHONE-BASED PAN & ZOOM FOR HEAD-MOUNTED AR

3D pan & zoom is a task requiring four degrees of freedom (DoF): a 3-DoF translation of the view position (x, y, z) and 1-DoF uniform scaling, presuming that the zoom focal point is preselected, e. g., in the center. The design space of possible solutions for mapping pan and zoom to available input channels would be endless without rigorous, but well-considered restrictions. In the following we discuss our design choices as a basis for the selection of candidate techniques.

Design Goals

Our goal is to study relevant input techniques for panning and zooming within a virtual 3D data space in head-mounted AR using a smartphone as exclusive input device. For this purpose we defined general requirements that the selected and composed pan & zoom techniques should all address:

Unimanual Input: Although the use of bimanual input techniques is generally recommended in literature to achieve high levels of parallelism [12, 43], we decided to investigate one-handed interaction to allow for a more casual and flexible interaction, leaving room for the use of additional tools, e. g., for selection or inspection of items.

Eyes-free Interaction: We seek to design for eyes-free interaction in order to keep the users' visual focus to the AR visualization and create minimum distraction. Thus, the touch display of the smartphone is not used for explicit visual feedback, but the device offers common haptic feedback like vibrations to confirm successful actions or mode changes.

Smartphone-only Implementation: We decided to omit additional input through specialized controllers or head- or gaze-tracking techniques as, e. g., in [41]. This decision is motivated by the same reasons as for unimanual and eyes-free interaction, but also by the idea of selecting and composing smartphone-based techniques that can be realized with current, out of the box mobile devices using inertial sensors or the device camera to provide spatial awareness.

High Degree of Compatibility: The *degree of compatibility* "measures the similarity between the physical actions of the users on the instrument and the response of the object" [3]. The physical action should resemble the response as much as possible. This implicates a desirable "symmetry of actions" in all needed DoF (back-forth, left-right, up-down, in-out) with no evident bias to any direction.

Robustness and Conciseness: It should be avoided that actions (accidentally or by their nature) interfere with each other or are misapplied, e. g., because of fatigue. This is typically a matter of human fine motor skills, how close gestures are designed to be potentially misinterpreted, and whether mode switches are inserted to clearly separate mappings.

Design Dimensions

While these design goals guided our iterative design of navigation techniques, the following generic properties or design dimensions (inspired by [3]) helped us balance the choice of techniques to cover a range as wide as possible. They also facilitate comparing efficiency and suitability of the techniques for specific tasks.

(D1) Degree of spatiality: *How many DoF are controlled through spatial input?* Previous research [11, 62] has shown that the movement of a tangible, spatially-aware input device can outperform touch gestures. On the other hand, input in free space can easily fatigue users [28]. Hancock et al. [25] found out that touch can outperform tangible interaction for moving and rotating objects in 2D while tangibles were more effective for navigating a visualization.

(D2) Degree of simultaneity: *How many DoF can be controlled in parallel/simultaneously?* According to Jacob and Sibert [34], panning and zooming are strongly related, forming an integral task which is better performed with an integral device providing a high number of DoF to execute required actions in parallel. However, they only investigated 2D panning and zooming and did not consider aspects of simultaneity for 3D pan & zoom, involving an additional dimension and a higher number of DoF.

(D3) Degree of guidance: *How many DoF are controlled through gestures with some sort of alignment to give guidance to the user?* Similar to [51] one can differentiate between 1D (path in space), 2D (surface), and 3D (totally free) guidance, accordingly corresponding to decreasing degree of guidance. Alignment or guidance in one or two dimensions is especially important for specific 3D visualizations in AR which are coupled to physical surfaces (e. g., placed on a table) and are used to examine the characteristics of a dataset along selected axes or cut surfaces. Of course, if guidance means to deactivate unneeded DoF, there is a correlation between D2 and D3 regarding separability. However, a high D2 does not necessarily imply a low D3 and vice versa.

Input Modalities & Mappings

The decision to study phone-based, unimanual interaction and the requirements described above limit the design space of possible solutions for mapping pan and zoom to two basic channels, spatial interaction and touch input.

Spatial Interaction. Mid-air interaction with a mobile device provides six DoF: up-down, left-right, back-forth, yaw, pitch, and roll. It is very evident to apply the first three of them to 3D panning because of the high *degree of compatibility* [3]. Furthermore, all of them can be used for zooming. To allow for clutching and repositioning, the prevalent strategy is a hold gesture with the thumb on the phone's display. Of course, other techniques are imaginable but not necessarily more efficient.

Touch Input. Single-touch input on a 2D surface – which is effectively the case for unimanual operation – together with a minimum temporal offset provides two DoF: up-down, left-right (*Drag*). Circular, clutch-free touch gestures have been proposed in literature [33] as an alternative to linear swipe gestures, which can be seen as an additional DoF. However, Nancel et al. [51] assessed a higher efficiency of linear gestures compared to circular gestures, and a combination of both may introduce interferences and misinterpretations. As a consequence, touch input on a mobile device can be easily mapped to zooming (1D) and planar panning (2D), but not applied for full 3D panning without additional input or specific gesture sets.

As single-touch input is applicable for planar panning (XY, YZ, XZ, or any cutting plane), another input modality is needed to pan along the orthogonal axis or to specify or change the plane. This could be either a "clicking through options" with a tap or double-tap (touch) or discrete spatial actions like pointing the smartphone towards a plane or rotating the smartphone so that its orientation is aligned with the plane to be selected.

Proposed Techniques

In an iterative process, we identified four applicable phone-based solutions for panning (Move, Rotate, 2D-Drag with plane selection, 2D-Drag with orthogonal translation) and three for zooming (Move, Rotate, Drag) which can be combined to 12 compound techniques. We selected a subset of four techniques based on our above-described requirements and aiming at a wide coverage of D1, D2 and D3 (Figure 2).

Move+Drag *(M+D)*: This technique is a combination of moving the device in free-space as direct input for 3D translation of the data space and a touch-based indirect input for zooming through an up/down-drag gesture (zoom in/out) on the phone's touch screen. This combination is a mix of spatial and on-device interaction (moderate degree of spatiality, D1) with a high degree of simultaneity (D2), but low degree of guidance (D3) because of the free-space input.

Move+Rotate *(M+R)*: Quite similar to M+D, this technique uses free-space input by moving the device for 3D translation of the data space. Zooming is realized by rotating the smartphone to the left (zoom out) and right (zoom in) along the forward facing axis. This combination has a high D1 as all DoF are controlled through spatial input and a high D2. Similar to M+D it has a low D3.

Drag+Drag *(D+D)*: Here we combine 2D-Drag on either the XZ-, YZ- or XY-plane with tracking the discrete device orientation for plane selection. A double tap is used as mode switch to enable zooming through an up/down-drag gesture (zoom in/out). This touch-oriented technique has a low D1, and through mode switch and discrete plane selection a very low D2. In contrast, through the alignment of the touch surface to the virtual planes it has a very high D3.

DragRotate+Drag *(DR+D)*: Similar to D+D, this technique provides 2D-drag but just for translation in the XZ-plane (horizontal), while rotating the phone along the forward facing axis simultaneously translates the data space up and down on the Y-axis. As in D+D, a double tap is used to enable zooming through an up/down-drag gesture (zoom in/out), resulting in a moderate D2. This technique is similarly spatially-oriented as D+D, and also has a high D3.

As a baseline, we included **AirTap *(AT)***, a free-hand technique based on the air-tap gesture provided by the HoloLens (Figure 2, left) using the internal camera for gesture recognition. Activated by a single air tap, 3D translation of the data space follows the free-space motion of the hand. A double air tap activates zooming, moving the hand back and forth results in zooming in and out. This pure free-hand gesture has a high D1, but a moderate D2 and low D3.

All techniques were initially implemented using only the internal devices' sensors. However, for more reliable and precise logging, we decided to use an external IR tracking system for the smartphone-based techniques during the study.

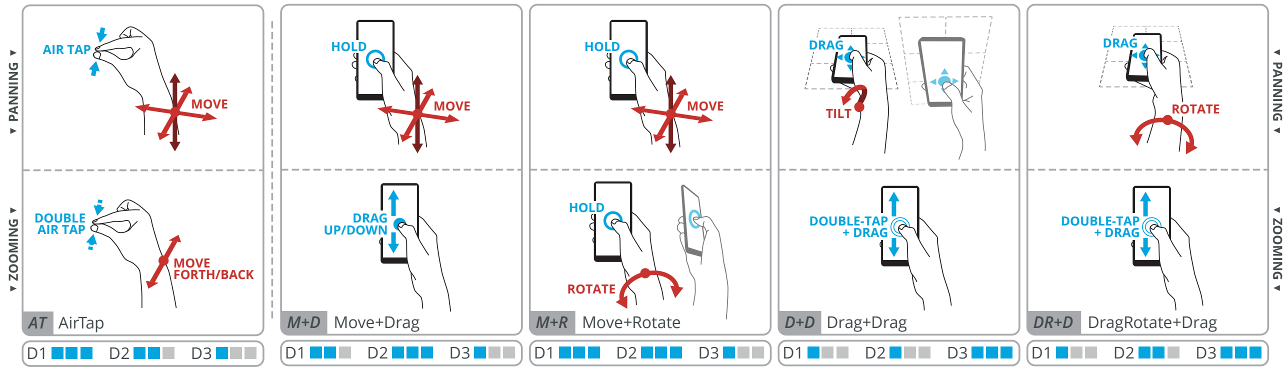

Figure 2: The five composed techniques for pan & zoom selected for evaluation. For each technique, the degrees of spatiality (D1), simultaneity (D2), and guidance (D3) are shown.

4 STUDY

Our goal was to gain insights into different 3D pan & zoom techniques and their combination for the exploration of a 3D data space. Particularly, we wanted to assess our five selected techniques regarding their performance in speed, efficiency, learning effects, and fatigue.[1]

Hypotheses

In accordance with the literature [62], and in consideration of their high degree of compatibility due to the extensive use of spatial input (D1) for panning, we assume interaction techniques based on spatial input for panning to be generally faster than touch techniques. Furthermore, as the potential of techniques to simultaneously perform actions to control as much DOF as possible influences their overall efficiency [34], techniques with a high D2 should perform significantly faster than those with lower D2. Thus, we expected that techniques M+D and M+R outperform techniques D+D and DR+D regarding task completion time (***H1***).

Free-space gestures with no alignment or guidance for specific paths or surfaces heavily depend on the user's body-awareness and proprioception [49] to efficiently perform the optimal relative motion. This means that for tasks that only involve planar panning those techniques with a high D3 should be superior. We therefore assumed that for planar panning tasks techniques D+D and DR+D outperform AT, M+D and M+R (***H2***). Furthermore, free-space gestures are known to be more physically demanding. Thus, we expected that users perceive techniques AT, M+D and M+R to be more tiring and physically demanding than D+D and DR+D (***H3***). In contrast, we predicted that D+D and DR+D will be perceived to be mentally more demanding (***H4***).

Another interesting question is: How fast are users able to learn the different techniques? While the concept of spatial interaction is understood to be quite intuitive [28], our experience shows that the air tap gesture is initially hard to perform for many users and needs considerable training. Thus, we expected a particular strong learning effect for AT in comparison to the other techniques (***H5***).

Study Design

We designed a user study as a controlled lab experiment. The independent variables were *interaction technique* (the five techniques of AirTap (AT), Move+Drag (M+D), Move+Rotate (M+R), Drag+Drag (D+D), and DragRotate+Drag (DR+D)), *task type* (two levels, 2D and 3D), and *target zoom level* (three levels, 2.5, 1.0, 0.75). The dependent variables were *task completion time* and *efficiency* (the ratio of the shortest translation path and the actual path). We chose a within-subject design where each participant completed all tasks with all techniques. The order of the techniques was counterbalanced, the order of the tasks for each technique was randomized.

Participants

We recruited 28 participants from an entry level HCI course at the local university, 25 of which completed the study successfully. The results of three participants had to be excluded because of technical and personal issues. 15 participants were male, 10 were female. Their mean age was 21.8 (SD = 3.5). All had normal or corrected-to-normal vision.

Apparatus and Tasks

Our experimental setup included a Microsoft HoloLens and a smartphone. In order to bring both devices into a common coordinate system and allow for precise and solid data logging and analysis, we attached IR markers to them and tracked both devices with a 12 camera IR tracking system (tracking volume: $4.0m \times 3.2m \times 1.7m$).

[1]Further details and study data available at https://imld.de/ar-pan-zoom

Our software architecture consisted of individual clients on smartphone and HoloLens and a server application on a PC. This server streamed tracking data to the HoloLens devices. The smartphone streamed interaction data (touch and accelerometer) to the HoloLens. All software was written in C# using the Unity 3D engine.

The scene presented to the user consisted of a cube with a size of $0.8m \times 0.8m \times 0.8m$. This cube served as a clipping volume for a larger scene, effectively serving as a volumetric window into this scene. The scene represented a zoomable information space. It was filled with randomly positioned clusters of colored spheres that served as both a depth cue and as mock "data items". In the center of the cube, a target volume was highlighted. For each trial, a small red cube was placed at predefined locations in the scene (Figure 1, right).

The goal of each task was to move this target object into the target volume in the center of the presentation space and adjust its size to the target volume. To this end, participants had to pan & zoom the scene. When the target was placed and scaled correctly, the next trial started. For 12 trials per task, the target was placed on one of the coordinate planes (2D tasks); for the remaining 24 trials, the targets were equally distributed between the scene octants (3D tasks). The target zoom level relative to the initial zoom was set to 2.5 (zoom in), 1.0 (no zoom), and 0.75 (zoom out) for 12 trials each.

Procedure and Data Collection

After an entry questionnaire (including prior AR experiences and physical/mental state) and signing informed consent, we calibrated the interpupillary distance of each participant using the HoloLens' calibration app. Afterwards, each participant completed five blocks of study trials, one block per interaction technique. First, each technique was explained in detail. Afterwards, the participants had time to train the technique until they were satisfied. This training period typically took around 5 minutes. The participants then completed 36 trials of the task described above in randomized order. They were instructed to solve the tasks one-handed and as quickly as possible but in a comfortable way. After each block, a questionnaire on task load and perceived performance was filled out by the participants. After completion of all five blocks, participants filled out an exit questionnaire. In total, participants needed 60 to 90 minutes to complete the study.

We logged task completion times as well as the target position and scale over time. We also logged the position and orientation of the HoloLens and, if applicable, the phone. All questionnaires were completed on paper and later digitized. The experiment was monitored by at least two researchers/assistants at any time, who also took notes on their observations. We opted against a screen capture of the HoloLens, as it would have been detrimental to the framerate.

Results and Analysis

We analyzed both task completion times and efficiency. However, for efficiency, we found results similar to those for the task completion times. We therefore do not report the results here and focus on the times instead. We also looked into the questionnaire data and examined learning effects.

Task Completion Times. We analyzed the task completion times using a three-way repeated measures ANOVAs. The three factors were the interaction technique, the task type, and the zoom level. We applied Greenhouse-Geisser corrections whenever the assumption of sphericity was violated. Post hoc tests were conducted using pairwise t-tests with Bonferroni corrections and $\alpha = .95$. Shapiro-Wilk tests showed several violations of the normality assumption for the residuals but we decided against changing our analysis approach based on evidence for the robustness of the ANOVA (e. g., [23]) and because ANOVAs on log-transformed data showed the same general trends.

We found significant main effects of the interaction technique [$F(4, 96) = 27.79, p < .001, \eta_p^2 = .54$], task type [$F(1, 24) = 16.55, p < .001, \eta_p^2 = .41$], and zoom level [$F(1.43, 34.34) = 167.07, p < .001, \eta_p^2 = .87$]. We also found interactions between technique and task type [$F(2.18, 52.38) = 6.42, p = .003, \eta_p^2 = .21$], technique and zoom level [$F(5.59, 134.05) = 7.08, p < .001, \eta_p^2 = .23$], and task type and zoom level [$F(2, 48) = 3.31, p = .045, \eta_p^2 = .12$].

Post hoc tests showed several significant differences between the techniques, with M+D [$M = 7.97, SE = .41$] and M+R [$M = 9.96, SE = .72$] being faster than AT [$M = 14.93, SE = 1.04$], D+D [$M = 12.88, SE = .84$], and DR+D [$M = 15.35, SE = .99$] (Figure 3a). In general, the 2D tasks [$M = 11.73, SE = .62$] were completed faster than the 3D tasks [$M = 12.71, SE = .67$], however this effect is not as pronounced as we expected (Figure 3b).

Looking into the interaction of technique and task type in more detail, we found that the 2D tasks were significantly faster for M+D, D+D, and DR+D but no significant difference showed for AT and M+R (Figure 3d). Especially for D+D and DR+D, the difference between 2D tasks and 3D tasks was big, indicating that both the added step of choosing a plane of reference in D+D and the separated input dimensions of DR+D negatively effect 3D panning performance.

For the zoom levels, we found that the tasks requiring zooming in were completed significantly slower [$M = 16.37, SE = .84$] than the tasks requiring zooming out [$M = 10.21, SE = .50$] or not zooming at all [$M = 10.07, SE = .62$] (Figure 3c). Based on the data we believe that this is due to participants zooming out (to get an overview or simplify panning) even when not strictly necessary. Also, the required zoom change was bigger for zooming in, making these tasks

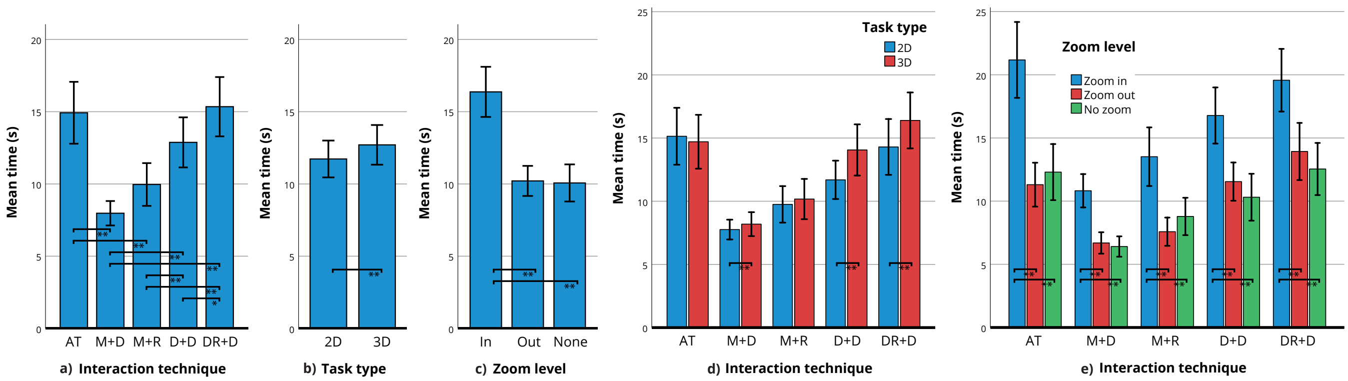

Figure 3: Mean task completion times for the five interaction techniques (a), the two task types (b) and the three zoom levels (c). Also, mean task completion times for technique × task type (d) and technique × zoom (e). Brackets mark statistical significance (: p<0.01, *: p<0.05), error bars show 95% confidence intervals.**

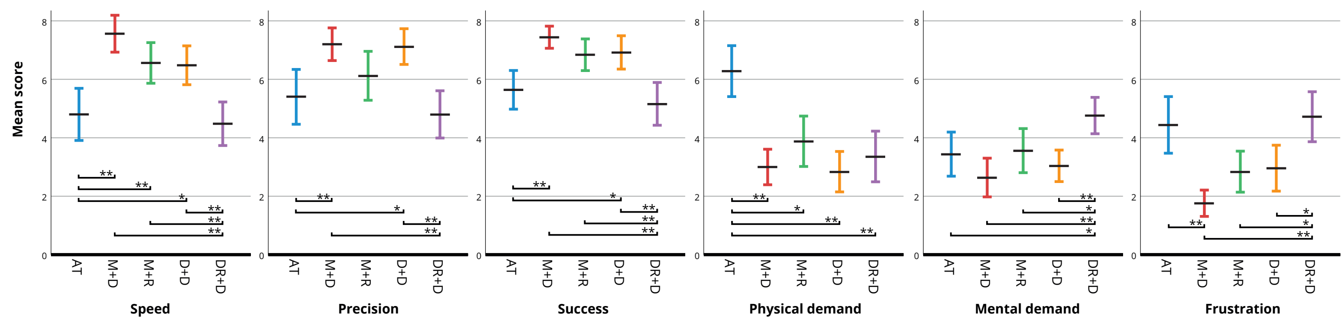

Figure 4: Mean scores for the task load questionnaire. All scores in range [1,10], higher scores are better for speed, precision, and success, lower scores are better for physical and mental demand and frustration. Brackets mark statistical significance (: p<0.01, *: p<0.05), error bars show 95% confidence intervals.**

potentially harder to complete. Post hoc analysis of the interaction of technique and zoom level confirm these findings: For all techniques, zooming in was significantly slower but there is no technique for which there is a significant difference between zooming out and not zooming (Figure 3e).

For the interaction of task type and zoom level, we found significant differences between the task types for *zoom in* and *no zoom* but not for *zoom out*. A possible explanation is that zooming out compresses the distances to the coordinate planes, effectively making 2D and 3D tasks very similar.

Questionnaires. Using Friedman's ANOVA to analyze the questionnaire data (task load and user preferences), we found significant differences between the techniques for perceived speed $\chi^2(4) = 50.33, p < .001$, precision $\chi^2(4) = 32.67, p < .001$, frustration $\chi^2(4) = 37.81, p < .001$, mental demand $\chi^2(4) = 29.25, p < .001$, physical demand $\chi^2(4) = 40.00$, $p < .001$, and task performance $\chi^2(4) = 38.62, p < .001$. Post hoc tests showed that AT and DR+D were rated significantly slower than M+D, M+R, and D+D. They were also rated as less precise than M+D and D+D. Physical demand was rated significantly higher for AT $[M = 6.28, SD = 2.11]$ than the other techniques. Mental demand, on the other hand, was highest for DR+D $[M = 4.76, SD = 1.51]$, all other techniques were rated as significantly less demanding. Consequently, DR+D was also rated significantly more frustrating than M+D, M+R, and D+D. In terms of success/task performance, M+D, M+R, and D+D were rated higher than DR+D. M+D and M+R were also rated higher than AT (Figure 4).

Analysis of the reported suitability (Figure 5) of the techniques confirmed these results, showing a significant difference between suitability for the task in general $\chi^2(4) = 39.688, p < .001$, for panning $\chi^2(4) = 19.42, p = .001$, and for zooming $\chi^2(4) = 34.97, p < .001$. Post hoc tests show that DR+D was rated significantly worse than M+D, M+R, and D+D for general task suitability. AT was rated significantly higher than DR+D for panning. For zooming, AT was rated significantly worse than M+D, M+R, and D+D. Also, M+D was rated better than DR+D. Asked for their overall preference, most participants (14 of 25) voted for M+D as their most favorite technique.

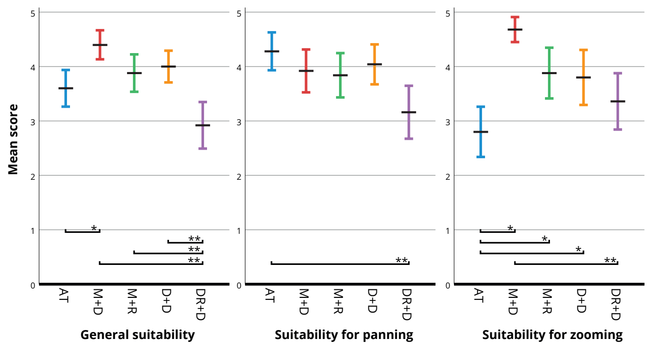

Figure 5: Mean scores for suitability. Scores in range [1,5], higher scores are better. Brackets mark significance (: p<0.01, *: p<0.05), error bars show 95% confidence intervals.**

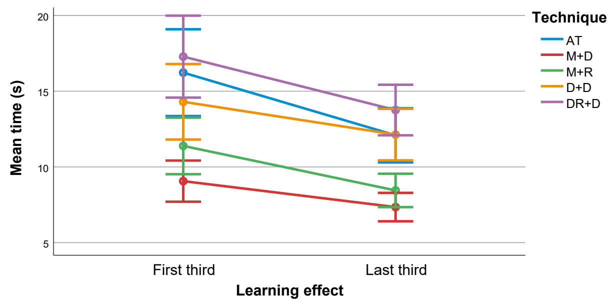

Figure 6: Mean task completion times for the first (left) and last (right) batch of twelve trials. Note the learning effect for all techniques. Error bars show 95% confidence intervals.

Learning Effects. To detect potential learning effects for the techniques, we compared the median task completion time of the first twelve and the last twelve trials for each task (Figure 6). We analyzed this data with a two-way ANOVA of the independent variables interaction technique and trials batch (early, late). Again, we used Greenhouse-Geisser corrections when sphericity was violated. In addition to the expected main effect of interaction technique, we also found a main effect of trials batch [$F(1,24) = 51.59, p < .001, \eta_p^2 = .68$], indicating a significant learning effect. However, the interaction of technique and trials batch was not significant, which means we cannot confirm any difference between the learning effects for the individual techniques.

Additional Findings from Observations. We observed that almost none of the participants moved during the trials. Most did not even shift their viewpoint for a better depth perception. The significantly high physical demand of AT was not only mentioned by several participants during the trials, but could also be observed by the investigators taking notes (e. g., subjects shook or rubbed their arm, rotated their shoulders between trials). Some participants also changed from the dominant to the non-dominant hand between trials to relax.

5 DISCUSSION AND RECOMMENDATIONS

We expected that the spatial interaction techniques M+D and M+R would outperform the touch-based techniques D+D and DR+D (***H1***). Our results strongly support this hypothesis, confirming earlier evidence in the related work [11, 62] that 3D interaction benefits from a high *degree of spatiality (D1)*. Several participants also highlighted the intuitiveness of M+D and M+R and that all actions could be performed simultaneously. In cases when the touch display is available as input modality for zooming, M+D should be preferred.

Regarding the difference between 2D and 3D tasks for pan & zoom, we expected D+D and DR+D to outperform the other techniques for 2D tasks (***H2***). Our results show that D+D and DR+D, although still slower than M+D and M+R, are far more suitable for 2D tasks than 3D tasks. We believe that this highlights the role of *guidance (D3)* for such tasks.

Based on the literature, we expected the mid-air gesture techniques AT, M+D, and M+R with a high *D1* to be more physically demanding than the touch techniques (***H3***). In fact, our results indicate that AT is by far the most physically demanding technique. However, there is no significant difference between the other techniques. Clearly, the added weight of a typical smartphone (including tracking markers in our study) is less of an issue than the unergonomic mid-air gestures on head level for the air tap interaction. It should be noted that in cases of large distance panning the necessary clutching may lead to a higher physical demand for M+R and M+D. However, our results do not reflect this.

We expected the mental demand to be highest for D+D and DR+D (***H4***). Here, our results show that DR+D is most challenging for the users. However, D+D rated fairly well despite being less direct than the other techniques. Of course, the design of the tasks necessitate a frequent turn of the smartphone to operate all DoF, due to the low *degree of simultaneity (D2)*. In application scenarios where users try to explore and compare items in a data space changing the reference plane would not happen that often. In this case we assume that D+D performs much better.

Finally, we expected a strong learning effect for AT (***H5***), as our experience showed that many users initially struggle with the air tap gesture. While we found a strong learning effect in general, there were no significant differences between the techniques. It is possible that the particularly strong physical demand of AT partially counteracted the learning effect for this technique because of gradual symptoms of fatigue. Looking at the reported suitability of the techniques, zooming was the main issue of AT. A possible explanation is that the double air tap gesture to trigger zooming was especially hard to execute for our participants, perhaps also induced by the limited camera field for gesture recognition.

As already mentioned, we observed that nearly all of the participants stayed at the same place during each trial and did rarely move. We anticipated that the users would walk around to inspect the visualization and therefore installed additional logging of the head position for later analysis. We assume that the stereoscopic presentation of the 3D data space and especially the relative panning and zooming was sufficient for the spatial depth perception required to solve the tasks. This in turn means that 3D visualizations in AR can be comfortably and successfully explored with our techniques even if the available physical space or existing obstacles do not allow to freely walk around.

Limitations

There are several limitations of our study. All participants have a similar background and nearly the same age. As such, they are not necessarily representing the population at large. During our study we experienced some tracking jitter and lags that may have influenced some of the results. However, they would have had the largest impact on M+D and M+R and as those techniques did perform especially well, we do not believe this issue to be of much concern. Finally, while we chose the interaction techniques systematically based on three design dimensions to cover a wide range of interaction styles, other techniques and variants could possibly have been studied. We believe that our findings also extend to most of those interaction techniques.

We focused on 3D pan & zoom, i. e., translation and uniform scaling. However, most of our techniques could easily be extended for rotation or other interactions like, e. g., focus selection for zooming. For example, M+D could be extended to use device orientation to control rotation. 3D zoom focus point selection could be realized via spatial input: for M+D and M+R using mode switches, for D+D and DR+D without. Furthermore, in accordance with our design goals, all our techniques are unimanual, leaving room for use-case specific extensions of the interaction techniques.

Recommendations

Overall, we recommend spatial interaction techniques such as M+D or M+R using a spatially aware mobile device for 3D exploration tasks involving pan & zoom. These techniques are feasible with today's hardware. They exhibit comparably high *degrees of spatiality and simultaneity*. Touch on the device is especially suitable for zooming but might not be available if other actions are mapped to touch input or if the smartphone is used as an additional display.

Mid-air hand gestures are also suitable in general. However, in comparison they are physically more demanding and we only recommend them for short interactions, or when device-free interaction is required. We believe this to be true not only for the air tap gesture, which we specifically tested, but also mid-air hand gestures in general, most of which would be even more complicated than the air tap.

The two touch-centric techniques D+D and DR+D exhibit a greater *degree of guidance* and more separated DoF with lower *degrees of spatiality and simultaneity*. Based on our study results, we cannot recommend them for general purpose pan & zoom of 3D data spaces. However, they show much better performance for 2D tasks, where guidance becomes an important factor. As such, they might be suitable for use cases where two dimensions are favored and the third dimension is only rarely used or secondary in nature. As described earlier, this may be the case when a visualization is oriented along a surface or wall, or if the data space is flat.

6 CONCLUSION AND FUTURE WORK

In this paper, we investigated the use of smartphones as interactive unimanual controllers for the exploration of 3D data spaces in head-mounted Augmented Reality (AR). On the basis of a well-considered design space we identified available input channels for spatial and touch-based interaction and their combination for the integral task of 3D pan & zoom. We designed four compound techniques and implemented them as well as an air-tap baseline technique. We then evaluated these five techniques in a controlled lab experiment with 25 participants regarding performance in completion time, efficiency, learning effects, and usability aspects. The results show that spatial interaction using smartphones as tangible input outperforms the other techniques for 3D pan & zoom regarding efficiency and user preferences.

Our work lays the foundations for efficient 3D data exploration in head-mounted AR using casual smartphone-based input techniques. We encourage researchers to build on our findings and design and conduct further studies and experiments, e. g., addressing differences in age and background of users, but also the comparison of the proposed techniques across AR/VR different settings. Other use cases of our techniques are conceivable that involve 3D translation and scaling, e. g., virtual object positioning in AR, also in combination with rotation. Another direction for future research is the investigation of domain-specific, real-world data sets and their individual exploration tasks.

ACKNOWLEDGMENTS

We thank Ricardo Langner for his invaluable feedback as well as all our participants for making this study possible. This work was funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden and DFG grant CollabWall (DA 1319/11-1).

REFERENCES

[1] Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. 2010. Reconfigurable Tangible Devices for 3D Virtual Object Manipulation by Single or Multiple Users. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology (VRST '10)*. ACM, New York, NY, USA, 227–230. https://doi.org/10.1145/1889863.1889913

[2] Adrien Arnaud, Jean-Baptiste Corrégé, Céline Clavel, Michèle Gouiffès, and Mehdi Ammi. 2016. Exploration of Virtual Environments on Tablet: Comparison Between Tactile and Tangible Interaction Techniques. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*. ACM, New York, NY, USA, 357–361. https://doi.org/10.1145/2993148.2993186

[3] Michel Beaudouin-Lafon. 2000. Instrumental Interaction: An Interaction Model for Designing post-WIMP User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 446–453. https://doi.org/10.1145/332040.332473

[4] Louis-Pierre Bergé, Emmanuel Dubois, and Mathieu Raynal. 2015. Design and Evaluation of an "Around the SmartPhone" Technique for 3D Manipulations on Distant Display. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 69–78. https://doi.org/10.1145/2788940.2788941

[5] Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. 2014. Exploring Smartphone-based Interaction with Overview+Detail Interfaces on 3D Public Displays. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI '14)*. ACM, New York, NY, USA, 125–134. https://doi.org/10.1145/2628363.2628374

[6] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. 2017. Hybrid Tactile/Tangible Interaction for 3D Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 881–890. https://doi.org/10.1109/TVCG.2016.2599217

[7] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. 2017. Mouse, Tactile, and Tangible Input for 3D Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4727–4740. https://doi.org/10.1145/3025453.3025863

[8] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272. https://doi.org/10.1561/1100000049

[9] Frédéric Bourgeois, Yves Guiard, and Michel Beaudouin-Lafon. 2001. Pan-Zoom Coordination in Multi-Scale Pointing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*. ACM, New York, NY, USA, 157–158. https://doi.org/10.1145/634067.634163

[10] Wolfgang Büschel, Annett Mitschick, and Raimund Dachselt. 2018. Here and Now: Reality-Based Information Retrieval. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 171–180. https://doi.org/10.1145/3176349.3176384

[11] Wolfgang Büschel, Patrick Reipschläger, Ricardo Langner, and Raimund Dachselt. 2017. Investigating the Use of Spatial Interaction for 3D Data Visualization on Mobile Devices. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 62–71. https://doi.org/10.1145/3132272.3134125

[12] William Buxton and Brad Myers. 1986. A Study in Two-handed Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '86)*. ACM, New York, NY, USA, 321–326. https://doi.org/10.1145/22627.22390

[13] Tom Chandler, Maxime Cordeil, Tobias Czauderna, Tim Dwyer, Jaroslaw Glowacki, Cagatay Goncu, Matthias Klapperstueck, Karsten Klein, Kim Marriott, Falk Schreiber, et al. 2015. Immersive Analytics. In *2015 Big Data Visual Analytics (BDVA)*. 1–8. https://doi.org/10.1109/BDVA.2015.7314296

[14] Huiwen Chang and Michael F. Cohen. 2017. Panning and Zooming High-Resolution Panoramas in Virtual Reality Devices. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 279–288. https://doi.org/10.1145/3126594.3126617

[15] Aurélie Cohé, Fabrice Dècle, and Martin Hachet. 2011. tBox: A 3D Transformation Widget Designed for Touch-Screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3005–3008. https://doi.org/10.1145/1978942.1979387

[16] Raimund Dachselt and Robert Buchholz. 2009. Natural Throw and Tilt Interaction Between Mobile Phones and Distant Displays. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. ACM, New York, NY, USA, 3253–3258. https://doi.org/10.1145/1520340.1520467

[17] Florian Daiber, Marco Speicher, Sven Gehring, Markus Löchtefeld, and Antonio Krüger. 2014. Interacting with 3D Content on Stereoscopic Displays. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*. ACM, New York, NY, USA, 32:32–32:37. https://doi.org/10.1145/2611009.2611024

[18] Dragoş Datcu, Stephan Lukosch, and Frances Brazier. 2015. On the Usability and Effectiveness of Different Interaction Types in Augmented Reality. *International Journal of Human–Computer Interaction* 31, 3 (2015), 193–209. https://doi.org/10.1080/10447318.2014.994193 arXiv:https://doi.org/10.1080/10447318.2014.994193

[19] David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2017. PocketThumb: A Wearable Dual-Sided Touch Interface for Cursor-based Control of Smart-Eyewear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 9 (June 2017), 17 pages. https://doi.org/10.1145/3090055

[20] Neven AM ElSayed, Bruce H Thomas, Kim Marriott, Julia Piantadosi, and Ross T Smith. 2015. Situated Analytics. In *2015 Big Data Visual Analytics (BDVA)*. 1–8. https://doi.org/10.1109/BDVA.2015.7314302

[21] Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction (SUI '16)*. ACM, New York, NY, USA, 99–102. https://doi.org/10.1145/2983310.2985757

[22] Jinbo Feng, Isaac Cho, and Zachary Wartell. 2015. Comparison of Device-Based, One and Two-Handed 7DOF Manipulation Techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 2–9. https://doi.org/10.1145/2788940.2788942

[23] Gene V Glass, Percy D Peckham, and James R Sanders. 1972. Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of educational research* 42, 3 (1972), 237–288.

[24] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, Luciana Nedel, and Anderson Maciel. 2017. Design and Evaluation of a Handheld-Based 3D User Interface for Collaborative Object Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5881–5891. https://doi.org/10.1145/3025453.3025935

[25] Mark Hancock, Otmar Hilliges, Christopher Collins, Dominikus Baur, and Sheelagh Carpendale. 2009. Exploring Tangible and Direct Touch Interfaces for Manipulating 2D and 3D Information on a Digital Table. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '09)*. ACM, New York, NY, USA, 77–84. https://doi.org/10.1145/1731903.1731921

[26] Chris Hand. 1997. A Survey of 3D Interaction Techniques. *Computer Graphics Forum* 16, 5 (Dec. 1997), 269–281. https://doi.org/10.1111/1467-8659.00194

[27] Anders Henrysson, Joe Marshall, and Mark Billinghurst. 2007. Experiments in 3D Interaction for Mobile Phone AR. In *Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia (GRAPHITE '07)*. ACM, New York, NY, USA, 187–194. https://doi.org/10.1145/1321261.1321295

[28] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A Survey of Design Issues in Spatial Input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology (UIST '94)*. ACM, New York, NY, USA, 213–222. https://doi.org/10.1145/192426.192501

[29] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 625, 12 pages. https://doi.org/10.1145/3290605.3300855

[30] Paul Issartel, Florimond Guéniat, and Mehdi Ammi. 2014. A Portable Interface for Tangible Exploration of Volumetric Data. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST '14)*. ACM, New York, NY, USA, 209–210. https://doi.org/10.1145/2671015.2671130

[31] Paul Issartel, Florimond Guéniat, Tobias Isenberg, and Mehdi Ammi. 2017. Analysis of Locally Coupled 3D Manipulation Mappings Based on Mobile Device Motion. *Presence: Teleoper. Virtual Environ.* 26, 1 (Feb. 2017), 66–95. https://doi.org/10.1162/PRES_a_00287

[32] Bret Jackson, Tung Yuen Lau, David Schroeder, Kimani C. Toussaint Jr., and Daniel F. Keefe. 2013. A Lightweight Tangible 3D Interface for Interactive Visualization of Thin Fiber Structures. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2802–2809. https://doi.org/10.1109/TVCG.2013.121

[33] Bret Jackson, Tung Yuen Lau, David Schroeder, Kimani C. Toussaint Jr., and Daniel F. Keefe. 2013. A Lightweight Tangible 3D Interface for Interactive Visualization of Thin Fiber Structures. *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2802–2809. https://doi.org/10.1109/TVCG.2013.121

[34] Robert J. K. Jacob and Linda E. Sibert. 1992. The Perceptual Structure of Multidimensional Input Device Selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 211–218. https://doi.org/10.1145/142750.142792

[35] Jacek Jankowski and Martin Hachet. 2013. A Survey of Interaction Techniques for Interactive 3D Environments. In *Eurographics 2013 - STAR*. Girona, Spain. https://hal.inria.fr/hal-00789413

[36] Brett Jones, Rajinder Sodhi, David Forsyth, Brian Bailey, and Giuliano Maciocci. 2012. Around Device Interaction for Multiscale Navigation. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 83–92. https://doi.org/10.1145/2371574.2371589

[37] Konstantin Klamka and Raimund Dachselt. 2015. Elasticcon: Elastic Controllers for Casual Interaction. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 410–419. https://doi.org/10.1145/2785830.2785849

[38] Konstantin Klamka and Raimund Dachselt. 2018. ARCord: Visually Augmented Interactive Cords for Mobile Interaction. In *Proceedings of the 2018 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, 6. https://doi.org/10.1145/3170427.3188456

[39] Sven Kratz, Michael Rohs, Dennis Guse, Jörg Müller, Gilles Bailly, and Michael Nischt. 2012. PalmSpace: Continuous Around-Device Gestures vs. Multitouch for 3D Rotation Tasks on Mobile Devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*. ACM, New York, NY, USA, 181–188. https://doi.org/10.1145/2254556.2254590

[40] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2018. Augmented Reality versus Virtual Reality for 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 24, 2 (Feb. 2018), 1038–1048. https://doi.org/10.1109/TVCG.2017.2658570

[41] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 81:1–81:14. https://doi.org/10.1145/3173574.3173655

[42] Chi-Jung Lee and Hung-Kuo Chu. 2018. Dual-MR: Interaction with Mixed Reality Using Smartphones. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. ACM, New York, NY, USA, Article 79, 2 pages. https://doi.org/10.1145/3281505.3281618

[43] Andrea Leganchuk, Shumin Zhai, and William Buxton. 1998. Manual and Cognitive Benefits of Two-handed Input: An Experimental Study. *ACM Trans. Comput.-Hum. Interact.* 5, 4 (Dec. 1998), 326–359. https://doi.org/10.1145/300520.300522

[44] David López, Lora Oehlberg, Candemir Doger, and Tobias Isenberg. 2016. Towards An Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* 22, 5 (May 2016), 1616–1629. https://doi.org/10.1109/TVCG.2015.2440233

[45] Paul Lubos, Gerd Bruder, and Frank Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 11–18. https://doi.org/10.1109/3DUI.2014.6798834

[46] 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

[47] Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and J. Jorge. 2018. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum* 0, 0 (2018). https://doi.org/10.1111/cgf.13390

[48] Alexandre Millette and Michael J. McGuffin. 2016. DualCAD: Integrating Augmented Reality with a Desktop GUI and Smartphone Interaction. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 21–26. https://doi.org/10.1109/ISMAR-Adjunct.2016.0030

[49] Mark R. Mine, Frederick P. Brooks, Jr., and Carlo H. Sequin. 1997. Moving Objects in Space: Exploiting Proprioception in Virtual-environment Interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 19–26. https://doi.org/10.1145/258734.258747

[50] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human movement science* 20, 6 (2001), 791–805.

[51] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air Pan-and-zoom on Wall-sized Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 177–186. https://doi.org/10.1145/1978942.1978969

[52] Anh Nguyen and Amy Banic. 2015. 3DTouch: A Wearable 3D Input Device for 3D Applications. In *2015 IEEE Virtual Reality (VR)*. 55–61.

https://doi.org/10.1109/VR.2015.7223324

[53] Alex Olwal and Steven Feiner. 2009. Spatially Aware Handhelds for High-Precision Tangible Interaction with Large Displays. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (TEI '09)*. ACM, New York, NY, USA, 181–188. https://doi.org/10.1145/1517664.1517705

[54] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. 1997. A Framework and Testbed for Studying Manipulation Techniques for Immersive VR. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '97)*. ACM, New York, NY, USA, 21–28. https://doi.org/10.1145/261135.261141

[55] Alejandro Rodríguez and Alejandro León. 2015. Smartphone-based remote 3D interaction for digital heritage applications. In *2015 Digital Heritage*, Vol. 1. 297–300. https://doi.org/10.1109/DigitalHeritage.2015.7413886

[56] Shun Sekiguchi and Takashi Komuro. 2017. VolGrab: Realizing 3D View Navigation by Aerial Hand Gestures. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2064–2069. https://doi.org/10.1145/3027063.3053108

[57] Mickael Sereno, Lonni Besançon, and Tobias Isenberg. 2019. Supporting Volumetric Data Visualization and Analysis by Combining Augmented Reality Visuals with Multi-Touch Input. EuroVis. , 3 pages. https://hal.inria.fr/hal-02123904 Poster.

[58] Jie Song, Gábor Sörös, Fabrizio Pece, Sean Ryan Fanello, Shahram Izadi, Cem Keskin, and Otmar Hilliges. 2014. In-Air Gestures Around Unmodified Mobile Devices. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 319–329. https://doi.org/10.1145/2642918.2647373

[59] Peng Song, Wooi Boon Goh, Chi-Wing Fu, Qiang Meng, and Pheng-Ann Heng. 2011. WYSIWYF: Exploring and Annotating Volume Data with a Tangible Handheld Device. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1333–1342. https://doi.org/10.1145/1978942.1979140

[60] Marco Speicher, Florian Daiber, Sven Gehring, and Antonio Krüger. 2016. Exploring 3D Manipulation on Large Stereoscopic Displays. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*. ACM, New York, NY, USA, 59–66. https://doi.org/10.1145/2914920.2915018

[61] Martin Spindler, Wolfgang Büschel, and Raimund Dachselt. 2012. Use Your Head: Tangible Windows for 3D Information Spaces in a Tabletop Environment. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 245–254. https://doi.org/10.1145/2396636.2396674

[62] Martin Spindler, Martin Schuessler, Marcel Martsch, and Raimund Dachselt. 2014. Pinch-Drag-Flick vs. Spatial Input: Rethinking Zoom & Pan on Mobile Displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1113–1122. http://doi.acm.org/10.1145/2556288.2557028

[63] Sophie Stellmach and Raimund Dachselt. 2012. Investigating Gaze-supported Multimodal Pan and Zoom. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 357–360. https://doi.org/10.1145/2168556.2168636

[64] Junwei Sun, Wolfgang Stuerzlinger, and Dmitri Shuralyov. 2016. SHIFT-Sliding and DEPTH-POP for 3D Positioning. In *Proceedings of the 2016 Symposium on Spatial User Interaction (SUI '16)*. ACM, New York, NY, USA, 69–78. https://doi.org/10.1145/2983310.2985748

[65] Vanessa Vuibert, Wolfgang Stuerzlinger, and Jeremy R. Cooperstock. 2015. Evaluation of Docking Task Performance Using Mid-Air Interaction Techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 44–52. https://doi.org/10.1145/2788940.2788950

[66] 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. https://doi.org/10.1145/3290605.3300849

[67] Chadwick A. Wingrave, Yonca Haciahmetoglu, and Doug A. Bowman. 2006. Overcoming World in Miniature Limitations by a Scaled and Scrolling WIM. In *3D User Interfaces (3DUI'06)*. 11–16. https://doi.org/10.1109/VR.2006.106

[68] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko. 2018. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1653–1660. https://doi.org/10.1109/TVCG.2018.2794222

[69] Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Difeng Yu, and Diego Monteiro. 2019. DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 444, 14 pages. https://doi.org/10.1145/3290605.3300674

[70] Lingyun Yu, Pjotr Svetachov, Petra Isenberg, Maarten H Everts, and Tobias Isenberg. 2010. FI3D: Direct-Touch Interaction for the Exploration of 3D Scientific Visualization Spaces. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov. 2010), 1613–1622. https://doi.org/10.1109/TVCG.2010.157