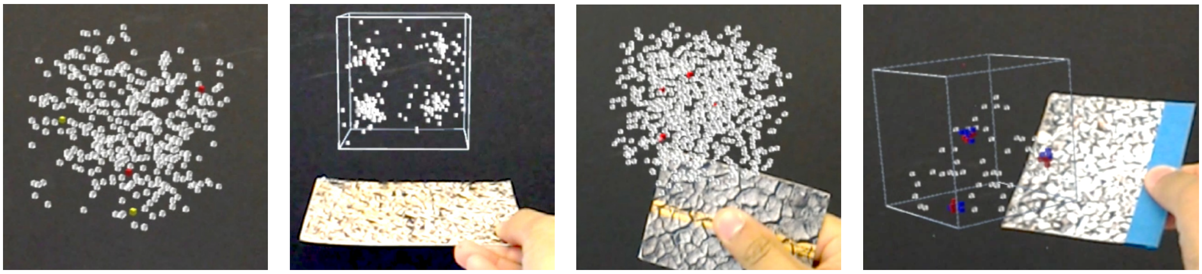
The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality?

Benjamin Bach, Ronell Sicat, Johanna Beyer, Maxime Cordeil, Hanspeter Pfister



(a) Distance task (b) Cluster task (c) Selection task (d) Cutting plane task

Fig. 1. Monoscopic and low-resolution approximations of hologram visualizations of 3D scatterplots using immersive tangible augmented reality with the HoloLens. Actual perception through the HoloLens provides stereoscopic images and higher resolution.

Abstract—We report on a controlled user study comparing three visualization environments for common 3D exploration. Our environments differ in how they exploit natural human perception and interaction capabilities. We compare an augmented-reality head-mounted display (Microsoft HoloLens), a handheld tablet, and a desktop setup. The novel head-mounted HoloLens display projects stereoscopic images of virtual content into a user's real world and allows for interaction in-situ at the spatial position of the 3D hologram. The tablet is able to interact with 3D content through touch, spatial positioning, and tangible markers, however, 3D content is still presented on a 2D surface. Our hypothesis is that *visualization environments that match human perceptual and interaction capabilities better to the task at hand improve understanding of 3D visualizations*. To better understand the space of display and interaction modalities in visualization environments, we first propose a classification based on three dimensions: perception, interaction, and the spatial and cognitive proximity of the two. Each technique in our study is located at a different position along these three dimensions. We asked 15 participants to perform four tasks, each task having different levels of difficulty for both spatial perception and degrees of freedom for interaction. Our results show that each of the tested environments is more effective for certain tasks, but that generally the desktop environment is still fastest and most precise in almost all cases.

Index Terms—Augmented Reality, 3D Interaction, User Study, Immersive Displays

1 Introduction

Driven by new display and interaction technologies, information visualization is rapidly expanding beyond applications for traditional desktop environments. Technologies such as virtual and augmented reality, tangible interfaces, and immersive displays offer more natural ways in which people perceive and interact with data by leveraging their capabilities for perception and interaction with the real world. For example, tangible interfaces provide higher degrees-of-freedom (DOF) in interaction, stereoscopic displays can provide a sense of depth, and augmented reality can connect virtual content to real-world objects and create strong affordances for interaction. This raises questions with respect to the benefit of natural interfaces for understanding and interactive exploration of data visualizations (e.g., [33, 35, 56]).

The traditional desktop environment, composed of 2D screens, mouse, and keyboard, is often criticized as being less effective for tasks concerned with visualization of 3-dimensional (3D) content [50]. However, 3D visualizations of both spatial and abstract data can be desired where two-dimensional projections and representations fall short (e.g., multi-dimensional scaling). Consequently, research in augmented and virtual reality (AR/VR) and HCI has contributed a variety of studies and techniques targeting 3D visualization and interaction. These studies have reported many insights in the respective conditions, suggesting general benefits of novel technologies (e.g., [10, 34, 63]). However, for visualization the question remains *how efficient is direct tangible interaction with virtual 3D holograms in the real world?* as well as *how effective is current technology for such?*

In this paper, we focus on visualization environments composed of immersive and stereoscopic augmented-reality combined with tangible input through fiducial paper marker tracking, called *tangible AR* [12]. Immersive head-mounted AR displays, such as Meta [3] or the Microsoft HoloLens [2], project stable stereoscopic projections (holograms) that can be placed at deliberate positions in the user's natural environment. In addition to allowing people to directly interact with the visualization in the same space where the holographic presentation is perceived, people can freely walk around the hologram and can even "park" holographic visualizations in their environment for later use. We believe this scenario offers a wide range of novel applications and designs with the goal of improving humans' understanding of data.

As devices for immersive and tangible AR have reached an ever higher level of maturity, we expect that the number of visualization applications for these devices will increase in the future. Therefore, the

- *Benjamin Bach, Ronell Sicat, Johanna Beyer, and Hanspeter Pfister are with Harvard University. E-mails: {bbach,sicat,jbeyer,pfister}@seas.harvard.edu.*
- *Maxime Cordeil is with Monash University. E-mail: max.cordeil@monash.edu.*

Manuscript received 31 Mar. 2017; accepted 1 Aug. 2017.
Date of publication 28 Aug. 2017; date of current version 1 Oct. 2017.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2017.2745941

purpose of our study is to provide researchers and practitioners with some initial evidence about how visualization environments for immersive tangible AR (HoloLens with tangible markers, *ImmersiveAR*) and traditional AR on a handheld tablet (*TabletAR*) compare to the traditional desktop environment (*Desktop*).

To that end, we focus on three of the most prominent aspects of visualization environments and are interested in their *combined* effects: (i) stereoscopic perception, (ii) degrees of freedom for interaction, and (iii) proximity of the respective physical spaces for perception and interaction [49]. In each environment, we investigate four representative visualization and interaction tasks that vary in the degree to which they rely on perception and interaction: (1) estimate spatial distance between points, (2) count clusters, (3) select visual elements, and (4) orient a cutting plane (Fig. 1). As for the visualization, we chose to study 3D point clouds as they are similar to a variety of 3D visualizations (e.g., 3D scatterplots, space-time cubes), their respective visual patterns (e.g., clusters, outliers, trends, density), as well as common visualization challenges in 3D (e.g., occlusion, perspective distortion).

Our results show good performance for immersive tangible AR for tasks that can be solved through spatial perception and interactions with a high degree of freedom. We also observed a slight improvement in performance due to training for the *ImmersiveAR* environment over several days. However, overall the desktop environment offered superior precision, speed, and familiarity across most tasks. One possible interpretation of these results is to strive for a tighter integration of different visualization environments in order to support a wider range of tasks and interactions. While technical performance and precision of immersive technologies, such as the HoloLens, will likely improve over the next years, our results point to general trends and current drawbacks and serve as a timely reference point.

2 Related Work

3D visualizations have been found useful for inherently *spatial* data in many applications in biomedicine, science, and engineering. Using 3D visualizations for displaying *abstract* data has historically been a controversial topic [34,50] but some exploration tasks for high-dimensional data have been found to increase cognitive effort if only sets of 2D projections were provided [53,61]. Overall, the landscape of scientific and abstract 3D visualizations is very rich [13] including 3D scatterplots [1,22,40], 3D multi-dimensional scalings (MDS) [39,40], and space-time cubes [8,28]

Perception of 3D visualizations

The effectiveness of 3D visualizations has been extensively evaluated on different display technologies such as 2D (monoscopic) displays [48,59], stereo-displays [5,62], stereo-displays with head-tracking [16], data physicalizations [34], and immersive virtual reality environments [23,63]. Two surveys [41,42] of studies from different domains conclude that 3D stereoscopic displays increase task performance by 60% on average. For example, understanding relative positions was found to be better supported on 2D screens, while shape understanding is supported better by respective 3D projections (on 2D screens). Also, 3D visualizations have been found most useful for classification tasks, and in cases where manipulation (interaction) was required. On the other hand, 25% of the reviewed studies found no benefit of 3D stereoscopic displays over 2D projections and suggest that *kinetic depth* (i.e., a 3D impression through motion-parallax) is more important for stereoscopic perception [60] than actual stereovision. That means that movement, e.g., through rotation of a 3D visualization on a screen, is enough to improve the perception of a 3D model.

Interaction with 3D Visualizations

Interaction is required in cases where visualizations become dense and for tasks requiring a lot of exploration. Interactive exploration for 3D visualization can include camera rotation, visual-access lenses [20], the placement of cutting planes [15,24], as well as selection [65] and brushing [54]. Due to its higher spatial dimensionality, interaction with 3D content may require higher degrees of freedom (DOF) for view and visualization manipulation (i.e., along the three spatial dimensions and three spatial angles).

Mid-air gesture interaction [17] and tangible user interfaces (TUIs) [26,30] are examples that both provide higher DOFs for interaction. Furthermore, through sensing the position of one's limbs through interaction (*proprioception*) these interfaces can provide information about space, positions, and distances [47]. In addition to gesture interactions [17], TUIs employ physical objects and allow a 1-to-1 mapping of physical interaction devices and virtual objects.

TUIs also allow control of multiple DOF simultaneously [66] and have been found to be more effective for interaction with 3D content compared to touch interaction on a tablet or a mouse [11,43]. For example, Hinckley explored tangible cutting planes that a user could move freely in space and whose rotation and position would be propagated to the visualization system on a screen [29]. Jackson presented a tangible interface for querying vectors in 3D vector fields [32] and Cordeil et al. [18] list further examples of tangible interfaces for 3D visualization. These examples include sophisticated technical devices such as dynamic tangible and physical barcharts [57], a Cubic Mouse for pointing tasks in 3D space [27], and Paper Lenses for exploring volume data through a spatially tracked sheet of cardboard and subsequent projection of virtual content onto the cardboard's surface [52].

General drawbacks of TUIs are fatigue and the need for extra physical objects [11], as well as the possible lack of coordination [66]. Moreover, TUIs coupled with 2D screens do not mean an automatic improvement in task efficiency. Besançon et al. [11], evaluated the efficiency for three interaction modalities on monoscopic screens: mouse, touch, and a tangible device for 3D docking tasks (rotation, translation, scaling of objects) and found that precise and well-known interaction with a mouse is outperforming TUIs. However, their study did not include any 3D visualization-specific exploration tasks.

One general problem of using TUIs in the context of 3D visualization may be the relative spatial distance between perceived interaction (i.e., fingers on the device) and the perceived output (i.e., visualization on the screen); in other words, the distance between the *interaction space* and the *perception space* may be too large [25,49].

Augmented Reality for Interactive Visualization

Augmented reality means the blending of virtual content into the real world [46] and has been used to couple tangible interfaces with virtual projections. *Tangible AR* [12] combines AR displays with tangible input modalities, most commonly based on vision-based fiducial marker tracking [67]. Fiducial markers are visual patterns, typically printed on paper, whose 3D position and orientation with respect to a camera are easily detected and tracked via vision-based techniques (e.g., [4]).

For displaying data in augmented reality, 3D scatterplots [21,44] and 3D graphs [51] have been implemented using fiducial markers for visualization placement and pointing. Tangible markers have also been used to simulate specialized tools, e.g., a virtual cutting plane that allows neurologists to explore 3D representations of the brain [58] and have been found faster than mouse and touch interaction [31]. While allowing for a high DOF and technically direct interaction with virtual content, in all these cases the virtual content is shown on the tablet screen while the interaction happens "behind" the screen, requiring cognitive mapping between interaction and perception [49].

Immersive Environments for Interactive Visualization

Immersion, such as through virtual reality, eventually is able to close the gap between perception and interaction space. While the sole effect of immersion, i.e., being surrounded by virtual content through a large field of view, has been both found useful [9,23] and questioned [43], environments that immerse the user into a virtual world are able to fully integrate action and perception space. Immersive environments have been used extensively to visualize 3D content, e.g., for 3D network visualization [19]. Interaction in virtual reality is often difficult, as real world objects (e.g., the users' hands) either need to be shown as video-overlays or re-modeled as completely virtual content [23,45,64]. AR, on the other hand, does not suffer from the missing visual feedback.

Headmounted displays for AR, such as the Microsoft HoloLens [2] or Meta [3], combine the best of both worlds: immersive stereoscopy as in VR and access to the real world, including desktop computers,

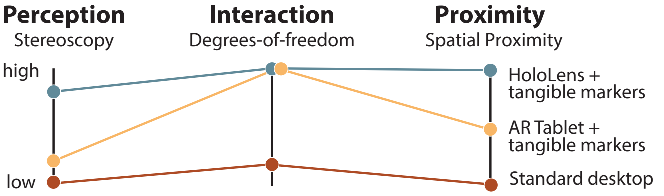


Fig. 2. Classification of the three visualization environments used in our study for perception, interaction, and proximity.

mobile devices, pen and paper, collaborators, large displays, and data physicalizations. The concept has been described as *immersive AR* [55] or—if used with tangible markers—*tangible AR* [12]. Belcher et al. [10] reported that stereoscopic AR is in fact more accurate but slower than a monoscopic display condition. Other than rotation, no interaction capabilities were tested in that study.

To the best of our knowledge, no study has directly investigated the combined effects of *immersive tangible AR*, i.e., immersive augmented reality displays coupled with tangible markers, for interactive 3D visualization. Many conditions have been tested individually in order to isolate the respective effects (e.g., mono vs. stereo displays, mouse vs. TUI). However, each combination of display and interaction technique creates a unique *visualization environment* and specific combinations of factors may perform better than others, independent of the respective factors in isolation. Our evaluation study aims to address this gap.

3 Study Rationale

We are interested in the benefits of an immersed tangible experience with 3D visualizations which promise to better match how humans perceive and interact with the real world. To that end we compare the HoloLens with tangible input (immersive tangible AR) to other visualization environments. While there are many conceptual and technical differences across visualization environments (e.g., resolution, field-of-view/screen size, physicality), we focus on the following three main aspects: *3D perception* (perception), *high-degrees-of-freedom for tangible interaction* (interaction), and the *spatial proximity* of perception and interaction (embodiment). These aspects represent characteristics that we consider to have a major influence on task performance for 3D visualization. Figure 2 shows each aspect as a dimension ranging from low to high and locates the respective visualization environments that we tested: traditional desktop, and tangible AR with HoloLens and tablet.

Stereoscopy (perception): The HoloLens enables a stable perception of a stereoscopic image. Users can freely move the hologram or move themselves around the hologram. Hence the ability to perceive 3D content is higher than with a desktop environment or tablet which provide only a flat screen without a stereoscopic image.

Degrees-of-freedom (interaction): The HoloLens allows the tracking of position and orientation of tangible fiducial markers, similar to AR on a tablet or mobile phone. This allows markers to become tangible tools and enables a high degree of freedom for interaction compared to a desktop environment equipped with mouse and keyboard.

Proximity (embodiment): Proximity means the extent to which interaction movements and interaction effects (rotation, movement, selection, etc.) are colocated and coordinated visually and in space [18]. The HoloLens allows users to "touch" the data by reaching with their hand inside the hologram. Given the right tracking technology, this allows direct manipulation of the hologram. In order to perform the same manipulations in a desktop environment, the user is constrained to mouse movement in two directions, while coordinating the movement of the mouse cursor on the screen with the movement of the hand approximately half a meter away from the screen.

Low and high values along each of these dimensions are highly approximative and do not imply lower or higher human task performance. However, the placement of an environment helps us formulate hypotheses about its suitability for a specific 3D visualization task and its expected performance. Moreover, we can describe additional visualization environments or their variations for a structured evaluation.

Factor	*Desktop*	*TabletAR*	*ImmersiveAR*
Perception			
Stereoscopic	×	×	✓
Screen Size	10.8"	10.8"	approx. 11"
Resolution	1268×720	1268×720	1268×720
Immersiveness	low	medium	high
Interaction			
Body movement	×	✓	✓
Vis. movement	✓	✓	✓
Tangibility	×	✓	✓
DOF	2	2+6	6
Proximity			
Interaction Space	2D	2D+3D	3D
Perception Space	2D	2D	3D
Spatial Proximity	far	medium	identity
Subjective			
Familiarity	high	medium	low
Physical Effort	low	medium	high

Table 1. Environment configurations as used in the study.

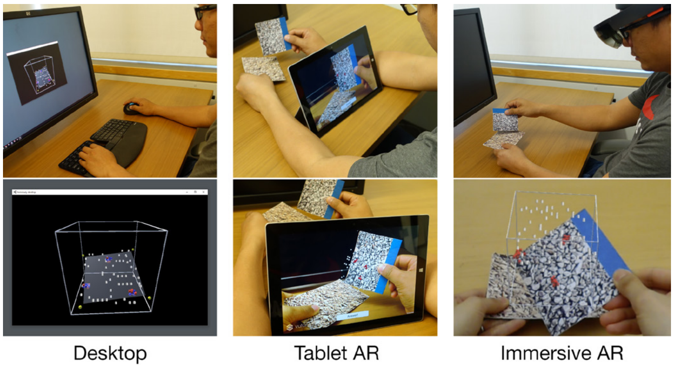


Fig. 3. Study setup (top row) and approximate user perspectives (bottom row) in each environment.

For example, we could have chosen to use a 3D monitor in the desktop environment, improving the user's ability to perceive 3D content. Alternatively, we could have used tangible interaction on a 2D screen.

In order to keep conditions low for a study, we decided to include augmented reality on a tablet with tangible markers as a third common visualization environment besides the HoloLens and the traditional desktop. A user perceives the visualization on the tablet while the visualization is placed onto a fiducial marker that is filmed by the tablet's camera. The tablet provides monoscopic perception and the same high DOF as in our HoloLens environment when manipulating the marker. In addition, touch-screen input allows one to select elements on the screen. Spatial proximity for the tablet environment is higher than for the desktop as people perceive their interaction in the perception space (the screen). A detailed description of the exact study conditions follows in the next section.

4 Study Design

We now explain the technical details of our three chosen visualization environments and describe tasks, hypotheses, and procedures of our controlled user study.

4.1 Environments

Table 1 summarizes the characteristics and parameters of the three visualization environments in our study.

Desktop, keyboard, and mouse (*Desktop*): Our desktop environment (Fig. 3, left) consisted of a standard 22-inch monitor with a maximal resolution of 2560×1600. We adjusted the size of the actual visualization (10.8") to match it across environments. Participants used a standard mouse and required only the left mouse button for interaction. The display showed a perspective projection of the visualization that could be rotated by dragging the mouse. We describe task-dependent mouse interactions in Sect. 4.3. Participants sat at a normal desk under the same lighting conditions as in the other conditions.

Tablet and markers (*TabletAR*): This environment (Fig. 3, center) featured a handheld tablet (Microsoft Surface 3) with a maximal resolution of 1920×1280 and a 10.8" 2D display. The tablet display showed the real-time video stream of the 8-Megapixel rear camera, filming a 9cm × 9cm cardboard with fiducial markers, used to render the visualization. Moving and rotating the cardboard propagated these interactions to the visualization. Markers were tracked using the Vuforia toolkit [4] and their recommended marker patterns. For other task-dependent interactions, e.g., element selection or cutting plane position, participants used special markers glued onto cardboard and representing tangible interaction tools (Sect. 4.3). In this setup, the visualization, user's hands, and tools, appeared to be *behind* the tablet. The built-in stand of the tablet allowed for bi-manual manipulation of the interaction tools.

Participants were seated at a table but were free to stand up and move if they preferred. We decided against requiring participants to strictly stay seated during the study and instead recorded the preferred user strategies (sitting, standing, moving around, etc.) for each task.

HoloLens and markers (*ImmersiveAR*): This environment (Fig. 3, right) consisted of Microsoft's HoloLens for binocular display of the visualization and the same tangible marker tools as in the *TabletAR* environment. For triggering and menu interaction, participants used the HoloLens clicker, which they held comfortably in their dominant hand together with the interaction tool. The clicker has a flexible strap so that it can be worn on any finger, allowing participants to grip and hold other tools. The HoloLens shows the content on two high-definition (1268×720) see-through displays and weighs about 579 grams, according to Microsoft's website. It is equipped with inertial measurement units, tracking cameras, stereo microphones, and one HD color camera on the front that we used for marker tracking. The HoloLens continuously tracks its environment and updating the environmental mesh. This helps in keeping holograms extremely stable in place and allowed participants to walk around the holograms. Similar to the *TabletAR* condition, users could sit at a table or stand up and lock the hologram anywhere in mid-air in the room.

Across all environments we controlled for the actual size and resolution of the visualization. On the desktop the application window was set to have the same size (10.8 inches) as the tablet and all environments showed the visualization in the same resolution (1268×720). The field-of-view of the HoloLens appears relatively small but results in approximately the same diagonal size as the tablet at the distance of a comfortable arm length from the hologram. Its display resolution was the same as the *TabletAR* and *Desktop*.

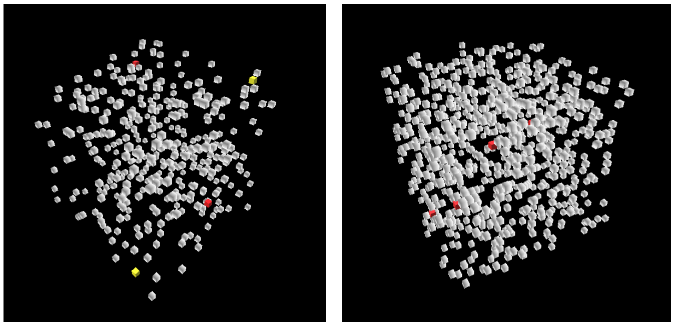
Marker images (e.g., see Fig 3 *middle* and *right*) were recommended by the Vuforia toolkit and taken from Vuforia's website. We tried different and less complex images but tracking performance was significantly reduced. Unfortunately, the 2D figures in this paper do not represent the proper conditions as seen in stereovision, in which we did not find the marker images to reduce perception of the hologram.

4.2 Measures

For all trials we recorded task completion time (*time*) and error (*error*). The timer started when the visualization had finished loading for each trial and stopped as soon as the participant hit one of the trigger keys (space-bar (*Desktop*), answer-button (*TabletAR*), or clicker (*ImmersiveAR*)). Participants were instructed to press the trigger key as soon as they knew the answer, stopping the task timer. After selecting the answer from a menu, the menu disappeared, then the next visualization appeared and the timer restarted.

4.3 Tasks and Data

We selected a set of four tasks representative of the exploration of 3D visualizations and balancing how much stereoscopic perception and direct interaction was required. To keep the number of conditions and the effort for learning low, we decided on a single and representative visualization technique: point clouds (Fig. 4). Point clouds represent a variety of 3D visualizations including 3D-scatterplots, specific space-time cubes, as well as biomedical images and even flow fields. Point clouds can contain individual points of interest, points of different types and sizes, areas of different densities, clusters, outliers, and can vary in their general density. Points in the scatterplot were rendered as equally-sized light-gray shaded cubes (Fig. 4). We found shaded cubes to be easier to perceive with depth and perspective distortion than, e.g., spheres. The dimensions of the visualizations were fixed across all environments and tasks to approximately 10×10×10 *cm*.



(a) Distance task (b) Selection task

Fig. 4. Example stimuli for two tasks. (a) *distance*: participants had to estimate which pair of colored points (red or yellow) had the smaller spatial distance. (b) *selection*: participants had to select the red points.

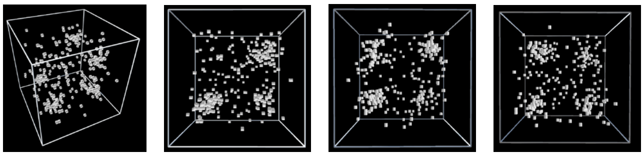
For each of our tasks, 9 data sets were created prior to the study (3 training + 6 recorded trials) and each participant was presented with all of the data sets in randomized order. Task order was kept fixed, ranging from more simple and perception-focused tasks to more complex interaction-focused tasks. In the following, tasks are described in the order they appeared in the study and were explained to the participants with examples on paper before each task condition.

Point Distance (*distance*): *Which of the point pairs are closer to each other: the red pair or the yellow pair?* The visualization showed randomly distributed points, colored light-gray, except for two pairs: a first pair was colored red, and the other one was colored yellow (Fig. 4(a)). The point cloud was dense yet sparse enough to prevent any interactions except changing the viewing direction being required. Participants had to rotate the visualization by dragging the mouse (*Desktop*), by rotating a tangible marker or by walking around the visualization (*TabletAR*, *ImmersiveAR*) The answer menu presented the user with two choices: *red* or *yellow*.

This distance task is representative for a variety of tasks related to visualization in 3D space. The proper perception of spacial relations is essential to the effectiveness of any 3D visualization, including the spotting of outliers and clusters. A variation of this task, requiring distance estimation in 3D space has been studied in cave VR environments [23] and for 2D displays [48].

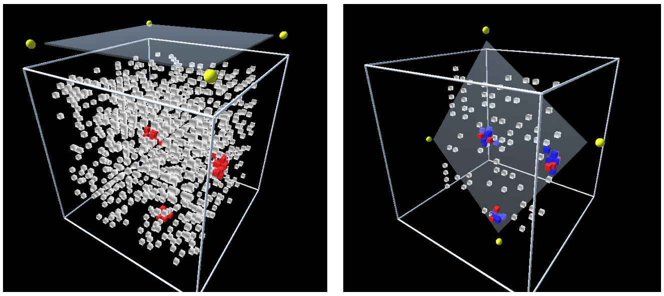
The data for this task consisted of randomly generated points on a regular grid of 30×30×30 possible discrete positions (the size of the visualization remained 10×10×10 cm). We used a point density of 1.5% (135 points). Out of these points, two pairs of points were randomly selected such that the spatial (Euclidean) distance between the first pair of points was 20% shorter than the distance between the second pair. The color assignment (red or yellow) of the two pairs was randomized. In a pilot study, we tried different distance differentials, including 10% and 50%. However, we found 10% difference caused too much effort for participants and was too error-prone, while 50% distance was too easy. Error for this task was binary and indicated if participants had found the correct pair.

Cluster Estimation (*clusters*): *What is the minimum number of clusters that you can count when looking from all three orthogonal viewing directions?* The visualization contained a set of gray points that formed sets of 3 to 7 point clusters, plus random noise (Fig. 5(a)). Participants saw data sets with 3, 4, 5, 6, and 7 clusters. Clusters were positioned in a way such that when the visualization was viewed



(a) 5 clusters (b) 3 clusters (c) 4 clusters (d) 4 clusters

Fig. 5. Example for *cluster* task: (a) perspective projection showing all 5 clusters. (b-d) seen from different sides where some clusters overlap. Participants had to report the lowest number of clusters observed.



(a) Initial state (b) Goal

Fig. 6. Example for *cuttingplane* task where participants had to intersect the three red clusters with the cutting plane. Yellow points served as handles for rotating the cutting plane.

from different orthogonal directions, different numbers of clusters overlapped and were perceived as a single cluster (Fig. 5). Participants had to view the data from three orthogonal directions, e.g., by rotating the visualization (via mouse drags or marker rotation) or physically moving around it. We visualized the wire-frame of the data bounding box to provide cues for the orthogonal viewing directions. The answer menu asked the user to select a number between 3 and 7. This task is representative of other tasks that require the inspection of projections in a 3D visualization. Information gleaned from these projections may include trends, outliers, and clusters.

Our data consisted of point clusters and points used as background noise. The noise points were randomly placed within a regular 30×30×30 grid (the size of the visualization remained 10×10×10 cm) with 0.25% density (approx. 67 points). Cluster count and cluster centroid positions were manually set to be evenly distributed and to provide the respective overlap-conditions required for this task. Placement of the 30 points per cluster were based on a random Gaussian distribution with a standard deviation of 2. We performed pilot tests with smaller clusters and different noise parameters before arriving at these parameters. Error for this task was binary and indicated if participants had found the correct answer.

Point Selection (*selection*): *Select all red points as fast as you can. Selected points turn blue.* The visualization showed randomly distributed points and 4 red target points (Fig. 4(b)). The density of the points was selected such that in some cases the red points were hidden, thus requiring participants to rotate the visualization in order to see them. For the *Desktop*, selection required point and click interaction, while for the *TabletAR*, selection required 2D touch interaction. We used the closest selected cube under the cursor and finger as selected item, respectively (2 DOF). For *ImmersiveAR*, selection required moving a marker in 3D space and placing a 3D pointer (3 DOF). The upper left corner of the selection marker served as a cursor and was marked with a small purple sphere (Fig. 1(c)). The clicker served as trigger.

Data consisted of random points in a 20×20×20 regular grid (the size of the visualization remained 10×10×10 cm) with 10% point density. The 4 target points were randomly selected from this set of points. In a pilot study we tried several point density settings, ranging from very sparse (1%) to very dense (50%) and found that 10% strikes a good balance between the points being too sparse so that no rotation is needed, and too dense so that it is impossible to find the red points in some cases. Error for this task is the number of clicks that did not select a target point.

Cutting Plane (*cuttingplane*): *Place the cutting plane so that it touches all three clusters of red points. Points in the cluster touched by the cutting plane turn blue.* The visualization showed noise points and 3 clusters of red points randomly positioned inside the visualization space. The wire frame of the data bounding box was provided for additional spatial cues. For the *Desktop*, the cutting plane was shown as a semi-transparent plane (Fig. 6) controlled by mouse interaction. After trying different interactions we decided on the following 3-DOF approach: dragging the mouse on the plane surface translates the plane along its normal; dragging the mouse on a plane corner (shown as yellow points) rotates the plane with respect to the axis defined by the two-neighboring corners. Participants issued the trigger once they were satisfied with their cutting plane placement. For *TabletAR* and *ImmersiveAR*, a tangible cutting-plane marker could be manipulated to directly place the cutting plane in 3D space with 6-DOF (Fig. 1(d)). Placing a cutting plane is common in many 3D visualizations (e.g. [29, 52]). It is especially useful for the exploration of very dense data, such as volume visualizations of medical scans or fluid flow.

The data consisted of randomly sampled points from a regular 15 × 15 × 15 grid (the size of the visualization remained 10×10×10 cm) with a density of 30%. The cluster centroid positions were randomly generated such that each pair was 10 units away from each other. Each cluster had 10 points whose positions are defined by a random Gaussian distribution with standard deviation of 0.5 units. Accuracy was calculated as sum of the distance of each cluster's geometric center to the cutting plane placed by the participants.

4.4 Hypotheses

Null-hypotheses for each task are that there will be *no differences* in time and error between all three environments. We developed our hypotheses based on our literature survey presented in Sect. 2 and our analysis of the three dimensions of our environments described in Sect. 3.

- $H_{distance-error}$: *For distance, we expect ImmersiveAR to be more precise than the other environments.* Stereoscopic vision may give a better impression of spatial distances between the points and does not require rotation to provoke kinetic depth [60].
- $H_{distance-time}$: *For distance, we expect Desktop to be faster than the two other environments (TabletAR, ImmersiveAR).* Even though participants may obtain a better first impression about point distance in *ImmersiveAR*, we expect participants to want to validate their answer by rotating the visualization or moving their head. We believe that rotation will be slower in *TabletAR* and *ImmersiveAR* as it requires physically moving the tangible markers or one's body (*ImmersiveAR*). Visual delay in correcting for head and marker movement is further expected to slow down *ImmersiveAR*.
- $H_{cluster}$: *For cluster, we expect Desktop to be faster than the two other environments (TabletAR, ImmersiveAR).* The *cluster* task requires both precise rotation and perception. We expect that mouse rotation on the desktop, due to low physical effort, allows participants to quickly and precisely rotate the visualization into the (three) positions required for this task. Again, we believe physical marker rotation to be slower in both *TabletAR* and *ImmersiveAR*. With respect to precision, we assume *Desktop* to increase precision because it already delivers a 2D-projection of the clusters while *ImmersiveAR* may prevent participants from perceiving proper 2D-projections from each side due to stereoscopic and perspective distortions.
- $H_{selection}$: *For selection, we expect ImmersiveAR to perform fastest.* It allows participants to directly select points in mid-air with 3-DOF without the need for rotation, while 2-DOF selection with a mouse in *Desktop* and touch on *TabletAR* would require frequently rotating the visualization in order to properly expose the target points.
- $H_{cuttingplane}$: *For cuttingplane, we expect ImmersiveAR to be both fastest and most precise.* This is because of the 6-DOF direct manipulation coupled with stereoscopic perception. Both factors can improve participants' perception of where the target plane (the plane spanned by the red clusters) and the current cutting plane are located

in space. We expect *TabletAR* to perform slower and potentially less precisely than *ImmersiveAR* due to the missing stereoscopic perception required for proper eye-hand coordination.

- $H_{training}$: *We expect participants to become both faster and more precise with increased familiarity over several days in the ImmersiveAR environment.* We assume that participants are well trained on the *Desktop* and that participants will get used to the *TabletAR* environment relatively quickly compared to *ImmersiveAR*.

4.5 Participants

We recruited 15 participants from the University's mailing list. 7 participants were undergraduates enrolled in an architecture program or related and were well trained in the usage of 3D CAD software on a traditional desktop. Four students were enrolled in a computer science program and well trained with mouse and keyboard interactions. Eight participants were male, seven were female. Two participants had previous experience with immersive VR technology and another two had previously used the HoloLens for a short time. Because the device is relatively new, our participants were novices with the HoloLens while all of them were well versed using the desktop. We do believe this reflects a typical scenario until wearable AR devices become truly ubiquitous. Yet, we were particularly interested in participants used to 3D visualization as immersive environments would be of special use to such users.

4.6 Procedure

We followed a full-factorial within-subject study design and blocked participants by environment. While environments were balanced using a Latin square (3 groups), task order was fixed to *distance*, *cluster*, *selection*, and *cuttingplane*. We decided on this order to increase perception and interaction complexity with each task. We report performance measures for each task individually.

Each condition (environment $\times$ task) started with 3 non-timed training trials followed by 6 timed study trials. Participants were told to be *as fast as possible*. Tasks were explained by the instructor using text instructions and examples printed on paper. During training, the instructor made sure participants correctly understood the task and could perform the required interactions to solve and finish the task. For each environment (*Desktop*, *TabletAR*, *ImmersiveAR*), the instructor explained the technology and helped participants with setting up.

During each of the 9 trials (including training) we measured task-accuracy and task-completion time from the start of the trial until the trigger event. We tracked positions of the visualization and the camera as well as the relative rotation between them. When participants clicked the trigger button to end a trial, an answering menu was brought up. In *Desktop* and *TabletAR*, the answer menu was shown in the center of the screen. In *ImmersiveAR*, the menu was shown always on the same wall in the study room for all participants, tasks, and trials. Participants were told to first issue the trigger—and hence stop the timer—and then turn to the menu to specify their answer.

Partcipants could take breaks between trials whenever the timer was not running. In *ImmersiveAR*, the instructor reminded participants to take breaks. Breaks could be taken as long as necessary in all conditions. The study was conducted in a quiet and well illuminated room with enough space for participants to freely walk around the hologram if desired. After the study, we asked each participant to fill out a questionnaire, indicating for each environment the participant's comfort and fatigue, the interaction's ease-of-use, as well as how each of the display conditions supported or hindered the tasks.

4.7 Long-term Training Condition

A random subset of 6 (out of 15) of the study participants was invited for a special condition to study the effects of familiarity/training on *ImmersiveAR* performance. Participants came back to the lab for 5 consecutive days to only perform the *ImmersiveAR* condition. Tasks, task order, and task difficulty remained the same. However, we generated new data for each session and for each of the 9 trials using the methods described in Sect. 4.3. On average, participants spent 15-20 minutes

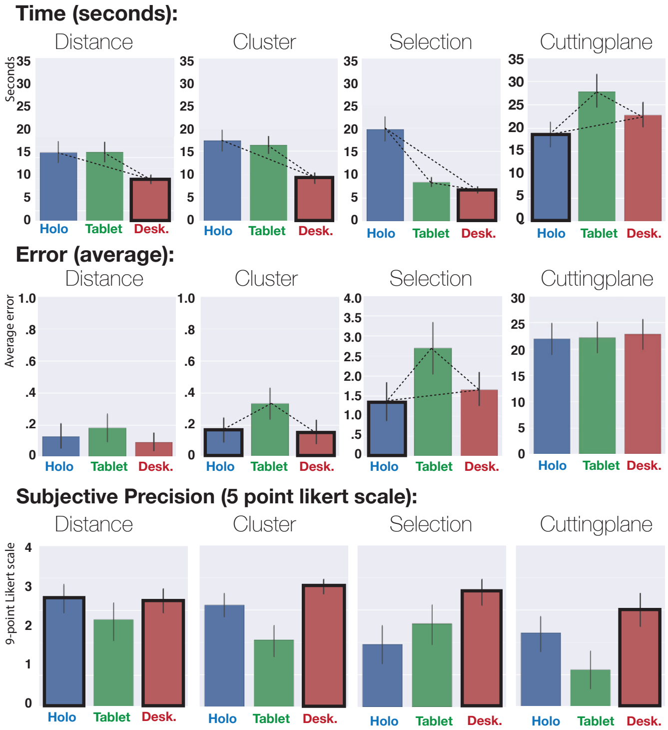


Fig. 7. Results for *time* (seconds), *error*, and subjective reported precision (5-point Likert scale) by task. Confidence intervals indicate 95% confidence for mean values. Dashed lines indicate significances for $p < .05$. Highlighted bars indicate significant best results.

on the *ImmersiveAR* condition each day. We report the results of this *long-term training* group separately in Sect. 5 and Sect. 6.

5 Results

We now report on the results of our user study with respect to time, accuracy, user strategies, and subjective user feedback.

5.1 Task Completion Time and Accuracy

On average, it took each participant 1.5 hours to complete the study on all three environments. For each of the 4 tasks, we obtained 270 recorded trials (15 participants $\times$ 6 trials $\times$ 3 environments), excluding the 3 training trials per condition. We found time and error to not be normally distributed and we were not able to correct to normal distribution through logarithmic or any other transformation. To find outliers, we hence visualized the individual distributions of values for both *time* and *error* for all tasks and environments. Some of these outliers were quite extreme and we decided on above 60 seconds and below 1 second to be good thresholds for removing outliers. In total, we found 19 outlier trials across all tasks. Trials taking longer than 60 seconds may have resulted from technical problems, such as clicker malfunction; trials below 1 second were attributed to accidental clicks ending the trial early. The distribution of outliers per technique was as follows: *Desktop*=2, *TabletAR*=6, *ImmersiveAR*=11. By removing outlier trials, we obtained an unequal number of trials and used the non-parametric FRIEDMAN-CHI-SQUARE test for null-hypothesis testing, as well as MANN-WHITNEY-U test for pairwise significance testing.

Significance values are reported for $p < .05$ ($*$), $p < .01$ ($**$), and $p < .001$ ($***$), respectively, abbreviated by the number of stars in parenthesis. Numbers in parentheses indicate mean values in seconds (time) and mean-errors in the specific unit for each task. Results for time and error are shown in Fig. 7. Confidence intervals indicate 95% confidence. We report time and error measures for each task separately.

Distance: We found significant ($***$) differences for *time*. *Desktop* (7.8s, SD=4s) was found to be faster ($***$) than both *TabletAR* (12.9s, SD=9.1s) and *ImmersiveAR* (12.9s, SD=9.4s). For *error*, FRIEDMAN-CHI-SQUARE test did not find significant differences. However, the pairwise comparison with MANN-WHITNEY-U test revealed *Desktop*

(.09, SD=.3s) to be more precise ($*$) than *TabletAR* (.18, SD=.4s). No significant difference was found between *ImmersiveAR* (.13, SD=.3s) and *TabletAR* (.18, SD=.4s), though *ImmersiveAR* was slightly faster than *TabletAR* on average. We thus can fully accept $H_{distance-time}$, but have to reject $H_{distance-error}$ due to the lack of significance. We conclude that users were faster *and* more accurate with *Desktop*, confirming earlier findings [60].

Clusters: We found significant differences ($***$) for *time*, with *Desktop* (9.2s, SD=5s) being the fastest ($***$) (*ImmersiveAR*=17.2s, SD=11s, *TabletAR*=16.2s, SD=9s). We can thus fully accept $H_{cluster}$. The time difference is likely due to the time required to physically move one's head or the marker. For *error*, FRIEDMAN-CHI-SQUARE test found a significant effect of the task ($**$). Pair-wise comparison with MANN-WHITNEY-U test found *Desktop* (.16, SD=.4) and *ImmersiveAR* (.16, SD=.4) to be more precise ($**$) than *TabletAR* (.33, SD=.5). This result came as a surprise. Since *TabletAR* featured components of the two other environments (monoscopic display and marker interaction) we attribute the increase in precision to the ease and precision in rotation in *Desktop* and stereoscopic vision coupled with head-movement in *ImmersiveAR*.

Selection: We found highly significant differences ($***$) between all environments for *time*. *ImmersiveAR* (19.7s, SD=9s) was slowest ($***$) while *Desktop* (6.7s, SD=3s) was fastest and significantly ($**$) faster than *TabletAR* (8.6s, SD=4.7s). We thus have to reject $H_{selection}$. We attribute the lack of speed of *ImmersiveAR* to the time required to move one's head and body around the visualization. For *error*, we found *TabletAR* (2.7, SD=2.8s) to require more (***) clicks (touches) than the other two conditions (*ImmersiveAR*=1.3 (SD=1.7), *Desktop*=1.7 (SD=2)). We attribute this to the fat-finger problem as people can more accurately pinpoint smaller targets with the mouse. This was crucial in the *selection* task where some points were partly hidden and which required rotation of the model and hence added time in *TabletAR*. *ImmersiveAR* required fewer clicks than *Desktop*, indicating that in real 3D space participants could better judge when a marked cube was hit. While the precision of *ImmersiveAR* came at the price of speed, we observed that in *Desktop* participants sometimes interacted too fast and hence missed the respective targets.

Cuttingplane: We found significant differences ($*$) for this task with respect to *time*. *ImmersiveAR* (18.6s, SD=11s) was faster ($**$) than both *Desktop* (22.2s, SD=13.4s) and *TabletAR* (27.7s, SD=17s). For *error*, we found a trend towards significance ($p = .056$) for *Desktop* (22.7, SD=14.2) being less precise than *ImmersiveAR* (21.9, SD=13.8). We can thus accept $H_{cuttingplane}$, but state that generally high precision is possible in all three environments.

5.2 Interaction and Task Strategies

We were interested in participants' exploration of different strategies and affordances of the visualization environments, especially for *ImmersiveAR*. We did not force participants into a single strategy, e.g., to remain seated and rotate the marker. For *ImmersiveAR*, only 2/15 (13.3%) remained seated during all tasks, while the rest (86.6%) stood up after the first training trial and locked holograms in free space (air-lock in Fig. 1(a,c-d)). More than half of the participants (8/13) placed the visualization at the height of their head and eyes, while the others (5/13) placed the visualization hologram at the height of their chest, i.e., lower than head height. For *cluster*, participants reported on the convenience of moving themselves or their head around the air-locked hologram to observe it from all three orthogonal directions. One participant explicitly reported that she placed the visualization so that she faced all orthogonal directions to an equal extent, effectively reducing her time to move around it.

During *TabletAR*, only 2/15 (13.3%) participants stood up and moved the tablet around the visualization or the marker, while the rest remained seated and instead moved the marker. One observed problem during the *TabletAR* condition was that the distance between the marker and the screen had to be large, making viewing the tablet screen for some participants hard. The prevalent strategies for *Desktop* were fast rotation of the visualization (71%), while 35% of the participants also or exclusively rotated slowly.

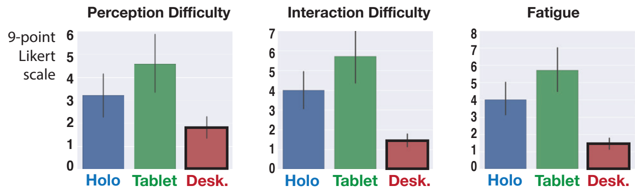


Fig. 8. Subjective ratings and perceived performances as indicated by the participants (0-9 Likert scale). Error bars: 95% CIs.

5.3 Subjective Feedback

After completing all tasks, participants filled out a questionnaire asking for subjective preferences and further feedback (Fig. 8).

Desktop was felt to be by far the easiest condition, while *TabletAR* was perceived harder than *ImmersiveAR*. Split into perception-difficulty and interaction-difficulty, *ImmersiveAR* was perceived to perform better than *TabletAR*. With respect to fatigue, *Desktop* caused the least fatigue, while *ImmersiveAR* caused less fatigue than *TabletAR*. This is perhaps surprising, given that 86.6% of all participants stood, walked, moved their arms in space, and had to wear the HMD, while most participants (86.6%) were sitting during *TabletAR*.

Subjective precision across tasks and environments mainly matched our measured results for *time* and *error* (Fig. 7). The only mismatch we found was for *TabletAR* in the *selection* task; participants reported high precision but in fact produced many clicks that did not target the marked points in the visualization. For the same task, however, *ImmersiveAR* was reported to be less precise than *Desktop*, though the recorded data indicated precision as high as for *Desktop*.

ImmersiveAR, understandably, was reported to be difficult to handle in the beginning but became more usable (*"I became accustomed to it in the end"*). The condition was also reported to be an attractive experience. Participants liked the spatial freedom and walking around (*"[I] loved the ability to walk around an object"*). They also positively reported on the stereovision (*"The hololens gives an instant understanding of depth [...]"*,*"ease with which I could judge distances/location"*, *"could see the space clearer"*), and spatial comprehension (*"comprehension was the highest with HoloLense* (SIC)*"*). However, interaction was more cumbersome (*"was very difficult to interact with."*,*"it's good as an experience but harder to for certain things beyond seeing, such as touching or slicing"*). One participant noted though she was prone to motion sickness, she did not feel any symptoms with the *ImmersiveAR*.

In *TabletAR* participants appreciated the easy and fast selection on the screen, but disliked the spatial mismatch between interaction and perception. One participant reported *"I felt constrained by the positioning of the camera in relation to the hologram markers. That could have been alleviated by moving the tablet just slightly but I didn't want to risk losing view of the marker"*. Others reported on the difficulty of manipulation *"holding tablet in hands contribute to imprecise manipulation."*, *"holding the tablet and markers involves a bit too much simultaneous manipulation to feel very precise."*. One participant suggested dragging objects on the screen, a setting common in other applications (e.g., [37]). In neither *TabletAR* nor *ImmersiveAR*, did the participants report complaints about the markers' patterns distracting them from the task or resulting in any visual interference with the visualization.

For *Desktop*, participants appreciated the ease and effectiveness of the environment (*"easier to complete [the tasks] on the flat screen"*, *"absence of Z parameter"*, *"the best interaction experience was with descktop* (SIC)*"*.) One participant summarized his/her experience as follows: *"To sum it up, Hololense* (SIC) *gives the best comprehension, Descktop* (SIC) *gives the best manipulation."*

5.4 Long-Term Training

To understand the effect of training for the *ImmersiveAR*, we analyzed the four additional sessions for the long-term training group (6 participants). Training was performed once a day, at the same time, for the four days following the participant's first session with the *ImmersiveAR*. We analyzed each task, data set, and participant individually as we believed there would be differences between participants. In particular,

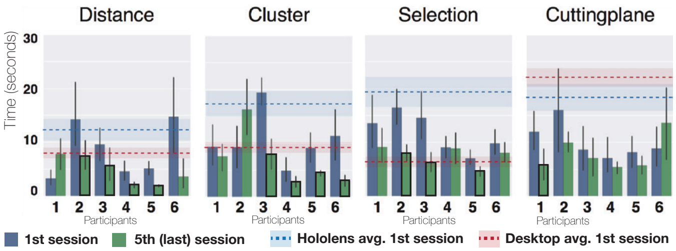


Fig. 9. Change in *time* with training for each participant (horizontal axes). Blue vertical bars indicate performance in the 1st session, while green vertical bars indicate performance in the 5th (last) session. Blue and red horizontal bars indicate mean and 95% CIs in the initial study for *ImmersiveAR* and *Desktop*, respectively.

some participants were quite fast in their first session and had little room to improve. In our analysis we excluded trials that took longer than 60 and less than 1 seconds. Results for *time* are shown in Fig. 9.

For *time* in each training trial (participant × trial × task), we calculated the linear polynomial fit over all sessions 1-5. Slopes were averaged across all trials for the same task and participant, resulting in 6 measures per task. Significance values were calculated between the results of the first session and the 5th (last) session. For *time*, out of the 24 measures (6 participants × 4 tasks), 22 showed a decrease in task completion time, 9 of which showed a significant decrease ($p < .05$) between their first and their last session. For *error*, we did not find any significant change in precision for any task.

We can partially accept $H_{training}$, though we think external conditions, such as personal performance and fatigue in the respective sessions may have caused participants to vary in their performance, and 5 sessions may be too few to obtain an effect. We also checked for difficulties that may have been introduced by specific data sets for specific days, but we did not find any evidence of such a variation. Subjective feedback from the training group showed that all participants rated their subjective improvement in *time* between 8 and 9 on a 9-point Likert scale. For precision, ratings varied between 6 and 8 on the same scale. The highest subjective increase was reported during the *selection* and *cuttingplane* tasks, which are the tasks that required the most interaction with markers. Our statistical results confirm this trend (Figure 9-*selection*). Participants reported a possible source of decrease in task-completion time could be their improved motor-control over time, as they learned to interact and navigate in 3D space. We also found a decrease in time for *distance* which may suggest that participants got used to perception with the HoloLens. Finally, all participants highly agreed they would further improve with more training.

After the training, we asked each of the 6 training participants again to rate all three environments on a 5-point Likert scale between “very inappropriate” and “very appropriate” for our set of tasks. Other than in the condition without training, participants rated the *ImmersiveAR* as appropriate as the *Desktop*. The *TabletAR* was rated worst. We see this as a positive sign that people can improve quickly with the HoloLens though being very novel to most participants, requiring very different motor actions, interactions, and strategies as on the desktop.

6 Discussion

6.1 Study Findings

We set out to find answers to the question *“How effective is interactive exploration of 3D visualizations in augmented reality?”*. The short answer is that *direct interaction with 3D holographic visualizations through tangible markers improves time and accuracy for tasks that require coordination between perception and interaction.* We found *ImmersiveAR* to outperform the other two visualization environments for tasks with high requirements for manipulation (*selection* (error), *cuttingplane* (time)). Across all tasks, *ImmersiveAR* was at least as precise as *Desktop*, despite the fact that only 2 participants (out of 15) had ever used the HoloLens before, and despite the quite new perception experience of the HoloLens provides. *TabletAR* led to the most errors in our study. Below, we report on finer grained findings from our study.

Immersive tangible AR is best for highly interactive tasks requiring detailed manipulation: We believe the low values for time and error resulted from the combination of conditions in the *ImmersiveAR* which matched both the spatiality of the visualization (3D), and the spatiality and high degrees of freedom of the interaction (3D with 6-DOF). The spatial proximity between input and output may have helped participants to coordinate interaction and perception. For *cuttingplane*, participants had a marker and head movement to rotate the view and another marker to orient the clipping plane, whereas for *Desktop* both view rotation and clipping plane orientation were performed with the mouse. As for *cuttingplane*, our findings partially confirm previous studies where TUIs have been found to be fastest for a similar task on a monoscopic screen and where interaction with a mouse has been found to be slowest [11].

Training can lead to further improvement in *ImmersiveAR*: *ImmersiveAR* led to generally slow performance across most tasks, which we attribute to a variety of reasons: *i)* participants preferred to actively move around the visualization that they had air-locked; *ii)* participants took extra time to explore holograms and verify their answers; *iii)* participants were new to the device, requiring time to adapt to the perception and motion-blur in fast head movements; and *iv)* technical delays in rendering and marker tracking, as well as occasional unresponsiveness of the clicker device. With respect to *i* and *iii*, our training data (Fig. 9) shows that some improvement is possible as people learn to coordinate perception and interaction for pointing in space. It may also be possible to gain time by more efficiently combining visualization and head movements as well as more training.

Proximity of perception and interaction spaces is important for manipulation tasks: We were surprised to find that the *TabletAR* environment led to the worst performance on almost all tasks, for both *time* and *error*. Our *TabletAR* actually was supposed to combine the best of two worlds: high DOF tangible interfaces with precise and fast interactions on a 2D touch screen. We believe the problems are due to low proximity between interaction and perception spaces, the mismatch between the two spaces' dimensionality (3D for interaction, 2D for perception), as well as the resulting visual offset and perspective mismatch between *i)* where the perceived outputs (hands, tools, visualization) appear (away from the hands), and *ii)* where they actually are (at the hands) (Fig. 3, bottom-center). However, the distance between the tablet and hands had to be large such that the pose was comfortable for participants. Subjective feedback largely reflected these conjectures.

Immersive environments afford engaging body motion: We observed that almost all participants preferred to stand while performing tasks with the *ImmersiveAR* and included their bodies into their navigation strategies. While this would be expected to increase fatigue, we did not find fatigue to be a problem during the approximately 40 minutes duration of the *ImmersiveAR* conditions. Instead, we conjecture that the ability to move can be an engaging experience compared to the otherwise passive sitting for *Desktop* environments.

***Desktop* performed generally well:** The traditional desktop environment overall led to good performance on all tasks. Beyond familiarity, we attribute the good performance to an appropriate match of the 2D interaction space (the mouse on the table) and the 2D perception (the monoscopic screen), combined with the effect of kinetic depth [48]. Another reason might be that the desktop requires minimal effort to interact with (e.g., small finger and hand movements, versus larger physical movements) and that people are well trained with mouse interactions on a desktop. We highlight that our participants were mainly young architecture students, spending much time interacting with CAD software, and may have had early access to 3D video games.

Performance in immersive environments may depend more on individual differences: Some participants subjectively reported on higher precision when using *ImmersiveAR* for perception but not for interaction, while others stated the inverse. More than in traditional desktop environments, we believe immersive environments, such as *ImmersiveAR*, increase personal variability in objective and subjective performance. For instance, they may benefit individuals differently or to varying extents, depending on the individual's abilities of spatial understanding and hand-eye coordination in general.

6.2 Limitations and Generalization

We explicitly limited our study to three visualization environments. We intentionally did not control for all factors by which they differ but focused on visual perception, interaction, and proximity. Here we discuss limitations with respect to technology, study design, visualizations, and participants as well as the generalizability of our results.

Technology Our findings are representative of the respective *combinations* of devices detailed in Table 1. Devices such as HoloLens or the choice of the marker-tracking framework (Vuforia) are individual choices meant to represent their respective class of devices and technology. They also represent the current *state of the art* of technology (early 2017) and may have imposed technical limitations to participants' performance with respect to future devices. For example, while holograms in *ImmersiveAR* were anchored in space extremely stably, we found marker tracking sometimes lagging and the HoloLens clicker did not always respond correctly to input. We found that speed, pressure, and position in which the clicker was triggered seemed to influence the success of a signal. However, correct reproduction of these parameters was not always possible. For the *selection* and *cuttingplane* tasks, very slow corrections in marker positions, crucial for pointing and positioning, were often not tracked. Further variability for *TabletAR* and *ImmersiveAR* performance might be the effect of reflections and shadows on the markers as reported elsewhere [36, 38].

In *ImmersiveAR*, participants initially complained about the small field-of-view but also reported that they got used to it during the study. Moreover, the HoloLens has been optimized for a viewing distance of between 1-2 meters. Testing holograms at arm-range, thus a distance significantly less than what the device was optimized for, may have negatively influenced user performance. However, interaction requires proximity and reachability, so being an inherent trade-off with HoloLens and similar head-mounted technology. Requiring participants to sit, thus restricting them from extensively moving around, may result in slightly different results with possibly shorter times for task completions. Touching the tablet during *selection* sometimes moved the tablet and thus prevented participants from correctly hitting the targeted point. During *cuttingplane*, participants sometimes had to switch between dominant and non-dominant hands in order to not hide the marker to which the visualization was attached, causing some discomfort. Additionally, though we planned to video-record participants' view port and manipulations through the HoloLens, streaming video data took too much of the HoloLens' performance, and effectively would have biased participants' performance records.

Our results are timely in addressing the questions of *how usable technology has gotten* and *what remains to be done*—given the renewed interest in AR and immersive technology [6, 7]. Our results, while highlighting the potential of this new technology, point to critical limitations of current technology for visualization: slightly lagging markers, the missing precision to track slight corrections in marker position and orientation, a limited field-of-view, and a reduced view quality when approaching a hologram for detailed inspection. These weaknesses remain to be addressed in future technological iterations in order to make immersive AR more competitive for the cases that we have tested. On the other hand, we found certain technological factors of the HoloLens—display resolution, stability, and complex marker pictures—to have less impact on task performance and subjective ratings of users. As we expect technology to improve over the coming years, we believe immersive AR can become a stronger competitor to traditional desktop environments for exploring 3D visualizations. Future studies may repeat our study with updated devices.

Participants Our study involved two types of preconditions in participants: i) familiarity in working with 3D content, and ii) familiarity with specific devices and visualization environments. Half of our participants were well versed in 3D graphics and manipulation as used in CAD systems and almost all participants had early experience with 3D video games. While we considered the first criteria representative for the targeted audience of 3D visualization, general familiarity with a state-of-the-art technique (*Desktop*) poses a common problem in evaluating novel techniques (*ImmersiveAR*) in visualization and HCI research. Familiarity with *Desktop* may have a positive effect on the respective results. In fact, there is always a cost in switching from a mastered technique to a novel technique as users have to master the new technology, which may take time. Our results indicate that even without training, results for *Desktop* and *ImmersiveAR* are comparable and that performance for *ImmersiveAR* can increase with training. We believe that comparing entirely untrained users on each condition does not help in assessing true performance in a real-world setup either. We encourage future studies assessing the effect of trained participants in immersive environments.

Visualization and Tasks Other 3D visualizations may feature different visual structures such as curves, complex shapes, and complete cubes [8], which we were not able to include in our study. These features may imply additional tasks such as finding the most straight or curved line, comparing shapes and shape volumes, and finding elements hidden within complete cubes. A structured overview of visualizations, visual structures, and tasks related to 3D visualization is yet missing, appart from perhaps the very high-level description in Bach et al. [8]. However, we consider our tasks—assessing distance, selecting elements, perceiving projections, placing cuttingplanes—generic and applicable to other visualizations in similar or slightly adapted form. That means our results are generalizable to other visualizations to a certain degree. Our study is meant as a first assessment and future studies may test for more specific conditions.

6.3 Bridging Visualization Environments

We believe different visualization environments, described here as combinations of input and output technologies and devices, can be seen as complementary, depending on the task; in the words of Bill Buxton: *"Everything is best for something and worst for something else"* [14]. Because visual exploration involves many different perceptual and interaction tasks it may be beneficial to mix and switch environments on the fly. We think immersive augmented reality is a promising technology but do not advocate it as a complete replacement of existing environments. Rather, we encourage *bridging* different visualization environments and allowing users to seamlessly switch between them. Moreover, while we have been focusing on each environments individually, it will be interesting to investigate combinations of augmented reality with touch-tables, physical visualizations, large walls, and desktop computers for visualization. For example, we can imagine HoloLens and paper markers being used for complex spatial interaction tasks such as placing cutting planes, brushing in 3D, or rotating a visualization. On the other side, pointing tasks, annotation, and setting visualization or filter parameters may be best supported by a desktop environment. Being able to bridge and integrate several environments is certainly a strength of augmented vs. virtual reality.

7 Conclusions and Future Work

We have presented a study comparing three visualization environments for interactive exploration tasks in 3D visualization. Our results suggest that each environment has specific strengths. More detailed investigations of individual factors will help in assessing their impact on user task performance. Our classification scheme can help in designing relevant future studies. For example, one could investigate the potential benefits of the HoloLens display combined with interactions on a multi-touch device or desktop environment. Another area of future work is an investigation of how to create easy-to-handle marker-based interaction tools to perform specific tasks for visualizations [8]. These tools could support, e.g., selecting orthogonal cutting planes, selecting arbitrary shapes, duplicating content, filtering, or annotations. We also believe that more powerful interaction and display capabilities will ultimately lead to novel or improved visualization designs and applications, integrating visualizations into the real-world to foster data literacy and collaboration [19].

Acknowledgments

This work was supported in part by NIH grant U01CA200059 and by the King Abdullah University of Science and Technology (KAUST) under award OSR-2015-CCF-2533-01.

References

[1] Matplotlib. online:http://matplotlib.org/mpl_toolkits/mplot3d/tutorial.html. last checked: March 2, 2017.

[2] Hololens. online:https://www.microsoft.com/microsoft-hololens/en-us, 2017. [last visited: March 24, 2017].

[3] Metavision. online:https://www.metavision.com, 2017. [last visited: March 24, 2017].

[4] Vuforia. online:https://www.vuforia.com, 2017. [last visited: March 24, 2017].

[5] B. Alper, T. Hollerer, J. Kuchera-Morin, and A. Forbes. Stereoscopic highlighting: 2d graph visualization on stereo displays. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(12):2325–2333, 2011.

[6] B. Bach, M. Cordeil, T. Dwyer, B. Lee, B. Saket, A. Endert, C. Collins, and S. Carpendale. Immersive analytics: Exploring future visualization and interaction technologies for data analytics. *IEEE VIS, Accepted Workshop*, 2017.

[7] B. Bach, R. Dachselt, S. Carpendale, T. Dwyer, C. Collins, and B. Lee. Immersive analytics: Exploring future interaction and visualization technologies for data analytics. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*, ISS '16, pp. 529–533. ACM, New York, NY, USA, 2016. doi: 10.1145/2992154.2996365

[8] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A descriptive framework for temporal data visualizations based on generalized space-time cubes. In *Computer Graphics Forum*. Wiley Online Library, 2016.

[9] A. Barrie, B. Cassell, and M. Cooper. Benefits of immersion for viewing 3d data. 2005.

[10] D. Belcher, M. Billinghurst, S. Hayes, and R. Stiles. Using augmented reality for visualizing complex graphs in three dimensions. In *Proceedings of IEEE and ACM Mixed and Augmented Reality*, pp. 84–93. IEEE, 2003.

[11] L. Besançon, P. Issartel, M. Ammi, and T. Isenberg. Mouse, tactile, and tangible input for 3d manipulation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 2017.

[12] M. Billinghurst, H. Kato, and I. Poupyrev. Tangible augmented reality. *Proceedings of ACM SIGGRAPH ASIA*, 7, 2008.

[13] R. Brath. 3d infovis is here to stay: Deal with it. In *IEEE VIS International Workshop on 3DVis*, pp. 25–31. IEEE, 2014.

[14] B. Buxton. Multi-touch systems that i have known and loved. online:http://www.billbuxton.com/multitouchOverview.html, 2007. [last visited: March 31, 2017].

[15] S. T. Carpendale, D. J. Cowperthwaite, M. Tigges, A. Fall, and F. D. Fracchia. The Tardis: A Visual Exploration Environment for Landscape Dynamics. In *Proceedings of Conference on Visual Data Exploration and Analysis*, number 3643, pp. 110–119, 1999.

[16] J. Chen, H. Cai, A. P. Auchus, and D. H. Laidlaw. Effects of stereo and screen size on the legibility of three-dimensional streamtube visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(12):2130–2139, 2012.

[17] S. Clarke, N. Dass, and D. H. P. Chau. Naturalmotion: Exploring gesture controls for visualizing time-evolving graphs. *Proceedings of IEEE VIS (poster session)*, 2016.

[18] M. Cordeil, B. Bach, Y. Li, E. Wilson, and T. Dwyer. A design space for spatio-coordinated interaction devices for immersive information visualisations. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, 2017.

[19] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas. Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE Transactions on Visualization and Computer Graphics*, 23(1):441–450, 2017.

[20] D. J. Cowperthwaite, M. S. T. Carpendale, and F. D. Fracchia. Visual access for 3D data. In *Proceedings on ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 175–176, 1996.

[21] R. M. C. do Carmo, B. S. Meiguins, A. S. G. Meiguins, S. C. V. Pinheiro, L. H. Almeida, and P. I. A. Godinho. Coordinated and multiple views in augmented reality environment. In *Proceedings of IEEE Information Visualization (IV'07)*, pp. 156–162. IEEE, 2007.

[22] N. Elmqvist, P. Dragicevic, and J.-D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1539–1148, 2008.

[23] R. Etemadpour, E. Monson, and L. Linsen. The effect of stereoscopic immersive environments on projection-based multi-dimensional data visualization. In *Proceedings of Information Visualization (IV)*, pp. 389–397, 2013.

[24] S. Fels, E. Lee, and K. Mase. Techniques for interactive video cubism (poster session). In *Proceedings of ACM Conference on Multimedia*, pp. 368–370. ACM, 2000.

[25] K. P. Fishkin. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing*, 8(5):347–358, 2004.

[26] G. W. Fitzmaurice, H. Ishii, and W. A. Buxton. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 442–449. ACM Press/Addison-Wesley Publishing Co., 1995.

[27] B. Fröhlich and J. Plate. The cubic mouse: a new device for three-dimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 526–531. ACM, 2000.

[28] T. Hägerstraand. What about people in regional science? *Papers in regional science*, 24(1):7–24, 1970.

[29] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 452–458. ACM, 1994.

[30] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 234–241. ACM, 1997.

[31] P. Issartel, F. Guéniat, and M. Ammi. Slicing techniques for handheld augmented reality. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 39–42. IEEE, 2014.

[32] B. Jackson, T. Y. Lau, D. Schroeder, K. C. Toussaint, and D. F. Keefe. A lightweight tangible 3d interface for interactive visualization of thin fiber structures. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(12):2802–2809, 2013.

[33] Y. Jansen and P. Dragicevic. An interaction model for visualizations beyond the desktop. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(12):2396–2405, 2013.

[34] Y. Jansen, P. Dragicevic, and J.-D. Fekete. Evaluating the efficiency of physical visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 2593–2602. ACM, 2013.

[35] D. F. Keefe and T. Isenberg. Reimagining the scientific visualization interaction paradigm. *Computer*, 46(5):51–57, 2013.

[36] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 3–12. IEEE, 2010.

[37] D. Lopez, 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 (TVCG)*, 22(5):1616–1629, May 2016. doi: 10.1109/TVCG.2015.2440233

[38] M. Luboschik, P. Berger, and O. Staadt. On spatial perception issues in augmented reality based immersive analytics. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, pp. 47–53. ACM, 2016.

[39] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[40] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[41] J. P. McIntire, P. R. Havig, and E. E. Geiselman. What is 3d good for? a review of human performance on stereoscopic 3d displays. In *SPIE defense, security, and sensing*, pp. 83830X–83830X. International Society for Optics and Photonics, 2012.

[42] J. P. McIntire and K. K. Liggett. The (possible) utility of stereoscopic 3d displays for information visualization: The good, the bad, and the ugly. In *IEEE VIS International Workshop on 3DVis (3DVis)*, pp. 1–9. IEEE, 2014.

[43] R. P. McMahan, D. Gorton, J. Gresock, W. McConnell, and D. A. Bowman. Separating the effects of level of immersion and 3d interaction techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 108–111. ACM, 2006.

[44] B. S. Meiguins, R. C. do Carmo, A. S. Goncalves, P. I. A. Godinho, and M. de Brito Garcia. Using augmented reality for multidimensional data visualization. In *Proceedings of Conference on Information Visualization (IV)*, pp. 529–534. IEEE, 2006.

[45] P. J. Metzger. Adding reality to the virtual. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pp. 7–13, Sep 1993. doi: 10.

1109/VRAIS.1993.380805

[46] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12):1321–1329, 1994.

[47] M. R. Mine, F. P. Brooks Jr, and C. H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of Conference on Computer graphics and Interactive Techniques*, pp. 19–26. ACM Press/Addison-Wesley Publishing Co., 1997.

[48] T. Sando, M. Tory, and P. Irani. Effects of animation, user-controlled interactions, and multiple static views in understanding 3d structures. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, pp. 69–76. ACM, 2009.

[49] E. Sharlin, B. Watson, Y. Kitamura, F. Kishino, and Y. Itoh. On tangible user interfaces, humans and spatiality. *Personal and Ubiquitous Computing*, 8(5):338–346, 2004.

[50] B. Shneiderman. Why not make interfaces better than 3d reality? *IEEE Computer Graphics and Applications*, 23(6):12–15, 2003.

[51] H. Slay, M. Phillips, R. Vernik, and B. Thomas. Interaction modes for augmented reality visualization. In *Proceedings of Asia-Pacific Symposium on Information Visualisation (APVis)*, pp. 71–75. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 2001.

[52] 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 (ITS)*, p. 7. ACM, 2009.

[53] M. St. John, M. B. Cowen, H. S. Smallman, and H. M. Oonk. The use of 2d and 3d displays for shape-understanding versus relative-position tasks. *Human Factors*, 43(1):79–98, 2001.

[54] R. Stenholt and C. B. Madsen. Poster: Brush, lasso, or magic wand? picking the right tool for large-scale multiple object selection tasks. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 163–164, March 2012. doi: 10.1109/3DUI.2012.6184212

[55] W. Steptoe, S. Julier, and A. Steed. Presence and discernability in conventional and non-photorealistic immersive augmented reality. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 213–218. IEEE, 2014.

[56] I. E. Sutherland. The ultimate display. *Multimedia: From Wagner to virtual reality*, 1965.

[57] F. Taher, J. Hardy, A. Karnik, C. Weichel, Y. Jansen, K. Hornbæk, and J. Alexander. Exploring interactions with physically dynamic bar charts. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 3237–3246. ACM, 2015.

[58] T. Tawara and K. Ono. A framework for volume segmentation and visualization using augmented reality. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 121–122. IEEE, 2010.

[59] M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Moller. Visualization task performance with 2d, 3d, and combination displays. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(1):2–13, 2006.

[60] B. W. Van Schooten, E. M. Van Dijk, E. Zudilova-Seinstra, A. Suinesiaputra, and J. H. Reiber. The effect of stereoscopy and motion cues on 3d interpretation task performance. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI)*, pp. 167–170. ACM, 2010.

[61] B. Wang and K. Mueller. Does 3d really make sense for visual cluster analysis? yes! In *IEEE VIS International Workshop on 3DVis (3DVis)*, pp. 37–44. IEEE, 2014.

[62] C. Ware and G. Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics (TOG)*, 15(2):121–140, 1996.

[63] C. Ware and P. Mitchell. Visualizing graphs in three dimensions. *ACM Transactions on Applied Perception (TAP)*, 5(1):2, 2008.

[64] Y. Yokokohji, R. L. Hollis, and T. Kanade. What you can see is what you can feel—development of a visual/haptic interface to virtual environment. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pp. 46–53, 265, Mar 1996. doi: 10.1109/VRAIS.1996.490509

[65] L. Yu, K. Efstathiou, P. Isenberg, and T. Isenberg. Cast: Effective and efficient user interaction for context-aware selection in 3d particle clouds. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):886–895, 2016.

[66] S. Zhai and P. Milgram. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pp. 320–327. ACM Press/Addison-Wesley Publishing Co., 1998.

[67] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 193–202. IEEE Computer Society, Washington, DC, USA, 2008. doi: 10.1109/ISMAR.2008.4637362