Breaking the Screen: Interaction Across Touchscreen Boundaries in Virtual Reality for Mobile Knowledge Workers

Verena Biener[1]   Daniel Schneider[1]   Travis Gesslein[1]   Alexander Otte[1]   Bastian Kuth[1]

Per Ola Kristensson[3]   Eyal Ofek[2]   Michel Pahud[2]   Jens Grubert[1]*

[1]Coburg University of Applied Sciences [2]Microsoft Research
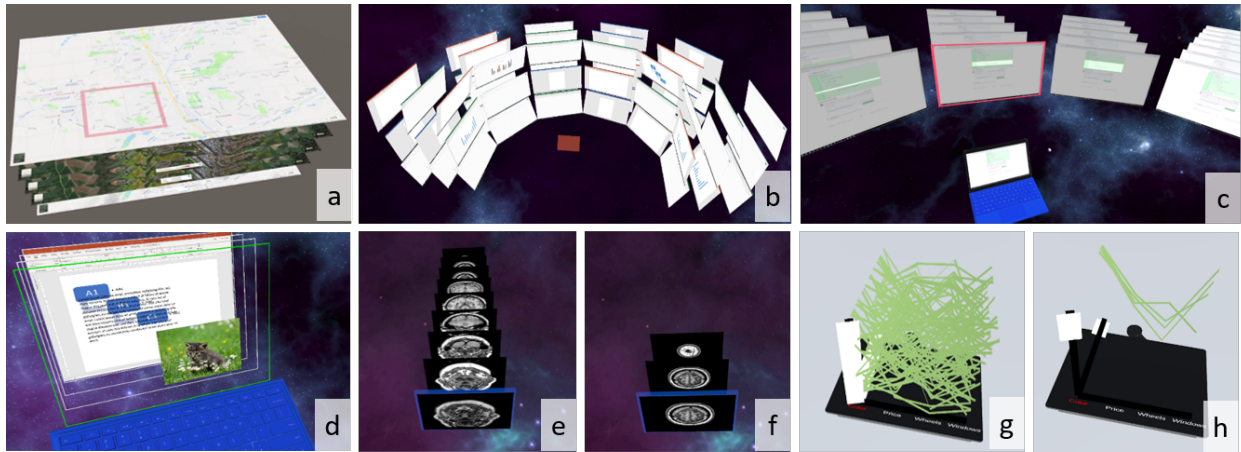[3]University of Cambridge

Fig. 1. Mobile knowledge worker applications using a touchscreen in VR implemented in this paper. a: map navigation, b: window manager, c: code version control, d: presentation editor, e-f: medical imaging, g-h: information visualization.

Abstract— Virtual Reality (VR) has the potential to transform knowledge work. One advantage of VR knowledge work is that it allows extending 2D displays into the third dimension, enabling new operations, such as selecting overlapping objects or displaying additional layers of information. On the other hand, mobile knowledge workers often work on established mobile devices, such as tablets, limiting interaction with those devices to a small input space. This challenge of a constrained input space is intensified in situations when VR knowledge work is situated in cramped environments, such as airplanes and touchdown spaces.

In this paper, we investigate the feasibility of interacting jointly between an immersive VR head-mounted display and a tablet within the context of knowledge work. Specifically, we 1) design, implement and study how to interact with information that reaches beyond a single physical touchscreen in VR; 2) design and evaluate a set of interaction concepts; and 3) build example applications and gather user feedback on those applications.

Index Terms—virtual reality, knowledge work, mobile office, window management, eye tracking, multimodal interaction

1 INTRODUCTION

Recent progress in virtual reality (VR) technology makes it possible to provide knowledge workers with a portable virtual office. This office can provide many potential advantages for mobile work, such as 1) providing a well-illuminated, private environment with wide display areas regardless of the physical surroundings; 2) enabling a virtual recreation of a consistent spatial workspace; and 3) relieving the user of physical limitations, using large and even distant displays while the user is remaining seated and the hands are resting on a table [39, 81].

The vision of a spatial user interface supporting knowledge work has been investigated for many years (e.g. [92, 94]). However, the recent emergence of consumer VR headsets now make it feasible to explore the design of deployable user interface solutions.

*contact author: jens.grubert@hs-coburg.de

However, while VR is promising for mobile office work it also brings its own additional challenges. For example, mobile workers often work in front of touchscreens, such as tablets or laptops. Within this work, we propose to extend the limited output and input space of such configurations by utilizing VR head-mounted displays (HMDs). Specifically, recent HMDs with inside out camera-based sensing allow for pose tracking of the user's head and hands. Through access to HMDs cameras, spatial tracking of further objects, such as screens or keyboards become feasible.

Given these opportunities, this paper explores how to leverage the large display area and additional three-dimensional display volume around the user afforded by an HMD while using established capacitive touch interfaces provided with tablets and some laptops.

This paper addresses this scenario with the following contributions. First, we explore and evaluate how to spatially arrange and manipulate information within the joint interaction space of HMD-tablet interaction. Second, we design, implement and evaluate six applications (window manager, code version control, parallel coordinates exploration, map navigation, volumetric image viewer, and a virtual PowerPoint), see Figure 1.

2 RELATED WORK

The work in this paper is underpinned by the following research areas: 1) mixed reality (MR) for knowledge work; 2) information windows in spatial environments; and 3) spatial interaction.

2.1 Mixed Reality for Knowledge Work

Recent prior work has begun exploring how to support knowledge work using MR [39, 44, 81, 96]. Early work investigated the use of projection systems to extend physical office environments for interaction with physical documents (e.g., [54, 88, 94, 122]). VR and AR HMDs have also been explored for interacting with physical documents (e.g., [36, 62]). Most prior work has focused on annotating documents displayed on a 2D surface while this work investigates the use of space surrounding a planar piece of information.

In addition to enhanced document interaction, prior work has also explored remote presence applications (e.g., [83, 92]). There is a number of publications investigating the use of VR in desktop-based environments for tasks such as text entry (e.g., [41, 53, 70]), system control [132, 133] and visual analytics [120]. Büschel et al. [15] surveyed a wide range of immersive interaction techniques for visual analytics. Previous research on productivity desktop-based VR has concentrated on the use of physical keyboards [99], controllers and hands [55, 133], and, recently, tablets [111]. The closest work to this paper is by Surale et al. [111], which focuses on spatial manipulation and creation of 3D objects. In contrast, we use the tablet for information management on virtual displays.

Complementary to this prior research, this paper aims to support mobile knowledge workers by extending commonly used tools such as tablets and notebooks through HMDs.

2.2 Information Windows in Spatial Environments

In 1993, Feiner et al. [31] introduced head-surrounding and world reference frames for positioning 3D windows in VR. In 1999, Mark Billinghurst and Thad Starner [8] introduced the spatial display metaphor, in which information windows are arranged on a virtual cylinder around the user.

Since then, virtual information displays have been explored in various reference systems, such as world-, object-, head-, body- or device-referenced [57]. Specifically, interacting with windows in body-centered reference systems [119] has attracted attention, for instance to allow fast access to virtual items [19, 61], mobile multi-tasking [27, 29] and visual analytics [28]. Gugenheimer et al. [43] introduced face touch, which allows interacting with display-fixed user interfaces (using direct touch) and world-fixed content (using raycasting). Yu et al. [129] investigated the use of body motions for switching interfaces in VR. Lee et al. [58] investigated positioning a window in 3D space using a continuous hand gesture. Petford et al. [84] compared the selection performance of mouse and raycast pointing in full coverage displays (not in VR). Recently, Jetter et al. [50] proposed to interactively design a space with various display form factors in VR. Wagner et al. proposed a desk metaphor for controlling visual analytics that reappropriates a physical desk in VR [120].

Prior work has also explored how to support user interaction across HMDs and mobile and wearable touch displays. Grubert et al. [37] presented body-aligned, device-aligned, and side-by-side modes for interacting between a touch display (smartphone or smartwatch) and an optical see-through HMD. Similar explorations have followed suit using video-see-through HMDs [80], an extended set of interaction techniques [131], using smartwatches [66, 123, 126], or with a focus on understanding smartphone-driven window management techniques for HMDs [95]. In a similar vein, prior work has studied the interaction across HMDs and stationary displays, such as tabletops [16, 93, 102].

Most prior work relate to the issue of field of view, that is, how to display and access more information by spreading it around the user in multiple windows. In this research, we are additionally interested in extending the information display in the depth direction. Such a display is suited for displaying different views of information (or layers) that are semantically connected by their 2D location. On the other hand, we use a very limited input space: While most referred prior work span all the angular range of the display as an input space, we only use the interaction space on or near the tablet in order to support interaction in constrained physical spaces [39, 71].

2.3 Spatial Interaction

A large number of techniques for selection, spatial manipulation, navigation, and system control have been proposed to support spatial user interfaces [57]. Regarding object selection, Argelaguet et al. [2] presented a survey on 3D selection techniques. For a recent survey on 3D virtual object manipulation, we refer to Mendes et al. [72]. Finally, recent surveys [14, 38] have extensively reviewed spatial interaction across mobile devices, mostly in non-VR settings.

In addition to unimodal techniques the combination of touch with mid-air has attracted attention from researchers. For example, outside of VR, Müller et al. [75] investigated the use of touch and mid-air interaction on public displays, Hilliges et al. [46] studied tabletop settings. Several researchers have proposed to use handheld touchscreens in spatial user interfaces for tasks such as sketching, ideation and modeling (e.g., [3, 23, 25, 33, 48, 91]), navigation of volumetric data [106], 3D data exploration [65] and system control [10]. Spatial manipulation has mostly been studied in single-user settings (e.g., [5, 51, 63, 68, 74, 111, 112]) but also in collaborative settings [35].

Evolving from the magic lens [7] and tangible interaction concepts [117], tangible magic lenses allow users to access and manipulate otherwise hidden data in interactive spatial environments. A wide variety of interaction concepts have been proposed within the scope of information visualization (e.g., recent surveys [114, 115]). Both rigid shapes (e.g., rectangular [107]) or circular [108] and flexible shapes (e.g., [109]) have been used, as well as various display media (e.g., projection on cardboard [17, 107]), transparent props [12, 98], handheld touchscreens [40, 59], or virtual lenses [64, 82]. In addition, the combination of eye-gaze with other modalities such as touch [85, 86], mid-air gestures [87, 97, 101] and head-movements [56, 103, 104] has been recently investigated for interaction in spatial user interfaces. For a recent survey on gaze-based interaction in AR and VR, see Hirzle et al. [47].

We draw on these rich sources of interaction ideas and adopt techniques in the context of VR interaction with touchscreens for mobile knowledge workers. Our work complements multimodal techniques combining touch and mid-air [46, 75], gaze-based techniques [85] and ideas for combining HMDs with touchscreens [37, 80, 131] through novel techniques for accessing virtual windows around or behind a physical touchscreen.

3 DESIGN SPACE

Currently most knowledge workers' applications are designed for 2D displays, which are the vast majority of displays in the world. VR and AR displays enable mobile workers to take their displays with them on the go. In addition, since HMDs are stereoscopic, they enable information presentation beyond the 2D display and the ability to manipulate this information by spatial manipulation. Researchers have already proposed various schemes for arranging (planar) information in a spatial context relative to the user. Specifically, prior work has proposed various reference systems, such as world-, object-, head-, body- or device-referenced [9], spatial window layouts in these reference frames, for example, scaffolding geometry such as spheres or cylinders [27, 28], and input modalities to access and manipulate information in these spaces (such as touch, gaze, mid-air interaction or multimodal interaction, c.f. [15, 47, 119]).

The guiding principle in this paper is that an information touchscreen display can break out of the screen in VR and transition from 2D to 3D— yet still be controlled by manipulating the original touchscreen. This is in part motivated by prior work reviewed in the previous section and in part motivated by the fact that a touchscreen allows the user to provide precise 2D input while displaying information in an HMD provides the ability to display information in 3D. This screen-centric manipulation using touch can be complemented with additional modalities, such as gaze tracking.

Two important aspects arising in this context are how to *spatially arrange* information relative to the touchscreen and how to *map the input from the physical touchscreen to the information in the virtual world*, given the joint input capabilities of 2D touch sensing (on the screen) and 3D spatial manipulation (around the screen).

3.1 Spatial Arrangement

Prior work has investigated options on how to align multiple 2D information windows around the user (e.g., [27,29]). One common approach is to arrange windows on a scaffolding geometry, such as spheres or cylinders [8](see Figure 2, b-d). In a spatially unconstrained environment, accessing this information is often realized using virtual hand or raycasting techniques. While using such extended display areas is possible, we should be aware that in constrained spaces, such as on an airplane or in touchdown spaces, it may not be usable for interaction as ample spatial movements can be unsuitable either due to social acceptability [1, 124] or simply due to lack of physical space to operate [39, 71].

An alternative display option is extending information along with the depth of the screen, extruding it into 3D (see Figure 2, a). This extrusion might be specifically suitable for additional information that is directly related to the document displayed on the physical screen. This information is often semantically different and should ideally be displayed separately from the main screen but with a spatial relation (specifically with corresponding x- and y-coordinates) to the document. For example, when reviewing a document, a common task consists of inserting comments. However, there is no natural space for the added comments within the document, and adding them in place can disrupt the layout of the document. Adding the comments in a separate layer, hovering in front of the main document, maintains the contextual relevance while enabling the user to focus on the layer of interest.

3.2 Input-Output Mapping

The touchscreen allows one layer of information at a time to align with the surface of the screen. This layer enjoys the easiest and most accurate input, as the screen supports the user's hand and enables the highest positional sensing, and, sometimes, pressure sensing. Next, we discuss, how to access layers of information that break out of the bounds of the physical screen.

3.2.1 Around the Screen Interaction

One challenge when interacting across screens is how to support both fine-grained selection on an active screen as well as efficient switching between screens. Naive implementations that use a fixed control-to-display gain (CD gain), such as stitching or pointer warping for bridging the space between displays [77, 121, 127], do not scale well to large displays due to the required trade-off between high precision pointing in a given region of interest on a display and fast switching between multiple regions of interest [78]. Hence, besides raycasting, prior work has proposed multiple strategies for controlling a large output space using a small input space on a touchscreen (e.g., [32, 69, 78, 79]).

Inspired by those techniques, we propose using a *bimanual selection technique* for allowing both precise input and efficient switching of the active window. This way, the user can move a cursor and select or deselect items inside the active window by simply using the tablet as a touchpad with a suitable CD gain for single screen interaction, see Figure 2, b. If the user touches the bezel of the touchscreen, which has a width of 2 cm, with their non-dominant hand, a coarse CD gain is activated to enable fluid switching between multiple windows, see Figure 2, c.

The CD gain for single screen interaction is a one-to-one mapping of the movement on the touchscreen. The CD gain for switching between multiple windows is set in such a way that the user has to move their finger 2 cm to switch to the next screen. We evaluated this CD gain along with values of 1 cm and 3 cm in an informal user study with five participants and found that 2 cm was the fastest and preferred by three participants while the other two CD gains were only preferred by one participant each.

As VR headsets also support head-pointing and partially eye-gaze tracking we implemented a second technique: *combined gaze and touch interaction*. Inspired by previous work on combining gaze and manual pointing [85, 130], we devised a technique in which the combined head and eye gaze provides the context for touch interaction, see Figure 2, d. With this technique regions of interest are discrete (the individual virtual screens). Hence, when a user gazes at a specific window, the cursor is transferred to that new screen, maintaining the same absolute position in the new screen as in the old screen (i.e., the same (x,y) coordinate).

In order to avoid unwanted cursor movements it is possible to use temporal thresholding (e.g., 200-300 ms [85, 130]), although we did not use this in our evaluation in order to allow for faster task completion times. To allow selection of items at the screen boundary, we use spatial thresholding with an empirically determined threshold of 5% of the window size in all directions. Specifically, to switch from one screen to another, the user first has to move their eyes beyond the second display boundary by that threshold. While we did not evaluate this threshold parameter in a formal user study, internal testing revealed that this threshold circumvents unwanted display switches and still allows for comfortable switching between display boundaries.

For both techniques, interacting with the now active virtual display can either 1) happen at the original location of the display, essentially turning the touchscreen into an indirect input device; or 2) occur by aligning the virtual display with the physical touchscreen. For example, dragging the finger on the bottom bezel of the touchscreen (or using a two-finger swipe) can rotate virtual windows around the x-axis of a scaffolding geometry (either display-by-display in discrete steps or continuously).

3.2.2 Depth Interaction

Accessing virtual displays along the depth dimension of the touchscreen can happen either in front, on, or behind the physical screen (see Figure 2, a). For virtual displays in front of the screen, direct mid-air selection might be used, but this may cause occlusion of content at layers further away and hinder interaction on the physical screen [13]. Accessing content behind the screen has to take into account the physical presence of the touchscreen, preventing direct access using virtual hand techniques. Users would either need to grab behind the screen using a virtual hand or use raycasting techniques that can handle object occlusion (e.g., [30]). Instead, we propose to use the touchscreen to select a desired virtual display. Depending on the number of virtual displays spread behind or in front of the physical screen, different mappings can be appropriate. We experimented with swiping along the bezel of the screen and two-finger swipe gestures, commonly available on multi-touch trackpads. Depending on the number of layers an absolute mapping for quick random access to layers or a relative mapping for navigating between adjacent layers might be appropriate. In an informal user study with three users, we found that swiping with two fingers outperformed bezel swipes and a relative mapping outperformed an absolute mapping for four and ten layers. Hence, we used this input mapping for the further development of the applications. We used the same CD gain as for switching between individual screens spread around the active screen, as described above.

While the input technique allows to quickly switch between layers it does not completely resolve potential occlusion issues. Occlusion can be mitigated (but not completely resolved) utilizing parallax if the user moves the head. We experimented with two visualization techniques to further mitigate potential occlusion. One option is to amplify the parallax effect when moving the head to the side of the screen, see Figure 7, f. Another option, inspired by explosion diagrams, is to temporarily scatter virtual screens around the physical screen, see Figure 7, e. While the explosion diagram technique allows the user to view all layers simultaneously and unoccluded, the number of concurrently visible layers is constrained by the available screen space. Also for both techniques, potential dependencies between layers are impacted (e.g., when there is a need for carefully arranging multiple graphical elements into a composite image). Another limitation of the explosion diagram technique is the reduction of the ability to associate
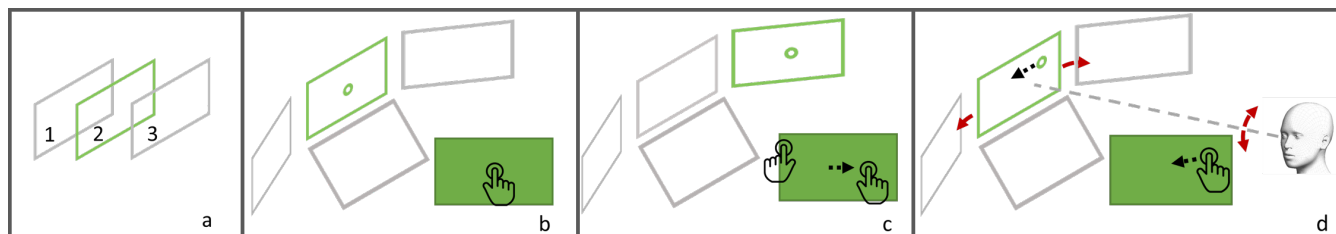
Fig. 2. a: Virtual displays (grey) arranged relative to the physical screen (green): behind (1), with (2), in front of (3). b and c: Changing between pointing on a single screen and switching between screens using a bimanual technique. b: If only a single finger is used on the touchscreen (filled green rectangle), it controls the cursor on the active screen (green boundary) with a suitable control-display gain. c: The screens are switched by pressing with a finger of the non-dominant hand on the bezel of the touchscreen and moving the finger of the dominant screen towards the second screen (in this case the right screen) using a second CD-ratio. d: Changing between pointing on a single screen and switching between screens using a combination of touch and gaze. The finger on the touch screen (filled green rectangle) controls the cursor on the active screen (green boundary). Gaze provides the region of interest. If the user gazes to the side, up or down, the corresponding screen gets activated (with appropriate temporal and spatial thresholds).

different objects between different layers based on their common 2D locations.

3.2.3　Single Screen Interaction

Even an individual layer that is aligned with the physical screen does not need to retain its physical bounds. The ability to display the view frustum according to the rotation of the user's head enables a virtual display that potentially spreads over a very large angular part of the user's field of view. It enables the simulation of screens that may be bigger than physical screens available to the user (see Figure 7, d). In this case, the mapping of the input space of the physical touchscreen can become indirect. For example, it is possible to use a different CD gain for absolute mapping of the physical finger position to the virtual finger position (following ideas from haptic retargeting [4, 20]), see Figure 7, or the touchscreen can be operated using a relative mapping. In some cases, it might be desirable to retain a direct mapping between the input and the output space. In this case, only the portion co-located with the physical touchscreen allows interactivity. Changing this active input area can either be realized by two-finger swipes or input on the bezels of the screen (analogous to implementations in desktop sharing applications, such as TeamViewer) or by redefining the active area by gaze (again requiring an appropriate clutch mechanism).

4　EVALUATION OF DESIGN PARAMETERS

While we designed the techniques following an iterative approach with multiple design iterations consisting of conceptualization, implementation, and initial user tests (eating your own dog food) [24, 118], we aimed to understand properties of the proposed interaction space in this paper in more detail. Within the proposed interactive space, users can both extend the display area around the current display on a two-dimensional proxy geometry as well as in-depth in front or behind the physical touch screen. For the evaluation, we investigated these two properties (around and in-depth) separately.

To this end, we first aimed at exploring, if using combined touch and gaze interaction has benefits over touch-only interaction when interacting across multiple virtual screens arranged on a proxy geometry around the touchscreen. Second, we wanted to quantify the benefits of viewing multiple stacked information layers in-depth behind a touchscreen compared to only showing a single layer at a time as is common in many applications today. While adding depth cues has been indicated to improve task performance in various settings [6, 89, 90, 105], we aimed at quantifying any performance improvement within the scope of the joint tablet-HMD interaction.

Hence, we investigated those two aspects in a user study. In the first part (subsequently called the CONTENT TRANSFER TASK), we compared the performance of the *bimanual selection technique* with the COMBINED GAZE AND TOUCH INTERACTION technique in a content transfer task for a small and large number of screens.
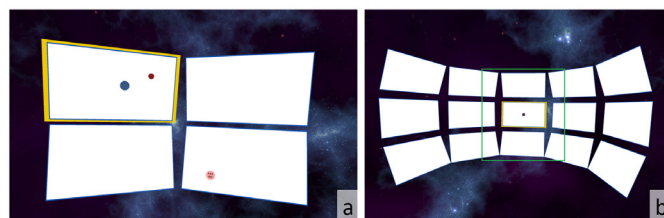

Fig. 3. Arrangements of the virtual screens for the CONTENT TRANSFER TASK. a: FOUR SCREENS: the blue dot can be acquired by the red cursor in the upper left screen and is to be placed on the goal in the lower right screen. b: FIFTEEN SCREENS: the green border indicates the field of view of the user.

In the second part (subsequently called the PUZZLE TASK), participants were asked to solve a puzzle task where each puzzle piece was displayed in an individual layer, mimicking tasks when composing presentations using multiple shapes or images using multiple layers in an image editing application. Note that we used the baseline technique described in Section *Depth Interaction* without amplified parallax or explosion diagram visualization due to the nature of the task having strong spatial dependencies between the *x*- and *y* position of the individual puzzle pieces.

For the CONTENT TRANSFER TASK, we followed a 2×2 within-subjects design. The first independent variable was INTERACTION TECHNIQUE for selecting the screens, which was done either GAZE-BASED or BIMANUAL as described above. The second independent variable was NUMBER OF SCREENS presented to the participant which was either FOUR SCREENS (Figure 3, a) or FIFTEEN SCREENS (Figure 3, b). We chose these configurations to better understand the performance of the techniques in presence of a few or many screens while still considering the field of view that can be comfortably covered when the user turns their head. This experimental design results in four different conditions, which are depicted in Figure 3. Dependent variables included tasks completion time (the duration from acquiring the content in one screen to placing it at the marked spot in another screen), accuracy (measured as Euclidean distance between the released item and the actual target location), usability as measured by the System Usability Scale (SUS) [11], workload as measured by NASA TLX (unweighted version) [45], simulator sickness (SSQ) [52] as well as user preferences. We hypothesized that the GAZE-BASED would outperform BIMANUAL in terms of task completion time, usability and workload, but not in terms of simulator sickness or accuracy (as the final movements were conducted with the finger of the dominant hand in both techniques).

For the PUZZLE TASK, we also followed a 2×2 within subjects design. The independent variables were VISUALIZATION with two levels: FLAT, a baseline condition in which all layers were displayed
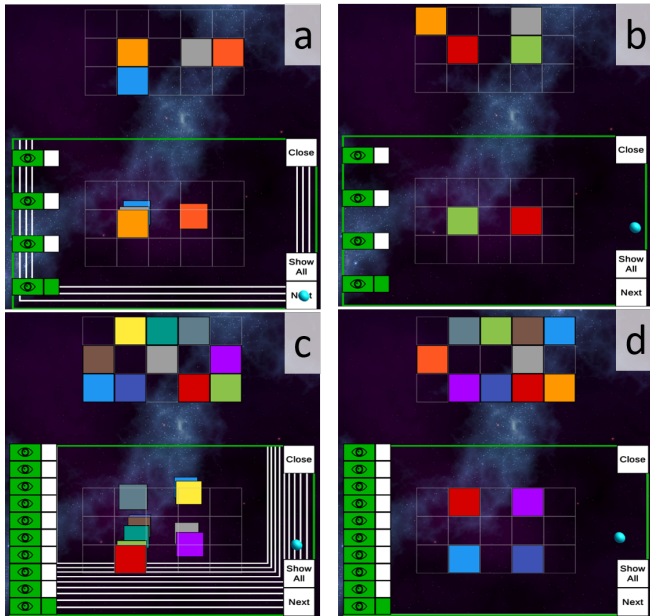
Fig. 4. Conditions for the PUZZLE TASK. a: DEPTH with FOUR LAYERS, b: FLAT with FOUR LAYERS, c: DEPTH with TEN LAYERS, d: FLAT with TEN LAYERS.

Fig. 5. The study setup. a: setup with motion tracked tablet, fiducials for the fingers and HTC Vive Pro Eye to the right. The lighthouse systems were placed on the ceiling around the table and the Optitrack system on a bar above the table. The monitor, keyboard and mouse to the left were operated by the experimenter to start individual conditions. b: Participant during the study.

at the same depth, see Figure 4, b and d, and DEPTH, where each layer was displayed with increasing z-distance, see Figure 4, a and c. The second independent variable was NUMBER OF LAYERS with two levels: FOUR LAYERS with four layers displayed (Figure 4, a) and TEN LAYERS in which ten layers were displayed (Figure 4, c). This was done to investigate if an increasing number of layers has an effect on the performance of the visualization techniques.

As can be seen in Figure 4, the active layer is highlighted with a green frame, the user's fingertip is visualized with a turquoise sphere. The target arrangement is shown above the puzzle's layered display as a flat image. To the left of the layered display there is an overview widget, inspired by similar views in image editing applications. The eye symbol toggles the visibility of the respective layer (green: visible) and the rectangles to the right of the eye symbol allow direct selection of a layer (green: active layer, white: inactive layer). By selecting a layer (pressing on the green square associated with it), the system moved this layer to align with the screen depth and made all layers in front of it transparent, thereby maximizing the current layer visibility. Switching between two adjacent layers can also be achieved by swiping vertically with two fingers on the touchpad. The button "Show All" collapses all layers into a single layer and activates the visibility of all layers. The button "Next" switches to the next task. The button "Close" was only visible during training and allowed participants to end the training session when they felt comfortable with conducting the task.

The dependent variables were, as in the CONTENT TRANSFER task, task completion time, usability (SUS), workload (TLX) as well as simulator sickness (SSQ). In addition, the system logged the number of incorrectly placed puzzle items. While we hypothesized that DEPTH would outperform FLAT in terms of task completion time due to the added depth cues, we were specifically interested in quantifying this difference. We also hypothesized that DEPTH would lead to significantly higher usability rating and lower workload compared to FLAT with no difference in simulator sickness. We still included SSQ to check if severe symptoms would occur.

4.1 Participants

We recruited 14 participants (5 female, 9 male, mean age 30.07 years, sd = 10.59, mean height 176.57 cm, sd = 8.54). All of them indicated prior VR experience. Four participants used head mounted VR devices very frequently, two often, four sometimes, three rarely and one only

once. One participant indicated she does not play video games, two rarely, four participants sometimes, three participants often and four participants very frequently. Seven participants wore contact lenses or glasses with corrected to normal vision. Eleven participants were right handed while three were left handed. All but one used their dominant hand to control the cursor and their non-dominant hand on the bezel. Thirteen participants used their index finger on the touchscreen and one used her middle finger.

4.2 Apparatus

The experiment took place in two locations: a laboratory environment in Europe and a home in the US. The study setup (except the PC) was replicated in both environments. A HTC Vive Pro Eye was used as VR HMD with built-in eyetracking. For touch input a Huawei Media Pad T5 was used (screen diagonal 10.1 inches, 16:10 aspect ratio). Velcro was attached to the left, right and bottom bezel of the tablet to support participants in identifying the touchscreen boundaries. The tablet was placed on a table in front of the participant in such a way that participants sitting in front of it on a chair could comfortably use it. The system was implemented in Unity 2019.3 and deployed on a PC with windows 10, an AMD Ryzen Threadripper 1950X 16-core processor with 3.4 Ghz, two Nvidia GeForce RTX 2080 graphics cards and 64 GB RAM in Europe and on a PC with Intel Core i9-9980HK 8-core processor, Nvidia GeForce RTX 2080 graphics card, 16 GB RAM in the US.

An OptiTrack V120 Trio tracking system was used for spatial tracking of index finger tips by using two rigid bodies with retro-reflective markers attached to them. Note that motion trajectories of the fingers were logged for comparative analysis in future experiments only, they were not analyzed within the scope of this paper. The Vive Lighthouse tracking system was used to track the VR HMD and the tablet. However, while the position of the tablet could have been changed by the participants during the experiment, no participant chose to do so. The study setup is shown in Figure 5.

For the CONTENT TRANSFER TASK, multiple screens were placed around the participant's head as shown in Figure 3, a and b. The virtual screen size was 24 inches and virtual screens were initially placed at a distance of 90 cm around the participant's head (if the participant moved their head forwards or backwards this distance changed). The horizontal angle spanned by the screens was at most 190 degrees and the vertical angle less than 30 degrees upwards and less than 35 degrees downwards. This allowed the participant to comfortably look at all screens. The cursor position was indicated as a disc with a diameter of 2.5 cm. For the present study, we chose to use no temporal threshold in order to obtain a comparable switching effect between both techniques.

The PUZZLE TASK used the same screen size and the same distance to the participant. The individual layers were placed 2.5 cm apart (this parameter was determined empirically).

4.3 Tasks

For the CONTENT TRANSFER TASK several windows were placed in a circle around the participant's head. Then they had to select a disk with a diameter of about 4 cm on one screen to a target area with the same diameter on another screen as shown in Figure 3 a. The window with the next target area was randomly selected while making sure that both small and longer distances were included during the condition. The disk and the target area appeared at one of eight randomly selected places around the center of a window. Both INTERACTION TECHNIQUES used the tablet as a touch pad to control the cursor within the active screen and to grab and release the dot via long-press. After successful placement of the disk onto the target area (with at least a partial overlap of the two), the task was repeated with a different arrangement of start and target areas.

For the PUZZLE TASK, participants were asked to arrange puzzle pieces (5×5 cm) as indicated by a template image, see Figure 4. Each puzzle piece was placed on a separate layer. The puzzle pieces snapped into a predefined 5×3 grid to facilitate accurate placement.

Participants were asked to advance to the next puzzle if they felt they had completed the task. The task was then repeated with a different template.

4.4 Procedure

After an introduction, participants were first asked to fill out a demographic questionnaire. Then the HTC Vive Pro Eye eye-calibration procedure was performed to ensure a working eye-tracking calibration for each participant. Thereafter the participants either started with the CONTENT TRANSFER TASK or the PUZZLE TASK (counterbalanced). For both tasks, the conditions were ordered using a balanced Latin square. This resulted in four permutations which could not be equally distributed among 14 participants. However, no significant order effects were detected. For each condition in the CONTENT TRANSFER TASK, participants performed a training phase where they completed ten content transfers. After that, they performed the actual task 32 times. The number of repetitions was chosen so that each distance and orientation in the FIFTEEN SCREENS condition appeared once (1–4 screens horizontally $\times$ 1–2 screens vertically $\times$ two directions each = $8 \times 2 \times 2 = 32$). For each condition in the PUZZLE TASK, participants also performed a training phase where they completed puzzle tasks until they felt comfortable (most users only conducted one practice task). Thereafter, they performed the actual puzzle task ten times per condition. In both tasks, after each condition, participants completed the SSQ, SUS and NASA TLX questionnaires. At the end of each block they also answered a questionnaire about their preferences and participated in a semi-structured interview. Finally, the participants were thanked and compensated with voucher worth 10 Euro.

For the CONTENT TRANSFER TASK, we collected 32 repetitions $\times$ 4 conditions $\times$ 14 users = 1792 data points for the performance data and 4 conditions $\times$ 14 users = 56 data points for the subjective feedback. For the PUZZLE TASK, we collected 10 repetitions $\times$ 4 conditions $\times$ 14 users = 560 data points for the performance data and 4 conditions $\times$ 14 users = 56 data points for the subjective feedback.

4.5 Results

Statistical significance tests for log-transformed target acquisition time was carried out using general linear model repeated measures analysis of variance (RM-ANOVA) with Holm-Bonferroni adjustments for multiple comparisons at an initial significance level $\alpha = 0.05$. We indicate effect sizes whenever feasible (η_p^2).

For subjective feedback, or data that did not follow a normal distribution or could not be transformed to a normal distribution using the log-transform (such as errors), we employed the Aligned Rank Transform [125] before applying RM-ANOVAs.

The analysis of results did not reveal significant differences between the participant pool in the US and the one in Europe. Hence, we will solely report the joint results.

Due to logging errors in the CONTENT TRANSFER TASK, we only obtained data from 31 instead of 32 tasks. For one participant we only received about half of the data in one GAZE-BASED condition due to

technical problems. Four values for the distance between goal and disc were excluded, because they were much higher than should be possible when successfully placing the disc. Note that slightly fewer values for some participants should not affect the overall results as we always used the mean performance for each participant in each condition.

The results in the following sections can be summarized as follows: For the CONTENT TRANSFER TASK, participants acquired targets significantly faster with the GAZE-BASED method compared to BIMANUAL (ca. 30%) and the GAZE-BASED resulted in significantly higher SUS ratings. No significant differences regarding accuracy, number of errors or simulator sickness ratings were detected between conditions. All but one participant preferred the GAZE-BASED technique.

For the PUZZLE TASK, participants performed the task significantly faster with DEPTH visualization compared to FLAT (approximately 15 %) but also made significantly more errors. The DEPTH visualization resulted in significantly lower mental demand and resulted in a significantly higher usability rating compared to FLAT.

4.5.1 Performance

For the CONTENT TRANSFER TASK there was a significant main effect of INTERACTION TECHNIQUE for task completion time, such that the GAZE-BASED method ($M = 3.70\ s, SD = 0.98$) resulted in a shorter task completion time than the BIMANUAL method ($M = 5.49\ s, SD = 2.33$). As expected, the main effect of the NUMBER OF SCREENS on the task completion time was also significant such that FOUR SCREENS ($M = 3.66\ s, SD = 1.20$) resulted in a shorter task completion time than FIFTEEN SCREENS ($M = 5.52\ s, SD = 2.20$). This was predictable because moving the target across multiple columns and rows takes longer than moving it only across two columns and rows. There was no significant interaction effect between INTERACTION TECHNIQUE and NUMBER OF SCREENS with respect to task completion time. There was no significant differences between the conditions for accuracy in placing the target on the goal. The performance results for the CONTENT TRANSFER TASK can be seen in Figure 6 and the results of the RM-ANOVA in Table 1.

For the PUZZLE TASK there was a significant main effect of VISUALIZATION for task completion time such that the DEPTH method ($M = 37.17\ s, SD = 22.79$) resulted in a shorter task completion time than the FLAT method ($M = 43.31\ s, SD = 29.52$). As expected, the main effect of NUMBER OF LAYERS for task completion time was significant such that the FOUR LAYERS ($M = 21.28\ s, SD = 8.87$) resulted in a shorter task completion time than TEN LAYERS ($M = 59.19\ s, SD = 24.23$). No significant interactions have been found between NUMBER OF LAYERS and VISUALIZATION. Analysis of the error data (using the total number of errors across all ten repetitions) indicated that both the NUMBER OF LAYERS and the VISUALIZATION method had a significant impact on the number of errors made. The conditions using the DEPTH method ($M = 0.29, SD = 0.6$) had significantly more errors than the conditions using the FLAT method ($M = 0.07, 0.26$). Furthermore, conditions with FOUR LAYERS ($M = 0.04, SD = 0.19$) resulted in significantly fewer errors than conditions with TEN LAYERS ($M = 0.32, 0.61$).

4.5.2 Simulator Sickness, Workload, Usability

For the CONTENT TRANSFER TASK we found that the Total Severity aspect of the Simulator Sickness Questionnaire was significantly influenced by the INTERACTION TECHNIQUE such that BIMANUAL ($M = 14.43, SD = 20.73$) resulted in a higher total severity than GAZE-BASED ($M = 13.35, 20.57$). Also the Oculo-motor aspect was lower for the GAZE-BASED than for the BIMANUAL. However, the values are very low and pairwise comparisons showed no significant differences. There was a significant difference for the overall TLX results, such that BIMANUAL resulted in a higher taskload than GAZE-BASED. Also the mental, physical, effort and frustration results of the TLX were significantly higher for the BIMANUAL conditions than for the GAZE-BASED conditions. The perceived performance of the participants, however, was significantly higher for the BIMANUAL method ($M = 27.14, SD = 19.31$) compared to the GAZE-BASED method ($M = 21.79, SD = 21.65$). This is in contrast to the findings

Content Transfer Task

	Task Completion Time					Accuracy					TS-SS					SUS					Overall Taskload				
	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2
I	1	13	63.66	<.001	.83		13	.79	.39	.06	1	13	11.11	.005	.46	1	13	38.42	<.001	.75	1	13	29.81	<.001	.7
S	1	13	99.41	<.001	.88	1	13	.53	.48	.04	1	13	.60	.45	.04	1	13	3.79	.07	.23	1	13	3.02	.11	.19
I × S	1	13	.86	.37	.06	1	13	1.2	.29	.08	1	13	2.0	.18	.13	1	13	.87	.37	.06	1	13	.14	.72	.01

Puzzle Task

	Task Completion Time					Errors					TS-SS					SUS					Overall Taskload				
	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2	df_1	df_2	F	p	η_p^2
V	1	13	5.32	.04	.29	1	13	8.97	.04	.28	1	13	12.45	.004	.49	1	13	10.39	.007	.44	1	13	12.34	.004	.49
L	1	13	1574.78	<.001	.99	1	13	5.14	.04	.28	1	13	4.04	.07	.24	1	13	.61	.45	.04	1	13	15.4	.002	.54
V × L	1	13	1.23	.29	.09	1	13	4.7	.05	.03	1	13	1.86	.2	.13	1	13	.08	.78	.0	1	13	.03	.86	.0

Table 1. RM-ANOVA results for both tasks. Gray rows show significant findings. I = INTERACTION TECHNIQUE, S = NUMBER OF SCREENS, V = VISUALIZATION, L = NUMBER OF LAYERS. TS-SS: Total Severity Dimension of the Simulator Sickness Questionnaire. SUS: System Usability Scale $df_1 = df_{effect}$ and $df_2 = df_{error}$.
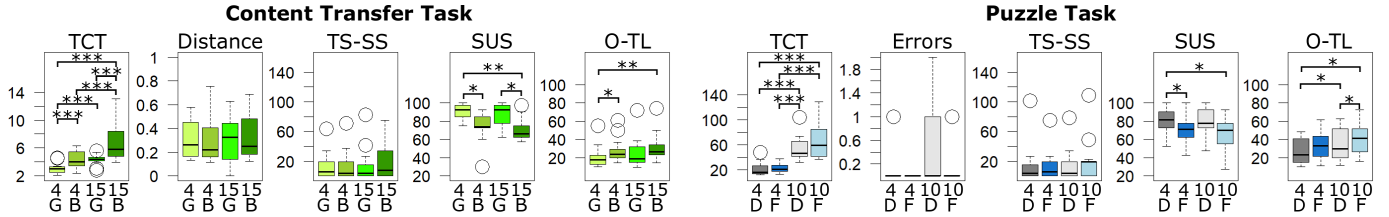

Fig. 6. Results of the task completion time (TCT), the Total Severity aspect of the Simulator Sickenss Questionnaire (TS-SS), the System Usability Scale (SUS) and the overall taskload (O-TL) for the CONTENT TRANSFER TASK and the PUZZLE TASK. Also for the CONTENT TRANSFER TASK the distance between disc and goal in cm upon task completion is shown and for the PUZZLE TASK the mean number of total errors. On the *x*-axis the digits indicates the NUMBER OF SCREENS (4 or 15) or the NUMBER OF LAYERS (4 or 10) as well as the INTERACTION TECHNIQUE G = GAZE-BASED, B = BIMANUAL and the VISUALIZATION F = FLAT, D = DEPTH. The number of stars indicates the level of significance between the conditions (*** < 0.001 ** < 0.01 * < 0.05).

of the task completion time and distance to goal shown above, where task completion time was significantly shorter for GAZE-BASED, and no significant difference was found in distance to the goal. Furthermore, INTERACTION TECHNIQUE had a significant impact on the SUS score in such a way that the usability of the GAZE-BASED method ($M = 88.75, SD = 10.26$) was higher than the usability of the BIMANUAL method ($M = 72.59, SD = 13.63$). The main results of the three questionnaires are shown in Figure 6. Plots and detailed results on the further dimensions of simulator sickness and task load can be found in the supplementary material.

For the PUZZLE TASK, the results of the simulator sickness questionnaire indicated that VISUALIZATION had a significant impact on the Total Severity such that it was significantly lower for DEPTH ($M = 14.29$, $SD = 23.85$) than for FLAT ($M = 16.7$, $SD = 24.76$). This was also true for the Oculo-motor and disorientation aspect. Also, the NUMBER OF LAYERS significantly influenced the disorientation aspect, such that the FOUR LAYER had a significantly lower value than TEN LAYER. However, the values are again very low and no significant differences were found in pairwise comparisons. The NASA TLX results indicated that the VISUALIZATION significantly influenced the overall task load such that it was significantly higher for the FLAT visualization ($M = 38.3, SD = 16.06$) than for the DEPTH visualization ($M = 30.92$, $SD = 15.92$). This was also the case for the mental and physical task load, as well as the frustration perceived by the participants. The perceived performance, however, was significantly higher for FLAT ($M = 29.64, SD = 21.73$) than for DEPTH ($M = 22.86, SD = 22.17$) which again, is in contrast to the results of the task completion time. In addition, we found that the number of layers had a significant impact on the overall task load such that it was lower for FOUR LAYERS ($M = 30.83$, $SD = 14.67$) than for TEN LAYERS ($M = 38.39$, $SD = 17.17$). The same was observed for frustration, physical task load and the temporal aspect. Also, we found that VISUALIZATION had a significant impact on the SUS results, in such a way that DEPTH ($M = 80.09$, $SD = 14.05$) resulted in a higher usability than FLAT ($M = 68.48$, $SD = 17.00$). The main results of the three questionnaires is shown in Figure 6 and Table 1. Plots and detailed results on the further dimensions of simulator sickness and task load can be found in the supplementary material.

4.5.3 Preference, Open Comments and Observations

For the CONTENT TRANSFER TASK, all but one participant preferred the GAZE-BASED method over the BIMANUAL, regardless of the number of screens. However, when asked about their preference for each NUMBER OF SCREENS separately this participant also liked GAZE-BASED more for FIFTEEN screens. When asked which interaction method the participants perceived as faster, all participants chose GAZE-BASED. When asked which method they could interact with more precisely, all participants chose GAZE-BASED for the conditions with FIFTEEN SCREENS. However, in the condition with FOUR SCREENS, only four participants thought that GAZE-BASED allowed for a more precise interaction and three participants felt that BIMANUAL was more precise. This is also supported by the objective measurements shown in Figure 6, even though no significant difference between conditions was indicated for distance to the target. We conjecture that these differences are due to the fact that in the GAZE-BASED method the selection of the target screen, and moving to the target within the screen, *can* happen concurrently. In contrast, in the BIMANUAL condition the selection of the target screen is *always* separated by an explicit mode switch (raising the finger of the non-dominant hand from the bezel) from the subsequent phase of moving towards the target within that screen. The lower precision in the BIMANUAL condition with FIFTEEN SCREENS compared to FOUR SCREENS might also be due to higher fatigue when moving the item across more screens. The participant who preferred BIMANUAL explained that it reminds him of the keyboard (alt+tab). Most other participants explained their preference for GAZE-BASED due to it being faster. One participant found GAZE-BASED more intuitive, and another participant found it less fatiguing. One participant suggested a hybrid solution where GAZE-BASED is only enabled when touching the bezel.

For the PUZZLE TASK, all but one participant preferred the DEPTH condition. However, one participant preferred FLAT. This participant later mentioned having trouble focusing the right eye and therefore found it more difficult to match the 3D state with the 2D target. Two participants thought the FLAT visualization was faster for FOUR LAYERS than the DEPTH visualization. All other participants felt that DEPTH was faster. Regarding the DEPTH visualization, participants also mentioned that "If you make a mistake it is easier to find the layer with the wrong element" (Participant P01), "I can better see which puzzle piece

is where" (P02), "you directly see where the puzzle piece is" (P04) and that DEPTH "provides more information" (P06).

Participants used different strategies to solve the puzzle task for the FLAT conditions. Three participants mainly identified the element belonging to the current layer by trial and error. They tried to move the element and if it was not possible they tested the next element until they found the element of that layer. Four participants started using this strategy but then switched to a more efficient one. One participant used it sporadically in the condition with FOUR LAYERS.

Some participants moved to the last layer (furthest away) at the start of each puzzle and continued to the front. In the last layer, only one element (the element of that layer) was visible, therefore the participant did not have to guess the correct element. When moving one layer, a new element appeared on the screen, and the participant could identify the next element with ease. Nine participants mainly used this strategy.

There were also participants who started from the front layer but shortly switched to the next layer and back in order to detect the active element as this element will then disappear and reappear. Two participants switched to this strategy after first using another strategy. Also, this strategy was basically always used when participants wanted to correct a mistake.

The fourth strategy observed was to click on the eye-symbols in the menu at the left to deactivate and activate the current layer. This would also show the user which element would disappear and reappear and therefore was the element of the active layer. This strategy was only used by one participant in his first FLAT condition. Thereafter he switched strategy. Two people also tried this strategy during the training phase.

5 APPLICATIONS

Based on the insights presented in the previous sections, we implemented six applications that we envision are applicable in mobile knowledge worker scenarios.

Window Manager: We implemented a window manager that allows the arrangement of multiple windows around the user (see Figure 1, b). Interaction with those windows is supported by joint spatial and touch interaction. For example, the active window is selected by head gaze (indicated by a red border around the window in Figure 7 a and b). The input space of the physical touchscreen (depicted as a red rectangle in Figure 7, b) is then mapped to the output space of the virtual window. Contrary to the content transfer task in the evaluation of design parameters, for the window manager, the selected window stays selected even if the user's gaze is switching to another window for as long as the user is touching the physical touchscreen. This design lets the user glance at other windows without losing the context of the task. Other clutching mechanisms are also feasible, such as locking the gaze-selected window on touch-down but then delaying the release on touch-up to let the user interact with the touchscreen.

To switch depth layers, the user can either swipe along the bezel or use a two-finger swipe, as described in the Section "Depth Interaction". Alternatively, to retrieve a temporary preview of the otherwise hidden layers, the user can lean towards the virtual screen to peak through (Figure 7, c). For interaction with virtual windows that are larger than the physical touchscreen, retargeting can be used, as described in the Section "Single Screen Interaction" (Figure 7, d).

Code Version Control: We implemented an interface for code version control (Figure 1, c) that uses a spatial layout around the physical screen, see Figure 1, c. Using on-screen pinches the user can select different scopes of code changes (line, block, function, class, file), swipe through the commits using the selected scope, and swipe the desired commit down to the physical screen for further editing.

Parallel Coordinates: Inspired by recent research in immersive analytics [21, 22, 67, 113], we built a parallel coordinates plot that is grounded on the tangible physical screen, see Figure 1, g, and h. Subranges of variables can be selected using on-surface swipes for the extent of the variable range and on-surface drags for the center of the variable range. Individual variables are selected by touch. Switching between coordinate axes (z, protruding the display, for the variable range,

y, along the height of the tablet, for the data items) can be achieved explicitly using a mode switch (e.g., pressing a soft button) or by mapping the swipe and drag along the x- and y-axis of the touchscreen.

Map: In the map application, users can navigate on a single layer without the map being cropped at the boundaries of the physical screen, see Figure 1, a . Available alternative map layers can be previewed by tilting the physical screen and be selected by swiping along the bezel of the screen.

Medical Imaging: In the medical imaging application, users can also swipe through different layers of information, see Figure 1, e–f. Additionally, the image slices can be previewed when the user moves their head to the side of the screen to look behind it.

PowerPoint: In the VR PowerPoint application users can arrange graphical items on a canvas, see Figure 1, d. Each item is associated with a separate layer. In the cases when items are occluded, they can be quickly accessed by swiping through the layers.

6 MOBILE IMPLEMENTATION

The system was implemented using Unity 2019.3. We used a HTC Vive Pro Eye as the HMD, which also enables hand tracking[1] as well as access to the camera streams[2]. While we used external OptiTrack outside-in tracking systems due to superior hand tracking accuracy compared to HMD-based inside out tracking [100] for our evaluations, we also implemented a system to track the user's screen using the HMD cameras, see Figure 7, g-h. While it is possible to build a computer vision algorithm to track the specific laptop model used by users, we used a model-independent approach. Since the laptop or tablet display screen is hidden from the user, wearing an HMD, we can use the screen to display a standard tracking pattern, such as ARUCO markers [76]. Displaying the pattern enables a robust detection and orientation using multiple wide-field-of view cameras of the screen and information about the laptop dimension are sufficient to display a virtual model that can guide the user touch gestures (see accompanying video). Other solutions, such as using an external camera, such as a laptop camera, to track the HMD are feasible but require additional instrumentation of the HMD [73].

To facilitate the community in further research and development of joint interactions between touchscreens and VR headsets, we make our code available under https://gitlab.com/mixedrealitylab/breakingthescreen.

7 USER FEEDBACK ON APPLICATIONS

The purpose of this user study was to learn from initial user reactions on the applications we previously introduced. To this end we demonstrated individual experiences to the user instead of carrying out task-based evaluations (following the approaches used in prior work, e.g., [18,99]). We recruited 17 participants (8 female, 9 male, mean age 31.9 years, sd = 9.2, mean height 173.9 cm, sd = 6.8) from a university campus with diverse study backgrounds. All participants were familiar with touch-sensitive screens.

7.1 Apparatus, Procedure, and Task

The same apparatus as in the evaluation of the design space parameters were used with the following change: instead of the Android tablet, a motion-tracked Surface Pro 4 tablet was used.

After an introduction, participants were asked to fill out a demographic questionnaire. Thereafter, we fixated OptiTrack trackers on the index finger of the dominant hand and the thumb of the other hand. Then individual applications were presented (counterbalanced). After trying out the applications the users rated them using a short three-item questionnaire capturing the user experience dimensions of ease of use, utility, and enjoyment (in line with similar approaches in prior work [18,60]). After trying out the applications, participants were asked

---

[1]https://developer.vive.com/resources/knowledgebase/vive-hand-tracking-sdk/ Last accessed April 24th, 2020.

[2]https://developer.vive.com/resources/knowledgebase/intro-vive-srworks-sdk/ Last accessed April 24th, 2020.
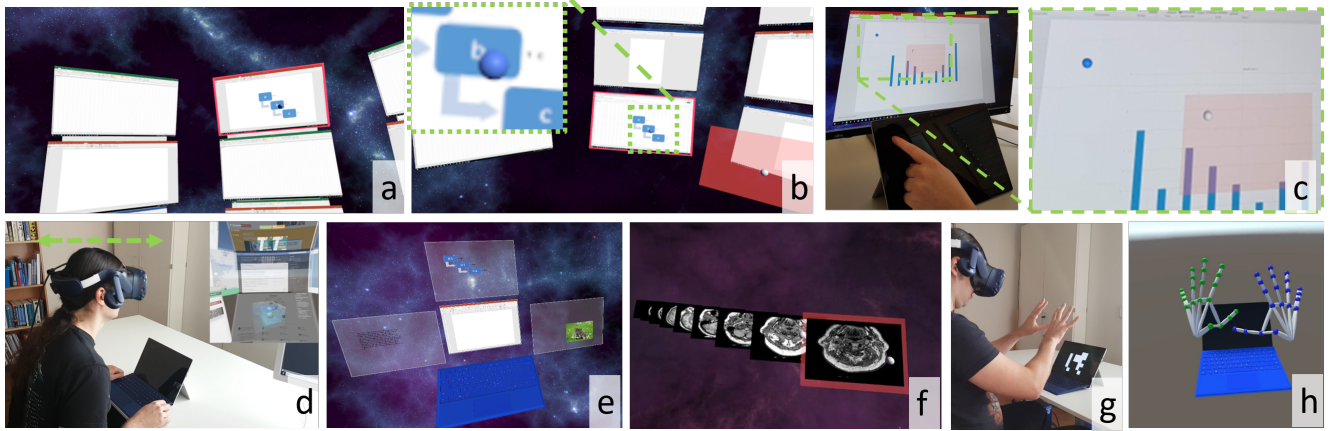
Fig. 7. a-d: Window Manager. a: An image is selected for movement to another window by head-gaze (selecting the window) + touch (selecting the image inside the window). b: The image is released on a different window at a different layer. The grey sphere on the red rectangle indicates the physical finger position. The blue sphere on the image depicts the projected finger position on the current window. c: The input on and above the physical touchscreen (red transparent rectangle, fingertip on touchscreen indicated by a white sphere) is retargeted to the virtual display (indicated by a blue sphere). d: A user previews windows in a hidden layer by leaning forward. e-f: Options to temporarily rearrange information layers. e: Virtual displays are temporarily arranged around the physical screen. f: If a user peeks behind the physical screen, virtual displays are extruded to that side. g-h mobile implementation: g: user wearing an HTC Vive Eye Pro looking on a tablet PC, which displays an Aruco marker. h: the user's fingers are spatially tracked with the HTC Vive Hand tracking SDK
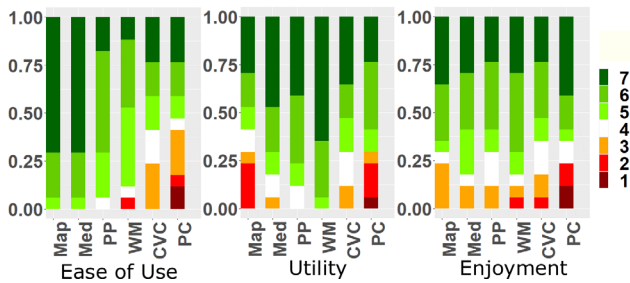

Fig. 8. Ease of Use, Utility and Enjoyment ratings for the evaluated applications on a range of 1 (strongly disagree) to 7 (strongly agree). MAP: Map Interaction, MED: MEDICAL, PP: POWERPOINT, WM: WINDOWMANAGER, CVC: CODEVERSIONCONTROL, PC: PARALLELCOORDINATES. The y-axis depcits the number of participants in percent (1.0 = 100% of participants).

to participate in a semi-structured interview. Finally, the participants were compensated with a 10 Euro voucher.

The order of the applications was balanced across participants, insofar as possible (a full permutation was not possible due to the number of applications).

7.2 Results

Figure 8 shows user ratings on seven-item Likert scales for questions on ease of use ("I found the application easy to use"), utility ("I found the application to be useful") and enjoyment ("I had fun interacting with the application"). Note that we did not run null hypothesis significance tests on these ratings as they should serve as a descriptive indication of these user experience dimensions only. Participants were asked to comment on individual applications. We followed top-down qualitative coding and structuring procedures to identify the benefits and drawbacks of the individual applications [110]. For WINDOWMANAGER the participants appreciated that the need for multiple physical monitors was mitigated ("Avoids the need of a monitor wall." participant P13) but also the layered view of the virtual displays ("You can view dozens of documents." P03) and it's ease of use ("It was easy to understand even for non-experts." P04). Regarding CODEVERSIONCONTROL, one participant mentioned that she had an "overview like in a airport control

tower." P13 and another one saw applicability to another domain: "I would like that for functional subunits in engineering" P17. Similarly, for POWERPOINT the participants saw a transfer to other domains with one mentioning that "these layers could also be used in drawing tools like Gimp or Photoshop" P14. Regarding, PARALLELCOORDINATES a participant mentioned that "it is a more entertaining way to present boring data" P06 and another participant, "The new visualization was nice" P13. Regarding MAP a participant mentioned that "I could need that for driving." P09 and regarding MEDICAL participants mentioned that "a 3D Model from the layers would be awesome" P14 but also that "It would be cool to pick the layers directly" P04.

7.3 Discussion

Using immersive HMDs enables mobile knowledge workers to bring their working environment with them everywhere they go [39]. Relative to the compact input and output space of today's mobile devices, the large field of view of HMDs allows to substantially extend the available display and increase privacy. Our design explorations and evaluations, explored different dimensions of this extended display space while focusing the input to the small area of a tablet screen. The first question we investigated was handling the large field of view of the virtual environment while using a very limited input space. To handle the large scale ratio between input and display, we tested two different techniques, One where we combined noisy eye gaze for large scale motions and touch for fine manipulation and compared it with a multi-touch technique. The results of the design parameter evaluation indicated that combination of gaze and touch input on a tablet as realized in the GAZE-BASED input method are beneficial for interaction in virtual multi-window environments. Specifically, the GAZE-BASED method outperformed the BIMANUAL METHOD by approximately 30%. The second parameter we studied was the use of the HMD's depth display for interaction above the touch screen. Using multiple layers has benefits of extending the amount of information that can be displayed and manipulated using a 2D screen. On the other-hand depth parallax and occlusions may hinder the user's performance. Our second study was designed to test it, by asking users to do a task both in 2D and in 3D. As indicated by the results, the use of depth can be both efficient and usable when compared to a 2D touch interaction. However, we were surprised by different strategies employed by some users to compensate for the lack of the depth structure in FLAT visualization, such as ordering their manipulations from back to front. Pairwise comparisons showed no

significant differences between the conditions with regard to errors and the overall number of errors was low (5.7% for DEPTH and 1.4% for FLAT). During interviews with participants, the five participants generating errors stated that they did not use the option to show all layers at the end of the task using the dedicated button shown in Figure 4. By design, the layers were arranged behind the physical screen, allowing the user to navigate them front to back. As previously noted, in the FLAT visualization nine participants first navigated to the last layer and then navigated to the front layer by layer. When they reached the front layer, all layers with all puzzle pieces were visible, hence there was no need to use the "Show all" button. The user feedback on the developed applications indicated that the prototypes were usable and enjoyable. Users could envision employing the prototypes in their work environments. We envision mobile VR to become an important tool for work, as it enables a large and private display on the go, regardless of the worker environment. However the physical environment of the user will not grow and the interactions should be designed accordingly. The presented applications are just a small probe of the complex design space encompassing all of the vast tasks information workers carry out, and this area is ripe for further research and development.

7.3.1 Use of Augmented Reality HMDs

The proposed interaction concepts, while being considered in VR, could be transferred to Augmented Reality (AR) and explored further. In AR, current generation optical see-through HMDs typically have a substantially smaller field of view compared to immersive VR HMDs. Hence, the proposed techniques might need adaptation. Potentially, off-screen visualization techniques could be integrated into an AR setting to compensate for this limited field of view [42]. Also, in AR displays without occlusion management, physical objects, such as the tablet or the user's hands are visible alongside the virtual objects, typically at a different focus distance, which can further lead to potential perceptual issues [26]. In contrast, combining an AR HMD with a physical tablet could make use of the potentially higher output resolution of the physical touch display and open up further possibilities for novel overview + detail techniques.

7.3.2 Limitations

Our work focused on a subset of a possibly large design space. While we foresee the combination of VR HMDs and tablets as one potentially promising direction on how mobile knowledge workers can be supported in the future, we are aware that it remains uncertain if and when such potential combinations will become products due to various factors such as technology, market, regulation, social acceptance and timing uncertainty [49]. For the CONTENT TRANSFER TASK, we tested both a small and a large number of virtual screens. However, other factors such as screen arrangement, screen size, or parameters of the interaction methods could impact the results. Although we foresee no substantial reversal of the effects indicated in the study, it may be interesting, as a future work, to test the limit of noisy eye gaze usage for very small screens and translation distances, following [116]. For the PUZZLE TASK we empirically determined the design parameters, such as distance between layers. Future work should investigate the effects of layer distance on various task types in more detail, as with increasing layer distance, the association between individual (x, y) coordinates on a layer could become harder for users to track. While we indicated the feasibility of mobile implementations using our prototype, we still opted for using an external tracking system for implementing our interaction techniques due to the limited accuracy of generation HMD-based hand tracking [100] at the time of the study. Newer technologies, such as Oculus Quest hand tracking offer higher accuracy and may enable sufficient performance. It remains to be evaluated how robust concurrent camera-based tracking of the hands, keyboard, the touchscreen, and, potentially, a pen is in real-world usage scenarios. An alternative approach may use a tablet to record the user's in air hand movement using a 45-degree mirror clipped above the tablet front camera [128].

8 CONCLUSIONS AND FUTURE WORK

Our work explored the opportunities arising out of a joint interaction space between VR HMDs and laptop/tablet capacitive touchscreens for supporting mobile information workers. We have demonstrated that such a joint interaction space is feasible to be implemented using consumer-oriented hardware by tracking the touchscreen and the user's finger, which could be useful in small constrained environments such as on an airplane or in touchdown spaces. We explored how to spatially arrange and interact with data within this joint interaction space. We further implemented a set of six applications (window manager, code version control, parallel coordinates exploration, map navigation, medical image viewer, and a virtual PowerPoint) making use of this design space. In future work, these interaction concepts should be studied in depth, such as recently done for spreadsheet interaction [34], to allow us to better understand their capabilities and limitations for supporting mobile knowledge workers in real-world working contexts. Also, additional study designs, e.g., for further investigations of the explosion diagram technique, comparisons with non-VR baseline techniques, effects of physical and virtual screen sizes, as well as the use of fully mobile implementations in real-world environments seem feasible. In particular, we see the potential for further multimodal interaction techniques that fit this shallow 3D space spanned by the hands resting on a 2D screen. One avenue of future work is to explore which particular document semantics are the most suitable for display above the documents (such as, for example, comments and other mark-up on a text document, notes on a code review), and which mappings fit different contexts. In addition, another avenue is to explore how to expand the interaction vocabulary with simultaneous pen and touch techniques.

REFERENCES

[1] D. Ahlström, K. Hasan, and P. Irani. Are you comfortable doing that?: acceptance studies of around-device gestures in and for public settings. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pp. 193–202. ACM, 2014.

[2] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013.

[3] R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, and K. Singh. 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*, p. 185. ACM, 2018.

[4] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pp. 1968–1979. ACM, 2016.

[5] T. Babic, H. Reiterer, and M. Haller. Pocket6: A 6dof controller based on a simple smartphone application. In *SUI'18: 6th ACM Symposium on Spatial User Interaction*, pp. 2–10, 2018.

[6] W. Barfield, C. Hendrix, and K.-E. Bystrom. Effects of stereopsis and head tracking on performance using desktop virtual environment displays. *Presence: Teleoperators & Virtual Environments*, 8(2):237–240, 1999.

[7] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: the see-through interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pp. 73–80. ACM, 1993.

[8] M. Billinghurst and T. Starner. Wearable devices: new ways to manage information. *Computer*, 32(1):57–64, 1999.

[9] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3D User interfaces: theory and practice, CourseSmart eTextbook*. Addison-Wesley, 2004.

[10] D. A. Bowman and C. A. Wingrave. Design and evaluation of menu systems for immersive virtual environments. In *Proceedings IEEE Virtual Reality 2001*, pp. 149–156. IEEE, 2001.

[11] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[12] L. D. Brown and H. Hua. Magic lenses for augmented virtual environments. *IEEE Computer Graphics and Applications*, 26(4):64–73, 2006.

[13] G. Bruder, F. Steinicke, and W. Sturzlinger. To touch or not to touch?: comparing 2d touch and 3d mid-air interaction on stereoscopic tabletop

surfaces. In *Proceedings of the 1st symposium on Spatial user interaction*, pp. 9–16. ACM, 2013.

[14] F. Brudy, C. Holz, R. Rädle, C.-J. Wu, S. Houben, C. N. Klokmose, and N. Marquardt. Cross-device taxonomy: Survey, opportunities and challenges of interactions spanning across multiple devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 562. ACM, 2019.

[15] W. Büschel, J. Chen, R. Dachselt, S. Drucker, T. Dwyer, C. Görg, T. Isenberg, A. Kerren, C. North, and W. Stuerzlinger. Interaction for immersive analytics. In *Immersive Analytics*, pp. 95–138. Springer, 2018.

[16] S. Butscher, S. Hubenschmid, J. Müller, J. Fuchs, and H. Reiterer. Clusters, trends, and outliers: How immersive technologies can facilitate the collaborative analysis of multidimensional data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[17] L. K. Chan and H. Y. Lau. Magicpad: the projection based 3d user interface. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 6(2):75–81, 2012.

[18] X. Chen, T. Grossman, D. J. Wigdor, and G. Fitzmaurice. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 159–168. ACM, 2014.

[19] X. A. Chen, N. Marquardt, A. Tang, S. Boring, and S. Greenberg. Extending a mobile devices interaction space through body-centric interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pp. 151–160. ACM, 2012.

[20] L.-P. Cheng, E. Ofek, C. Holz, H. Benko, and A. D. Wilson. Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3718–3728. ACM, 2017.

[21] M. Cordeil, B. Bach, A. Cunningham, B. Montoya, R. T. Smith, B. H. Thomas, and T. Dwyer. Embodied axes: Tangible, actuated interaction for 3d augmented reality data spaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020.

[22] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 71–83. ACM, 2017.

[23] T. Dorta, G. Kinayoglu, and M. Hoffmann. Hyve-3d and the 3d cursor: Architectural co-design with freedom in virtual reality. *International Journal of Architectural Computing*, 14(2):87–102, 2016.

[24] A. Drachen, P. Mirza-Babaei, and L. E. Nacke. *Games user research*. Oxford University Press, 2018.

[25] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio. Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020.

[26] A. Eiberger, P. O. Kristensson, S. Mayr, M. Kranz, and J. Grubert. Effects of depth layer switching between an optical see-through head-mounted display and a body-proximate display. In *Symposium on Spatial User Interaction*, pp. 1–9, 2019.

[27] B. Ens, J. D. Hincapié-Ramos, and P. Irani. Ethereal planes: a design framework for 2d information space in 3d mixed reality environments. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 2–12. ACM, 2014.

[28] B. Ens and P. Irani. Spatial analytic interfaces: Spatial user interfaces for in situ visual analytics. *IEEE computer graphics and applications*, 37(2):66–79, 2016.

[29] B. M. Ens, R. Finnegan, and P. P. Irani. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3171–3180. ACM, 2014.

[30] A. O. S. Feiner. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST'03*, pp. 81–82, 2003.

[31] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon. Windows on the world: 2 d windows for 3 d augmented reality. In *ACM Symposium on User Interface Software and Technology*, pp. 145–155, 1993.

[32] C. Forlines, D. Vogel, and R. Balakrishnan. Hybridpointing: fluid switching between absolute and relative pointing with a direct input device. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pp. 211–220, 2006.

[33] D. Gasques, J. G. Johnson, T. Sharkey, and N. Weibel. Pintar: Sketching spatial experiences in augmented reality. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, pp. 17–20. ACM, 2019.

[34] T. Gesslein, V. Biener, P. Gagel, D. Schneider, E. Ofek, M. Pahud, P. O. Kristensson, and J. Grubert. Pen-based interaction with spreadsheets in mobile virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020.

[35] J. G. Grandi, H. G. Debarba, L. Nedel, and A. Maciel. 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*, pp. 5881–5891. ACM, 2017.

[36] R. Grasset, A. Duenser, H. Seichter, and M. Billinghurst. The mixed reality book: a new multimedia reading experience. In *CHI'07 extended abstracts on Human factors in computing systems*, pp. 1953–1958. ACM, 2007.

[37] J. Grubert, M. Heinisch, A. Quigley, and D. Schmalstieg. Multifi: Multi fidelity interaction with displays on and around the body. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3933–3942. ACM, 2015.

[38] J. Grubert, M. Kranz, and A. Quigley. Challenges in mobile multi-device ecosystems. *mUX: The Journal of Mobile User Experience*, 5(1):5, 2016.

[39] J. Grubert, E. Ofek, M. Pahud, P. O. Kristensson, F. Steinicke, and C. Sandor. The office of the future: Virtual, portable, and global. *IEEE computer graphics and applications*, 38(6):125–133, 2018.

[40] J. Grubert, M. Pahud, R. Grasset, D. Schmalstieg, and H. Seichter. The utility of magic lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing*, 18:88–103, 2015.

[41] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Text entry in immersive head-mounted display-based virtual reality using standard keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 159–166. IEEE, 2018.

[42] U. Gruenefeld, A. E. Ali, W. Heuten, and S. Boll. Visualizing out-of-view objects in head-mounted augmented reality. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–7, 2017.

[43] J. Gugenheimer, D. Dobbelstein, C. Winkler, G. Haas, and E. Rukzio. Facetouch: Enabling touch interaction in display fixed uis for mobile virtual reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 49–60. ACM, 2016.

[44] J. Guo, D. Weng, Z. Zhang, H. Jiang, Y. Liu, Y. Wang, and H. B.-L. Duh. Mixed reality office system based on maslow's hierarchy of needs: Towards the long-term immersion in virtual environments. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 224–235. IEEE, 2019.

[45] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

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

[47] T. Hirzle, J. Gugenheimer, F. Geiselhart, A. Bulling, and E. Rukzio. A design space for gaze interaction on head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 625. ACM, 2019.

[48] K. Huo, K. Ramani, et al. 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*, pp. 37–45. ACM, 2017.

[49] H. Jalonen. The uncertainty of innovation: a systematic review of the literature. *Journal of Management Research*, 4(1):1, 2012.

[50] H.-C. Jetter, R. Rädle, T. Feuchtner, C. Anthes, J. Friedl, and C. N. Klokmose. "in vr, everything is possible!": Sketching and simulating spatially-aware interactive spaces in virtual reality. In *CHI Conference on Human Factors in Computing SystemsACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2020.

[51] N. Katzakis, R. J. Teather, K. Kiyokawa, and H. Takemura. Inspect: extending plane-casting for 6-dof control. *Human-centric Computing and Information Sciences*, 5(1):22, 2015.

[52] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simula-

tor sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[53] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 345. ACM, 2018.

[54] M. Kobayashi and H. Koike. Enhanceddesk: integrating paper documents and digital documents. In *Proceedings. 3rd Asia Pacific Computer Human Interaction (Cat. No. 98EX110)*, pp. 57–62. IEEE, 1998.

[55] P. G. Kry, A. Pihuit, A. Bernhardt, and M.-P. Cani. Handnavigator: Hands-on interaction for desktop virtual reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pp. 53–60. ACM, 2008.

[56] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 81. ACM, 2018.

[57] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[58] J. H. Lee, S.-G. An, Y. Kim, and S.-H. Bae. Projective windows: Bringing windows in space to the fingertip. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 218. ACM, 2018.

[59] S.-w. Leigh, P. Schoessler, F. Heibeck, P. Maes, and H. Ishii. Thaw: tangible interaction with see-through augmentation for smartphones on computer screens. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 89–96. ACM, 2015.

[60] J. R. Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq. *ACM Sigchi Bulletin*, 23(1):78–81, 1991.

[61] F. C. Y. Li, D. Dearman, and K. N. Truong. Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pp. 125–128. ACM, 2009.

[62] Z. Li, M. Annett, K. Hinckley, K. Singh, and D. Wigdor. Holodoc: Enabling mixed reality workspaces that harness physical and digital content. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 687. ACM, 2019.

[63] H.-N. Liang, C. Williams, M. Semegen, W. Stuerzlinger, and P. Irani. An investigation of suitable interactions for 3d manipulation of distant objects through a mobile device. *International Journal of Innovative Computing, Information and Control*, 9(12):4737–4752, 2013.

[64] J. Looser, R. Grasset, and M. Billinghurst. A 3d flexible and tangible magic lens in augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 51–54. IEEE, 2007.

[65] D. López, L. Oehlberg, C. Doger, and T. Isenberg. 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):1616–1629, 2015.

[66] Z. Lu, Y. Hu, and J. Dai. Watchar: 6-dof tracked watch for ar interaction. In *IEEE ISMAR 2019 Demonstrations*.

[67] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas. *Immersive Analytics*, vol. 11190. Springer, 2018.

[68] A. Marzo, B. Bossavit, and M. Hachet. 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*, pp. 13–16. ACM, 2014.

[69] D. C. McCallum and P. Irani. Arc-pad: absolute+ relative cursor positioning for large displays with a mobile touchscreen. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pp. 153–156, 2009.

[70] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 2143–2152. ACM, 2015.

[71] M. McGill, J. Williamson, A. Ng, F. Pollick, and S. Brewster. Challenges in passenger use of mixed reality headsets in cars and other transportation. *Virtual Reality*, pp. 1–21, 2019.

[72] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A

survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer Graphics Forum*, vol. 38, pp. 21–45. Wiley Online Library, 2019.

[73] P. Mohr, M. Tatzgern, T. Langlotz, A. Lang, D. Schmalstieg, and D. Kalkofen. Trackcap: Enabling smartphones for 3d interaction on mobile head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 585. ACM, 2019.

[74] A. Mossel, B. Venditti, and H. Kaufmann. 3dtouch and homer-s: intuitive manipulation techniques for one-handed handheld augmented reality. In *Proceedings of the Virtual Reality International Conference: Laval Virtual*, p. 12. ACM, 2013.

[75] J. Müller, G. Bailly, T. Bossuyt, and N. Hillgren. Mirrortouch: combining touch and mid-air gestures for public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pp. 319–328. ACM, 2014.

[76] R. Munoz-Salinas. Aruco: a minimal library for augmented reality applications based on opencv. *Universidad de Córdoba*, 2012.

[77] M. A. Nacenta, R. L. Mandryk, and C. Gutwin. Targeting across display-less space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 777–786, 2008.

[78] M. Nancel, O. Chapuis, E. Pietriga, X.-D. Yang, P. P. Irani, and M. Beaudouin-Lafon. High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 831–840, 2013.

[79] M. Nancel, E. Pietriga, and M. Beaudouin-Lafon. Precision pointing for ultra-high-resolution wall displays. 2011.

[80] E. Normand and M. J. McGuffin. Enlarging a smartphone with ar to create a handheld vesad (virtually extended screen-aligned display). In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 123–133. IEEE, 2018.

[81] E. Ofek, J. Grubert, M. Pahud, M. Phillips, and P. O. Kristensson. Towards a practical virtual office for mobile knowledge workers. *Microsoft New Future of Work Symposium*, 2020.

[82] J.-y. Oh and H. Hua. User evaluations on form factors of tangible magic lenses. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 23–32. IEEE, 2006.

[83] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pp. 1716–1725. ACM, 2016.

[84] J. Petford, M. A. Nacenta, and C. Gutwin. Pointing all around you: selection performance of mouse and ray-cast pointing in full-coverage displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 533. ACM, 2018.

[85] K. Pfeuffer, J. Alexander, M. K. Chong, Y. Zhang, and H. Gellersen. Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 373–383, 2015.

[86] K. Pfeuffer and H. Gellersen. Gaze and touch interaction on tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 301–311, 2016.

[87] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 99–108. ACM, 2017.

[88] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *International conference on ubiquitous computing*, pp. 315–331. Springer, 2001.

[89] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman. Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE transactions on visualization and computer graphics*, 19(5):886–896, 2012.

[90] D. Raja, D. Bowman, J. Lucas, and C. North. Exploring the benefits of immersion in abstract information visualization. In *Proc. Immersive Projection Technology Workshop*, pp. 61–69, 2004.

[91] D. Ramanujan, C. Piya, K. Ramani, et al. Mobisweep: Exploring spatial design ideation using a smartphone as a hand-held reference plane. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 12–20. ACM, 2016.

[92] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 179–188. ACM, 1998.

[93] P. Reipschläger and R. Dachselt. Designar: Immersive 3d-modeling combining augmented reality with interactive displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pp. 29–41, 2019.

[94] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 378–385. ACM, 1999.

[95] J. Ren, Y. Weng, C. Zhou, C. Yu, and Y. Shi. Understanding window management interactions in ar headset+ smartphone interface. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–8, 2020.

[96] A. Ruvimova, J. Kim, T. Fritz, M. Hancock, and D. C. Shepherd. "transport me away": Fostering flow in open offices through virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020.

[97] K. Ryu, J.-J. Lee, and J.-M. Park. Gg interaction: a gaze–grasp pose interaction for 3d virtual object selection. *Journal on Multimodal User Interfaces*, pp. 1–11, 2019.

[98] D. Schmalstieg, L. M. Encarnação, and Z. Szalavári. Using transparent props for interaction with the virtual table. *SI3D*, 99:147–153, 1999.

[99] D. Schneider, A. Otte, T. Gesslein, P. Gagel, B. Kuth, M. S. Damlakhi, O. Dietz, E. Ofek, M. Pahud, P. O. Kristensson, et al. Reconviguration: Reconfiguring physical keyboards in virtual reality. *IEEE transactions on visualization and computer graphics*, 2019.

[100] D. Schneider, A. Otte, A. S. Kublin, P. O. Kristensson, E. Ofek, M. Pahud, A. Martschenko, and J. Grubert. Accuracy of commodity finger tracking systems for virtual reality head-mounted displays. In *IEEE VR 2020*. IEEE, IEEE, March 2020.

[101] R. Schweigert, V. Schwind, and S. Mayer. Eyepointing: A gaze-based selection technique. In *Proceedings of Mensch und Computer 2019*, pp. 719–723. ACM, 2019.

[102] M. Serrano, B. Ens, X.-D. Yang, and P. Irani. Gluey: Developing a head-worn display interface to unify the interaction experience in distributed display environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 161–171, 2015.

[103] L. Sidenmark and H. Gellersen. Eye&head: Synergetic eye and head movement for gaze pointing and selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 1161–1174, 2019.

[104] L. Sidenmark, D. Mardanbegi, A. Ramirez Gomez, C. Clarke, and H. Gellersen. Bimodalgaze: Seamlessly refined pointing with gaze and filtered gestural head movement. In *Proceedings of Eye Tracking Research and Applications*, 2020.

[105] R. L. Sollenberger and P. Milgram. Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors*, 35(3):483–499, 1993.

[106] P. Song, W. B. Goh, C.-W. Fu, Q. Meng, and P.-A. Heng. Wysiwyf: exploring and annotating volume data with a tangible handheld device. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1333–1342. ACM, 2011.

[107] M. Spindler and R. Dachselt. Paperlens: advanced magic lens interaction above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, p. 7. ACM, 2009.

[108] M. Spindler, C. Tominski, H. Schumann, and R. Dachselt. Tangible views for information visualization. In *ACM International Conference on Interactive Tabletops and Surfaces*, pp. 157–166. ACM, 2010.

[109] J. Steimle, A. Jordt, and P. Maes. Flexpad: highly flexible bending interactions for projected handheld displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 237–246. ACM, 2013.

[110] A. Strauss and J. Corbin. *Basics of qualitative research*. Sage publications, 1990.

[111] H. B. Surale, A. Gupta, M. Hancock, and D. Vogel. Tabletinvr: Exploring the design space for using a multi-touch tablet in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 13. ACM, 2019.

[112] Z. Szalavári and M. Gervautz. The personal interaction panel–a two-handed interface for augmented reality. In *Computer graphics forum*, vol. 16, pp. C335–C346. Wiley Online Library, 1997.

[113] S. K. Tadeja, T. Kipouros, and P. O. Kristensson. Exploring parallel coordinates plots in virtual reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, p. LBW2617. ACM, 2019.

[114] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann. A survey on interactive lenses in visualization. In *EuroVis (STARs)*. Citeseer, 2014.

[115] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann. Interactive lenses for visualization: An extended survey. In *Computer Graphics Forum*, vol. 36, pp. 173–200. Wiley Online Library, 2017.

[116] J. Turner, J. Alexander, A. Bulling, and H. Gellersen. Gaze + rst: integrating gaze and multitouch for remote rotate-scale-translate tasks. In *Proceedings of 2015 CHI conference on Human Factors in Computing Systems*, pp. 4179–4188, 2015.

[117] B. Ullmer and H. Ishii. The metadesk: models and prototypes for tangible user interfaces. In *Proceedings of Symposium on User Interface Software and Technology (UIST??? 97), ACM*, 1997.

[118] R. Unger and C. Chandler. *A Project Guide to UX Design: For user experience designers in the field or in the making*. New Riders, 2012.

[119] J. Wagner, M. Nancel, S. G. Gustafson, S. Huot, and W. E. Mackay. Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1299–1308. ACM, 2013.

[120] J. A. Wagner Filho, C. M. D. S. Freitas, and L. Nedel. Virtualdesk: a comfortable and efficient immersive information visualization approach. In *Computer Graphics Forum*, vol. 37, pp. 415–426. Wiley Online Library, 2018.

[121] M. Waldner, E. Kruijff, and D. Schmalstieg. Bridging gaps with pointer warping in multi-display environments. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pp. 813–816, 2010.

[122] P. D. Wellner. Interacting with paper on the digitaldesk. Technical report, University of Cambridge, Computer Laboratory, 1994.

[123] D. Wenig, J. Schöning, A. Olwal, M. Oben, and R. Malaka. Watchthru: Expanding smartwatch displays with mid-air visuals and wrist-worn augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 716–721, 2017.

[124] J. R. Williamson, M. McGill, and K. Outram. Planevr: social acceptability of virtual reality for aeroplane passengers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[125] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146, 2011.

[126] D. Wolf, J. J. Dudley, and P. O. Kristensson. Performance envelopes of in-air direct and smartwatch indirect control for head-mounted augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 347–354. IEEE, 2018.

[127] R. Xiao, M. A. Nacenta, R. L. Mandryk, A. Cockburn, and C. Gutwin. Ubiquitous cursor: a comparison of direct and indirect pointing feedback in multi-display environments. In *Proceedings of Graphics Interface 2011*, pp. 135–142. 2011.

[128] C. Yoo, I. Hwang, E. Rozner, Y. Gu, and R. F. Dickerson. Symmetrisense: Enabling near-surface interactivity on glossy surfaces using a single commodity smartphone. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 5126–5137. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858286

[129] D. Yu, H.-N. Liang, X. Lu, T. Zhang, and W. Xu. Depthmove: Leveraging head motions in the depth dimension to interact with virtual reality head-worn displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 103–114. IEEE, 2019.

[130] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 246–253, 1999.

[131] F. Zhu and T. Grossman. Bishare: Exploring bidirectional interactions between smartphones and head-mounted augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020.

[132] D. Zielasko, M. Krüger, B. Weyers, and T. W. Kuhlen. Menus on the desk? system control in deskvr. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1287–1288. IEEE, 2019.

[133] D. Zielasko, M. Krüger, B. Weyers, and T. W. Kuhlen. Passive haptic menus for desk-based and hmd-projected virtual reality. In *2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–6. IEEE, 2019.