Augmented Reality Views for Occluded Interaction

Klemen Lilija
lilija@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Henning Pohl
henning@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Sebastian Boring
sebastian.boring@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Kasper Hornbæk
kash@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

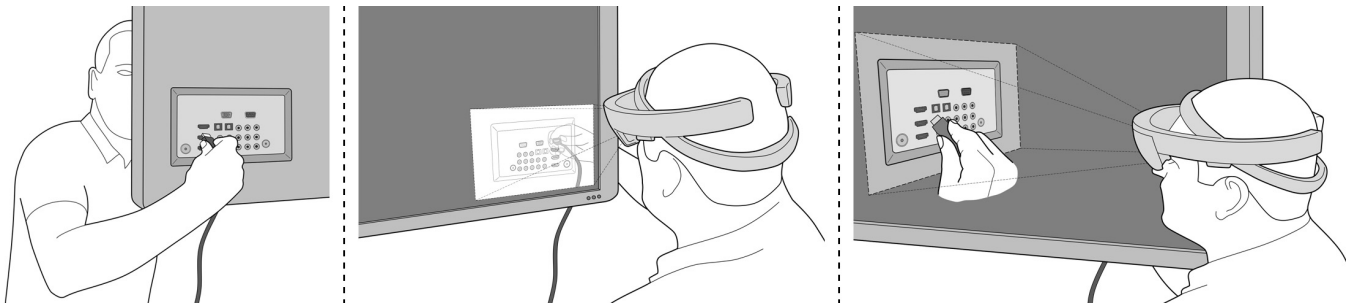

Figure 1: In some situations, we need to manipulate objects out of our sight. We investigate how different views of occluded objects support users during manipulation tasks. An example of such a task is plugging in an HDMI cable. While the port is normally out of sight, see-through view (middle) and displaced 3D view (right) enable visual feedback during interactions.

ABSTRACT

We rely on our sight when manipulating objects. When objects are occluded, manipulation becomes difficult. Such occluded objects can be shown via augmented reality to re-enable visual guidance. However, it is unclear how to do so to best support object manipulation. We compare four views of occluded objects and their effect on performance and satisfaction across a set of everyday manipulation tasks of varying complexity. The best performing views were a see-through view and a displaced 3D view. The former enabled participants to observe the manipulated object through the occluder, while the latter showed the 3D view of the manipulated object offset from the object's real location. The worst performing view showed remote imagery from a simulated hand-mounted camera. Our results suggest that alignment of virtual objects with their real-world location is less important than an appropriate point-of-view and view stability.

CHI 2019, May 4–9, 2019, Glasgow, Scotland, UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300676

CCS CONCEPTS

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

KEYWORDS

Augmented reality, manipulation task, finger-camera

ACM Reference Format:
Klemen Lilija, Henning Pohl, Sebastian Boring, and Kasper Hornbæk. 2019. Augmented Reality Views for Occluded Interaction. In *ACM Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland, UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300676

1 INTRODUCTION

Our hands can reach places and manipulate objects out of direct sight. This allows us to fish for keys under a car seat, scratch the back of our head, tighten a hidden engine bolt, or plug an HDMI cable in a port on the back of a TV (see Figure 1). During these interactions, users cannot rely on eye-hand coordination to accurately guide their reaching and grasping. Instead, they have to use their sense of proprioception, tactile feedback, and past knowledge of the object's shape, position, and orientation.

When the user's direct view of an object is occluded, the object may be observed from some other perspective. For example, endoscopic inspection cameras are commonly used in construction to provide a view inside of walls or confined spaces. Similarly, finger [29, 30, 37] and body-mounted [13, 15] cameras have been used to provide remote perspectives. We

envision camera and sensor technology to shrink further, allowing systems to collect real-time visual data from anywhere the user wants to interact. With such data, occluded objects can be rendered into the user's visual field through augmented reality (AR) headsets. However, how to best do that is unclear. Often, remote imagery is shown to users on a dedicated monitor or in a picture-in-picture (PIP) view. In contrast, work on see-through AR [21] assumes that keeping the remote view stable, in a position and orientation corresponding to remote imagery's real location, is beneficial. However, depending on the task, other types of views could be preferred.

We compared four views of occluded objects (with an additional baseline, where the object is not seen at all) to empirically identify the trade-offs with respect to performance and subjective satisfaction across a set of manipulation tasks of varying complexity. The views included two variants of PIP (a static and a dynamic, hand-mounted camera), a see-through view, and a displaced three-dimensional (3D) view. The views were implemented using virtual models of the occluded objects and virtual cameras, giving us full flexibility in manipulating the camera position and view content.

We contribute a set of views designed to support interaction with occluded objects, and empirical data on the performance and user satisfaction with those views. Our results show that the displaced 3D view performs on par with the see-through view and is often the preferred choice for occluded interaction. The worst performing view showed remote imagery from a virtual hand-mounted camera. The above two results suggests that alignment of remote imagery and its real-world location is less important than an appropriate point-of-view and view stability.

2 RELATED WORK

We build upon related work on use of remote cameras, ways of presenting the remote views, and perceptual adaptation to discrepancies between vision and proprioception.

Remote Cameras

In many situations, users require imagery of something that is not directly viewable; remote cameras provide this. For example, remote cameras in video surveillance systems enable the security officer to monitor multiple rooms. Similarly, rear view cameras in cars enable drivers to see the car's surroundings.

Cameras can also be placed on a user's body. Yang et al. explored the design space enabled by a finger-worn RGB camera [37]. They proposed using the device as an extension of the user's sight. Stearns et al. explored how a finger-mounted camera can aid users with impaired vision when reading [30]. Horvath et al. instead mapped the visual information from a finger-worn remote camera to haptic information [9]. Kurata et al. proposed a shoulder-mounted camera system for remote collaboration, where remote collaborators can control the camera's angle [15]. Depending on the application, the placement of wearable cameras can vary. Mayol-Cuevas et al. provided a model for choosing the camera's location on the body, based on the field of view and the resilience to the wearer's motion [19].

Remote cameras can also be mounted on tools. For example, endoscopic cameras are used for inspection in construction as well as by surgeons. However, endoscopic cameras are not easy to control as they move in counter-intuitive ways. This has inspired work on improving control of the remote camera's movement and ways of presenting the camera's view to the user [23, 25, 36].

Instead of showing a 2D RGB view of the scene, several papers explored the use of depth, infrared (IR), and stereoscopic cameras. For example, *Room2Room* used Microsoft's Kinect to capture a 3D point cloud of a remote participant for a telepresence application [24]. Several cameras can be combined for fully instrumented rooms, allowing for real-time volumetric capture of the scene [12].

Presenting the Views

Data from remote camera can be displayed in many ways. In diminished reality, the scene is altered to remove, hide, and see through objects and thus reveal the area of interest (see Mori et al. for a survey [21]). For example, Sugimoto et al. presented a method for removing a robotic arm from the remote camera view, enabling the operator to see more of the work area [31]. Researchers also investigated best ways to visualize hidden objects to keep the spatial understanding and enable depth judgments [6, 17, 39].

Commonly, the visual information is shown on dedicated displays or as a picture-in-picture view in a head-mounted display. Alternatively, AR can be used to combine the user's view with the camera data. Colley et al. use handheld projection to reveal the physical space on the other side of the wall [5]. The *Room2Room* used projection to render a volumetric capture into the user's space [24]. Similarly, Krempien et al. used projection to augment a surgeon's view of a patient with medical imagery [14]. Furthermore, such blending of virtual and real can enable new interaction techniques such as *The Virtual Mirror* [3] and transparency-enabling gestures as presented in *Limpid Desk* [10].

The most flexible way for presenting remote imagery is via virtual reality. An example of this is recent work by Lindlbauer and Wilson, who explored the concept of *remixed reality*, whereas user's viewpoint can be moved to an arbitrary location, while the scene's objects can be copied, moved, and removed on demand [16]. While many of the above ways of presenting the remote visual information work well for exploration tasks, they are not well suited for tasks that also require manipulation of the observed.

Perceptual Adaptation

Manipulating an object while viewing it from a uncommon perspective can cause a mismatch between the visual feedback, proprioception, and tactile feedback. Users can adapt quicker to some discrepancies than to others. A number of studies investigated pointing errors during the displacement of vision induced by prism spectacles [8, 11, 27, 35]. They found that perceptual adaptation is rapid, with pointing errors drastically decreasing within a few trials.

The adaptation is slower for the left-right reversal of the visual field. In Seklyama et al.'s study, the participants wore prism spectacles inducing left-right reversal of visual field and needed a month to adapt well enough to be able to ride a bicycle [28]. Interestingly, fMRI scans show that a new visuomotor representation emerges at about the same time. Subjects are then able to switch between the old and new mappings for eye-hand coordination.

Arsenault and Ware investigated the influence of a distorted perspective and haptic feedback on rapid interaction with virtual objects [1]. They found that accurate perspective and haptic feedback improve performance in fish tank VR. Ware and Rose [34] conducted a series of experiments to investigate the differences between virtual and real object rotations. They suggest that manipulation of virtual object is easier when the hand is in approximately the same location. Pucihar et al. investigated the perceptual issues related to rendering of AR content from the device perspective (versus user perspective) [33].

Remote imagery of interaction with occluded objects introduces even stronger perceptual mismatch than some of the prism spectacle experiments and thus likely comes with even larger adaptation costs. There are few studies investigating the influence of displacement, view stability, and point-of-view in dexterous manipulation tasks with haptic feedback.

Summary

Previous studies investigated application of remote cameras and ways of presenting the captured visual information to support exploration and visual tasks. However, few of those studies touched upon supporting manual interaction. Similarly, there are many studies on perceptual adaptation that give insight into separate aspects of manual interaction (e.g., eye-hand coordination [28]). However, it is unclear how to unite these findings to design views that best support manual interaction with occluded objects.

3 TYPES OF OCCLUDED INTERACTION

We define occluded interaction as interaction where users manipulate objects that are partially or fully occluded. Users can self-occlude parts of the scene with their body (e.g., fingers covering part of the touch screen). Similarly, tools can occlude as well (e.g., a handheld drill covering part of the drilling area). In many situations, the occluder is a separate object, located between the user and the target. For example, a couch occludes what is underneath for people sitting on it. Not only do such objects block the sight, they also constrain the users' movement and make it hard to *reach* what is occluded. For example, tightening a bolt at the back of a sink requires reaching around it. While self-occlusion and tool-occlusion can often be remedied, occlusion by the environment or larger objects is generally too costly or impossible to remove.

Our everyday life is full of tasks that require occluded interaction, and complexity of these interactions varies. For example, a relatively simple task is pressing a switch hidden under a table. However, occluded interaction can also involve complex movements (e.g., a mechanic working on a part in the back of an engine compartment).

To structure the space of occluded interaction, we looked into task taxonomies [4, 7, 18, 26, 38]. From this we built a selection of tasks that is representative of the range of movement complexities and constraints [7]. Furthermore, we selected tasks that commonly occur in everyday life (see *A taxonomy of everyday grasps in action* [18]).

All tasks we identified as representative for occluded interaction require acquisition of an object, followed by manipulation. These tasks are:

Pressing, where users toggle an object, such as a light switch, power button, or door bell.

Rotating, where users grasp an object and then twist it (with one DoF of rotation) to the desired orientation. Examples of such objects are radiator valves, thermostats, and dimmer switches.

Dragging, where users grab an object and move it (constrained to one DoF) to the desired position. Common examples are chain locks, mounted at the back of doors.

Plugging, where users have to slide one object into another. This requires matching orientation as well as position. A common example is plugging of a USB cable or stick at the back of TV or computer.

Placing, is a task where one object is put onto another. This requires positional alignment, but can also include orientation constraints. For example, hanging a key on a hook, or an umbrella onto a door handle.

4 VISUALIZING OCCLUDED OBJECTS

Removing occlusion (i.e., bringing back the object into the user's view) to enable manual interaction in the occluded area can be done in several ways. The simplest approach is having 2D cameras, placed *statically* within an environment, showing the user an unmodified view of the occluded scene. In this case, the camera maintains its spatial relation to the occluded objects, whereas the perspective the user sees depends on the camera's position and orientation. If placed wrong, the

user's hand or used tool might eventually become an occluder. Some of the techniques from diminished reality deal with this problem by removing the object of occlusion from the scene (e.g., [31]). Another issue that may arise by using static cameras to reveal the occluded scene is the mirror-effect. For example, when the camera is placed on the opposite side of the user (i.e., an angle of 180°), moving one's hand to the left in the user's frame of reference, results in movement to the right in the camera's view. This can impede manual interactions in which a user relies only on a remote camera's view.

To avoid the constraints of statically placed cameras, cameras can be placed on the user's body (e.g., on fingers as in [30]) or onto a tool controlled by the user (e.g., [31, 36]). Such *dynamic* cameras enable users to control the camera's orientation and location, allowing the users to select a perspective relevant to the task at hand. Some techniques from augmented reality allow for arbitrary viewpoint positioning (e.g., [16, 22]). Being able to control the remote camera's perspective can alleviate the frame-of-reference issues, while at the same time causing a new set of problems related to viewpoint instability and counter-intuitive control of the remote view.

With the advent of AR headsets as well as depth cameras, showing occluded objects is no longer restricted to two dimensions. One option is to simply remove the occluder, as in diminished reality, while keeping the user's frame-of-reference. That is, the occluded object becomes visible through the occluder in a *see-through* fashion (e.g., [2]). This has the advantage that the spatial relation between the object (which is now visible), the user, and the interacting hand remains the same, which should allow for interactions as if the occluder would not be present. One issue here, however, is that the object might occlude itself. This is the case when interaction occurs on the side of the occluded object, facing away from the user (e.g., user is frontally facing a TV while searching for a slot on it's backside, as in Figure 1). In this case the irrelevant parts of the occluded object could occlude the relevant parts, or the interacting hand. In cases when virtual models of the scene are available, this can be alleviated by varying the amount of transparency depending on relevance and depth of the objects.

The 3D virtual representation of the occluded object can also be shown to the user, as if the user would stand in front of it, independently of the actual location of the physical object. Such *cloned 3D* view can be positioned in a way to reveal the most relevant perspective to the user and/or enable the most ergonomic positioning of user. The cloned virtual object maintains its 3D properties and appearance. The user can move freely in front of that view, for example, to explore the sides of the occluded object or to get a different perspective, once the interacting hand is present. However, as the view is offset from the object's actual location we expect difficulties as noted in the related work on perceptual adaptation (see Section 2).

While each of these views mitigates the actual problem of occlusion, we expect that their advantages and disadvantages would render some of them more useful than others. We predict that the see-through view will perform the best on all the occluded interaction tasks, especially on the ones that involve complex manipulation and orientation of handheld objects (i.e. placing and plugging). The cloned 3D view should perform slightly worse than see-through view because of the view's displacement from the area of manipulation. However, cloned 3D should still perform better than the static 2D view as the latter does not provide depth cues nor does allow adjustment of the viewing angle. Our last assumption is that the dynamic camera view will perform the worst. While it allows adjustment of the viewpoint by moving the camera, it does so on expense of camera instability. We believe that its benefits will be outweighed by the drawbacks.

5 EVALUATING OCCLUDED INTERACTION VIEWS

To investigate the views and assumptions mentioned in the previous section, we designed an experiment in which we compared five views of occluded objects across a set of five everyday manipulation tasks.

Tasks

We base the experiment tasks on the task categories described in Section 3. Within each category we chose a familiar object encountered in everyday life. The tasks and their corresponding objects (see Figure 2) were as follows:

Pressing a Light Switch: Required either pressing (changing the state of) the left button, right button or both of the buttons on a two-button light switch.

Rotating a Dial: Required participants to rotate a dial knob with a small arrow indicating its orientation. The starting position was always at 6 o'clock, and participants had to rotate the knob to either the 9 o'clock, 12 o'clock or 3 o'clock position.

Dragging a Slider: Required participants to position a slider at a specified position. Similar to the dial, the starting position was constant and the participants were instructed to positioned the slider to either 25 %, 50 % or 75 % of the full length of the slider's rail.

Plugging in an HDMI Stick: Required the participants to plug a HDMI stick into one of four vertically arranged HDMI slots. The participants were instructed which port to use at the beginning of the task.

Placing a Key Fob: Required the participants to hang a key fob onto one of three hooks. The participants were instructed to either hang the item onto the left, middle, or right hook.

Figure 2: During the study participants performed five different tasks (left to right): *pressing* one of two light switches, *rotating* a dial, *dragging* a slider, *plugging* an HDMI stick into one of four ports, and *placing* a key fob onto one of three hooks.

Views

The experiment compared five views. Four of the views were already mentioned in the Section 4. We added a baseline condition in which participants did not receive additional visual help. All views were implemented using exact virtual copies of their corresponding physical objects. This allowed us full control over rendering of the objects into the participants' view. The views were:

No Visualization: Here, the participants did not receive any visual help from the AR headset.

Static Camera: In this view, we showed a virtual remote camera view, rendered as a picture-in-picture (PIP) in the participant's visual field via an AR headset (see Figure 3a). The PIP followed the participant's head movement similarly to the windows in operating system of Microsoft Hololens. The remote virtual camera (60 FOV, 16×9) was positioned at a static location, 30cm from the object of manipulation.

Dynamic Camera: In this view, we showed a virtual remote camera view, rendered as a PIP, similarly to the static camera view (see Figure 3b). The virtual remote camera (90 FOV, 16×9) was attached either to the tip of the participant's index finger or the tip of the handheld object (i.e., HDMI device or the key fob).

Cloned 3D: In this view, we rendered the 3D models of the occluded objects at a static location in the proximity of the participant. More specifically, the virtual model of the occluded object was rotated 70° around the vertical axis and moved 65 cm away from the object's physical location (see Figure 3c), to the left of the participants.

See-Through: Here, the 3D model of the occluded object is rendered at its actual position. The participants get an impression of seeing through the occluder (see Figure 3a).

In all of the views, we also either rendered the fingertips of the participants' index finger and thumb (in *Pressing*, *Rotating* and *Dragging* task), or the tracked handheld object (in *Plugging* and *Placing* task).

Apparatus

We used a 1x1x1 meter aluminum frame as a base for mounting the task objects, the tracking setup, and the panel that occluded the participants' view (see Figure 4). The task objects were mounted into a wooden panel placed at 30° angle to the cube's surface facing the participants.

Participants were seated in front of the cube, with a numpad placed on the left of them. The enter key on the numpad served as a trigger to start and end the trial. To track the participants' hands we used an OptiTrack setup (eight Flex 13 cameras, 1280×1024 pixels, 120 fps). In the *pressing*, *rotating* and *sliding* task, we used OptiTrack markers placed on the index finger and thumb. In the two tasks that involved a handheld object, we placed the OptiTrack markers on the object itself (see Figure 2).

Figure 3: Four of the views used in the experiment: A) Static camera, b) Dynamic camera, c) Cloned 3D and d) See-through view. No visualization view is not shown, as it does not render anything in user's field of view.

Figure 4: During the study, participants were seated in front of a $1\,m^3$ frame. They reached in from the right side and interacted with objects placed on a wall, that was tilted at a 30° angle. Movement inside the frame was tracked with an Optitrack setup and participants received visual feedback in a HoloLens headset.

We used Microsoft HoloLens mixed-reality headset (2.3M Holographic resolution, 30°H and 17.5°V FOV) to display the views. The HoloLens was connected wirelessly to a host computer which captured the movement data and tracked the objects' state via a Teensy 3.2 microcontroller. To align the coordinate systems of the HoloLens and OptiTrack, we used a one-time calibration procedure in which we placed a HoloLens's spatial anchor at the origin of OptiTrack's coordinate system.

Design

We used a repeated-measures within-subjects factorial design. Independent variables were **Task** (*Pressing*, *Rotating*, *Dragging*, *Plugging* and *Placing*), and **View** (*None*, *See-Through*, *Cloned 3D*, *Static*, and *Dynamic*).

We split the tasks into two blocks and counter-balanced across participants both the tasks within the block and the blocks itself. Tasks were split in blocks to minimize the tasks preparation time, as each of the blocks required a distinct hand tracking setup. One block contained tasks which used finger tracking (*Pressing*, *Rotating* and *Dragging*), while another contained tasks with a tracked handheld object (*Plugging* and *Placing*).

For each *Task*, participants used each of the *Views*. We used one practice repetition and two timed repetitions per *Task*. The *Task* goal was randomized (e.g., which hook to place the key fob onto) and the *Views* were presented in random order within each repetition. Each participant completed a total of 75 trials.

Measures

We collected three performance measures for each trial: (1) how long it took the participants to start the manipulation of the task objects, (2) the duration of manipulation, and (3) the error of the manipulation.

The temporal measures were calculated from the moment participants' hand entered the cube. The start of the manipulation for the pressing, rotating and dragging tasks, was the time till the first change of the object's state was detected (e.g., when the dial was rotated). For the plugging and placing conditions, this was the time till the first plugging or placement of the object. The duration of manipulation is measured from the time participants' hand entered the cube till the last state change (e.g., last rotation of the dial, or last placement of the key fob). Finally, manipulation error in the dragging and rotating conditions, was measured as the difference between the set position and the target position. The pressing, plugging, and placing tasks only allow for discrete state changes and thus error was registered accordingly. Error was recorded at the end of each trial, so intermittent wrong states were not penalized.

Participants

We recruited 24 participants (8 female, age 19–71, M = 33.6, SD = 11.9) via mailing lists and social media. All participants were right-handed, four participants wore glasses. When asked to rate their experience with augmented reality on a 1–5 scale, 13 participants stated no experience. Five participants each rated their experience as 2 and 3, and only one participant had higher than average (4) experience. For participation in the study, participants received gifts worth about $30.

Procedure

First, we explained the purpose of the study to the participants and presented an overview of the tasks and objects. Once participants understood the tasks and how to manipulate the objects, the experimenter introduced the HoloLens and the views. After being familiarized with the five views, the experiment began.

Each of the five tasks started with a practice trial for each of the views. Participant received instruction shown via the HoloLens (e.g., "See-through view: Rotate the dial to 12 o'clock."). The trial started once the participant pressed the enter key on a numpad placed in their proximity, and ended once they pressed the key again; this also initiated new instructions. Between the trials the experimenter reset the object to its starting state, and participant could start the next trial as described above. In the practice trials, participants were encouraged to take their time and ask the experimenter for any clarifications they needed.

After the practice trials, participants continued with the timed trials going through all the views twice in a randomized order. In the timed trials participants were instructed to complete the trials as fast and accurate as possible. Once they completed the task, participants had a short break during which they removed the HoloLens and were asked to fill out a questionnaire.

The questionnaire asked the participants to rate the view with respect to four Likert-scale questions: (Q1) whether they liked that visualization, (Q2) whether they could easily manipulate the object, (Q3) whether they felt supported by the visualization, and (Q4) whether the visualization allowed them to easily check the object's state. We also asked participants to describe advantages and disadvantages of the views, as well as challenges particular to the task and further comments. Participants were provided with a sheet showing the names and images of the five views as seen through the HoloLens.

After completing all tasks, participants filled out a final questionnaire where they provided an overall rating for each view and additional comments. The experiment took 60 to 90 min, depending on the participants' pace of going through the trials and how much time they used on the in-between the tasks questionnaires.

6 RESULTS

We separate the results into three sections: (1) analysis of the performance measures collected during the trials, (2) analysis of the ratings from the questionnaires, and (3) thematic analysis of the participants' comments. The overall results of the analysis show that the see-through and cloned 3D view preformed the best, with the latter being the preferred choice of participants. The worst performing and perceived view was the dynamic camera view.

Differences in Performance

To determine differences in performance, we analyze the 1200 timed trials, sans invalid ones. Trials are invalid if: (1) the participant did not interact with the object, or (2) the experimenter did not properly reset the setup for that trial. This was the case for 19 trials (i.e., 1.6 % of the trials).

Because each task required a different kind of interaction, measures are not directly comparable. For example, the average manipulation duration ranged from 3.7 s (pressing task) to 9.1 s (plugging task). To better show the differences per view, we hence normalized the data for visualization and show relative performance measures.

For statistical analysis of differences we used repeated measures two-way ANOVAs. We report on main effects of view and interaction effects, but not on effects of tasks. All post-hoc tests used permutational paired t-tests with Holm-Bonferroni correction and 1000 permutations. We ran post-hoc tests to compare the views, but not the tasks or interactions.

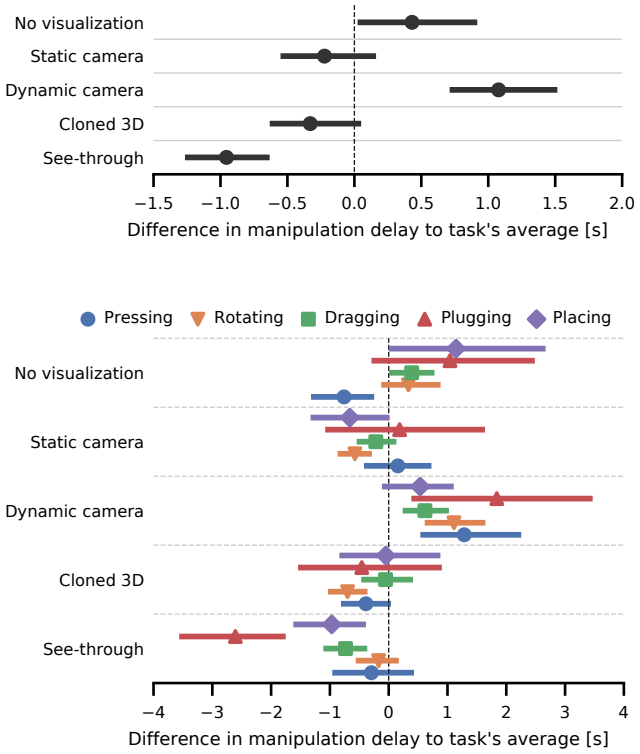

Figure 5: Delay till start of manipulation per view (top) and per view and object (bottom). Error bars show bootstrapped 95 % confidence intervals.

Time till Start of Manipulation. As shown in Figure 5, the manipulation delay differed between the views. It took participants more than 1 s longer than average to start manipulation when they used the dynamic camera view. On the other hand, with the see-through view they started manipulation almost a second earlier on average.

Figure 5 also shows how this delay differed depending on the task. This highlights interaction effects, such as the see-through view being particularly beneficial in the plugging task. Similarly, no visualization fared worse for all tasks but the pressing task, where participants only had to press a light switch.

We found a main effect of view ($F(4,92) = 12.0, p < 0.001, \eta^2 = 0.06$) as well as an interaction effect ($F(16,368) = 2.7, p < 0.001, \eta^2 = 0.04$). Post-hoc testing showed significant differences between the dynamic camera view and all other views ($p < 0.05$), but not with the no visualization condition ($p = 0.2$). Furthermore, the see-through view was significantly different from no visualization ($p < 0.05$).

Duration of Manipulation. For the duration of manipulation we also see differences between the views (see Figure 6). Here the dynamic camera view performed badly, resulting in task

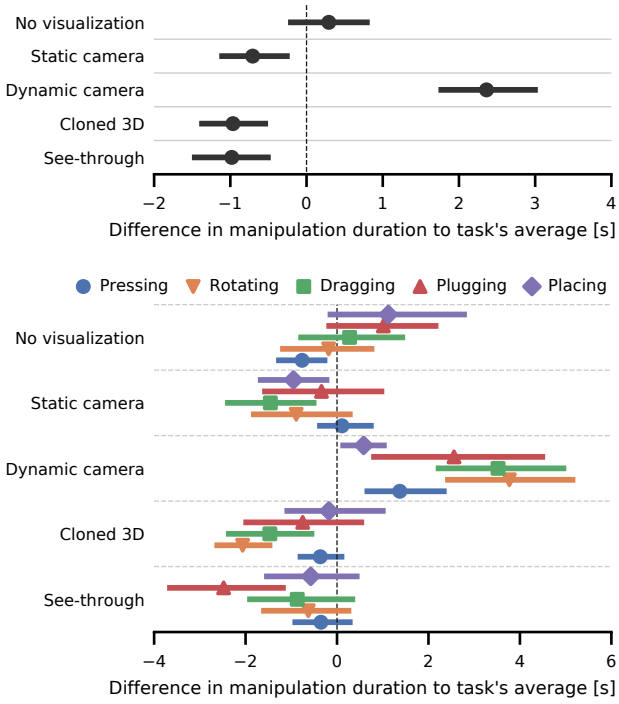

Figure 6: Duration of manipulation per view (top) and per view and object (bottom). Error bars show bootstrapped 95 % confidence intervals.

durations longer by an average of more than 2 s. Having no visualization also impacted performance, albeit not as much as the dynamic camera view.

Interaction effects between task and view are also visible in Figure 6. For example, the cloned 3D view worked especially well for the rotating task, while dragging and rotating tasks are much harder with the dynamic camera view than placing tasks.

We found a main effect of view ($F(4,92) = 24.9, p < 0.001, \eta^2 = 0.11$), as well as a significant interaction effect ($F(16,368) = 2.9, p < 0.001, \eta^2 = 0.05$). Post-hoc testing showed significant differences between the dynamic camera view and all other views ($p < 0.05$).

Manipulation Error. Finally, we investigated how precisely participants were able to manipulate the objects. In the dragging and rotating conditions, error is measured as the difference between the set position and the target position. The pressing, plugging, and placing tasks only allow for discrete state changes and we only compare the presence or absence of error. Error is recorded at the end of each trial, so intermittent wrong states are not penalized.

We separately analyze the error for dragging and rotating (relative) from the other three conditions (absolute). Figure 7 shows how the view influenced the two kinds of manipulation

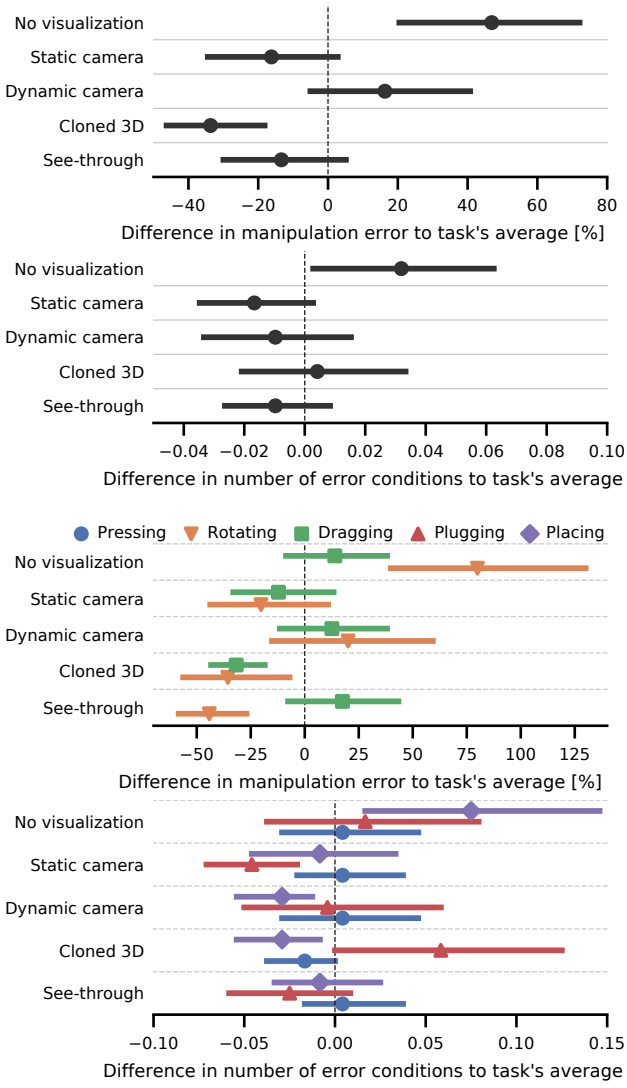

Figure 7: Relative error and absolute errors per view (top two) and per view and object (bottom two). Error bars show bootstrapped 95 % confidence intervals.

error. As can be seen, in the no visualization condition participants were more prone to errors than in other conditions. While participants could still use tactile and proprioceptive information, the lack of visual feedback made accurate manipulation challenging.

Figure 7 also shows the interactions between task and error. For example, with the see-through view the rotating task resulted in less error than the dragging task. With no visualization, this relationship is inverted. For the former, we found a main effect of view ($F(4,92) = 3.8, p < 0.001, \eta^2 = 0.05$). However, there was no interaction effect ($F(4,92) = 2.2, p = 0.07, \eta^2 = 0.03$). Post-hoc testing showed significant differences between the cloned 3D view no visualization ($p < 0.05$).

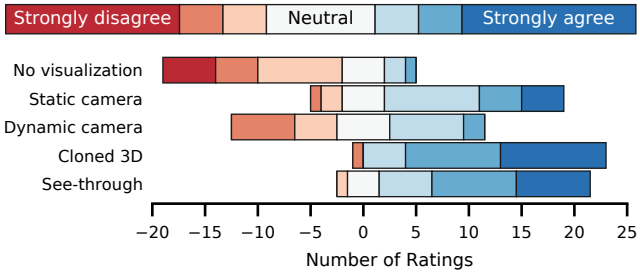

Figure 8: At the end of the study, participants provided an overall rating of each view, indicating on a 7-point Likert scale how well the view supported them in the tasks. Shown here are the number of ratings for each level of the scale, stacked horizontally to highlight overall trends.

For the absolute error conditions, we also found a significant effect of view ($F(4,92) = 2.5, p < 0.05, \eta^2 = 0.02$), as well as an interaction effect ($F(8,184) = 2.0, p < 0.05, \eta^2 = 0.05$). However, post-hoc testing showed no significant differences in any comparison of two views.

Differences in Ratings

Asked for their overall rating of each view at the end of the study, participants gave favorable ratings to all but the dynamic camera view and the no visualization condition (see Figure 8). A Friedman test confirmed the significant effect of view on overall rating; $\chi^2(4) = 54.644, p < 0.001$. Post-hoc testing with a pairwise Wilcoxon signed rank test with Holm-Bonferroni correction showed significant differences between the cloned 3D view and no visualization ($p < 0.001$), as well as the dynamic ($p < 0.001$), and static ($p < 0.05$) camera views. The see-through view also differed from the dynamic camera view ($p < 0.001$) and no visualization ($p < 0.001$). Finally, the static camera view was significantly different from the dynamic camera view ($p < 0.01$) and no visualization ($p < 0.01$). In addition to the overall preferences, we took a closer look at the ratings the participants provided after each block. There we noted relatively consistent ratings for all of the views except the dynamic view. The latter was rated more highly than average in all four questions after participants did the placing task.

For statistical analysis of the factors influencing these ratings, we used cumulative link mixed models that enabled regression on the ordinal rating data in repeated measure designs. We fit two models to the data: one including interaction effects between view and object, and one without these effects. Both models included the user id as random effect variable. A nested model ANOVA showed that the interaction between view and object is significant ($p < 0.001$). Analysis of the main effects on the model with Chi-squared tests showed no significant effect of object ($p = 0.2$), but a significant effect of view ($p < 0.001$).

Qualitative Differences

The questionnaire given to participants between the tasks and at the end of the experiment contained a set of open-ended questions. We did a thematic analysis on the participants' responses, Table 1 shows the reoccurring themes and the number of participants mentioning the topic.

Static Camera. Participants liked the static view as it provided an overview of the area of manipulation and enabled seeing the state of the manipulated object.

Participants were split over how supportive the view was for manipulation. Five participants found the view helpful for movement and "easier orientation of objects" (P1), while five had troubles "to determine distance and details" (10) and "finding the correct angle" when manipulating the occluded objects.

Four participants complained about the distance of the static camera (P3: "The static camera is too far away."). This is a limitation of the fixed camera perspective as it cannot at both provide a good overview and a detailed view.

Dynamic Camera. Participants disliked dynamic camera view for a number of reasons. Participants complained about difficulties of seeing the relevant parts of their interaction. P5 mentioned that "dynamic view was annoying because the camera kept shifting and got in the way of getting a good view." A number of participants also mentioned self-occlusion as a problem, because either the handheld object or "the fingers obscured the view" (P11).

The unstable camera perspective "was confusing to adjust to" (P3), and impeded interaction with occluded objects. This is supported by the qualitative results, which showed that dynamic camera view took participants the longest to finish the manipulation tasks (see Figure 6).

When using dynamic camera view in task where the virtual camera was showing the perspective of the handheld objects, three participants drew parallels with the point-of-views they use in games. P9 said "it was like playing a 1st person shooter" and P7 that "it's like driving a spaceship.".

Surprisingly, the dynamic camera view was less disliked in the placing task (Figure 8). Three participants mentioned that in the placing task the dynamic camera view was "not horrible anymore" (P23). Task-dependence of view preference was also expressed by P10, saying that "it's funny how different views are helpful for different tasks."

Judging by the participants responses the handheld object-mounted camera was more helpful than the finger-mounted camera because of more relevant and stable perspective.

Cloned 3D. The cloned 3D view was perceived as intuitive (P23: "It was intuitive, I didn't have to translate the experience to make sense of what to do next."), natural and real. One of the participants event tried reaching for the virtual object to

Table 1: Benefits and drawbacks of the views as expressed by the participants.

View	Benefits	Drawbacks
Static camera	good overview (10) view's stability (2)	distant view (4)
Dynamic camera	game-like (3)	hard to manipulate (10) bad overview (8) confusing (8) self-occlusion (7) unstable view (4)
Cloned 3D	natural, intuitive, real (10) easy to manipulate (9) good overview (9) depth perception (3)	confusing (4)
See-through	natural, intuitive, real (8) easy to manipulate (11) good overview (8) depth perception (2)	confusing (5)

manipulate it before realizing that the physical object is at a different location. Three participants explicitly mentioned that the "cloned 3D view was advantageous because it helped (the) depth perception" (P24).

While the cloned 3D view was intuitive for most, five participants mentioned difficulties when using cloned 3D view. P23 mentioned that "it was confusing, like trying to do everything mirrored." and P2 mentioned that "the angle (of the object) was hard to adjust" during the plugging task.

Two participants made an observation mentioning that they saw the offset of virtual object from the real object's location as a benefit (P19: "With the cloned view the fact that you were looking off to the side made it easy to abstract movement from visualization")

See-through. Similar to the cloned 3D, the see-through view was also perceived as intuitive by ten participants. After using it in a task for the first time P22 exclaimed "this was scary easy."

The see-through view was perceived as giving a good overview of the area of interaction and good at supporting object manipulation. P5 said that "it enables your brain to relax since you see it as you would."

Contrary to the above, some participants found the see-through view confusing. P12 called it "an odd perspective" and P9 mentioned that "seeing the object from behind makes it difficult". Seeing the object from behind can also cause confusion between left and right, as expressed by four participants. For example, P19 turned the dial to nine instead of three in one of the practice trials and after noticing it exclaimed: "Oh yeah, it's because I see it from behind."

7 DISCUSSION

We compared four views of occluded objects to identify the best support for occluded interaction. We found few differences between the views in performance (i.e., manipulation duration and manipulation error). However, the dynamic camera view performed the worst. With the cloned 3D and see-through views participants completed the tasks the fastest; these two views were also rated the highest on subjective satisfaction. Static camera view also supported occluded interaction well. In light of these results, we discuss three factors that contributed to these results: *point-of-view*, *view stability* and *view displacement*.

Point-of-View: The cloned 3D and see-through view were rated as the most liked and supportive. Many participants mentioned that those views felt natural and intuitive and that they showed the relevant part of the interaction. This was partly due to participants being able to choose their point-of-view by moving their head, just as we do in non-occluded interactions. This was not the case with the static camera view, thus some participants complained about the camera's view being too far away. Such limitations result from a fixed viewpoint, as it cannot provide both a good overview and a detailed view.

View Stability: Being able to change the point-of-view brings several benefits. However, if it comes at the expense of view stability, the drawbacks can quickly outweigh the benefits. Poor view stability was most noticeable in the dynamic camera view. While the participants could change the point-of-view, they were often confused by not knowing what will be shown next or how their hand movements affects the viewpoint. Issues with view stability can be only partly mitigated by smoothing the camera's movement. Good view stability requires intuitive mapping between the user's and viewpoint's movement (e.g., as in cloned 3D view).

View Displacement: Another factor that should have affected the performance of the view and users' satisfaction is view displacement. Past works suggest that view displacement negatively affects performance in manipulation tasks (e.g., [34]). Taking this into account, the cloned 3D view should have performed worse than see-through, as it was rotated and offset from the actual location of the occluded objects. We assume that the negative effect of the displacement is not noticeable in our study because of the use of everyday tasks. This allowed participants to rely on their tactile sense and past knowledge of familiar objects. The discrepancy between our results and past findings warrant further studies investigating the influence of view displacement on complex manipulation tasks, in which participants can rely on a spectrum of senses and skills used in their daily lives.

These factors are important, not only when considering the future research on interactions with occluded objects, but any research that deals with out-of-sight interaction. For example, dexterous input for AR headsets that involves hand movement out of user's direct sight.

Limitations

There are a few study limitations related to experiment design decisions. First, we conducted the experiment in a controlled setting to limit the external influences when collecting performance measures. This means that occluded interactions in the experiment only approximate the ones from real-life situations. However, we believe that the use of everyday tasks and objects, even if only in an artificial setting, is generalizable to many real-life situations.

Second, we chose one specific configuration and appearance for each of the view. A different appearance (e.g., transparency in see-through view) or configuration (e.g., the static camera placed at a different angle) might have influenced the results. Evaluating these aspects would require comparison of view variations, which would have extended our experiment duration (90 min per participant) even further.

Third, all the views were simulated by modeling the occluded objects and the use of virtual cameras. This gave us flexibility at expense of realism. We believe this was not a major issue considering the comments of participants' expressing how real the objects looked.

Last, participants had only a short practice run with each of the views. It is possible that with extensive training or longitudinal use participants could have adapted to more uncommon views.

Future Work

While our study shows that cloned 3D and see-through views support occluded interaction well, there are several open questions and directions to explore.

It is unclear what parts of an occluded scene and of the user's body should be rendered to best support occluded interaction. In our study, participants saw only the fingertips of their index finger and thumb, or the handheld object. Despite the minimalistic rendering of the hand's location, we did not note any complaints about showing too little. On the contrary, we had comments about self-occlusion interfering with the interaction. A systematic investigation into what hand features to render to best support eye-hand coordination would help make more informed decisions when designing the views for occluded interaction.

When considering the appearance of the occluded objects, investigating more extreme visual alternations can reveal new ways of supporting manipulation of occluded objects. For example, exploring planar abstractions [20], or hybrid visualization of see-through and cloned 3D view might offer additional benefits. Such visual alternations would also require unrealistic mapping between user's movement and visual feedback, as for example explored by Teather and Stuerzlinger [32].

8 CONCLUSION

We investigated how well different views can support manual interaction with objects that users cannot directly see. We evaluated the system in a lab study where we varied views and tasks in a controlled manner. We found out that see-through and cloned 3D view perform the best, with the latter one being preferred by the participants. The worst performing and most disliked view simulated remote imagery from a hand-mounted camera. We believe that alignment of remote imagery of the manipulated objects and their real-world location is less important than suggested by previous work. Furthermore, our results highlight the importance of view stability and an appropriate point-of-view.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 648785).

REFERENCES

[1] Roland Arsenault and Colin Ware. 2000. Eye-Hand Co-Ordination with Force Feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*. ACM Press, New York, New York, USA, 408–414. https://doi.org/10.1145/332040.332466

[2] P. Barnum, Y. Sheikh, A. Datta, and T. Kanade. 2009. Dynamic seethroughs: Synthesizing hidden views of moving objects. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 111–114. https://doi.org/10.1109/ISMAR.2009.5336483

[3] Christoph Bichlmeier, Sandro Michael Heining, Marco Feuerstein, and Nassir Navab. 2009. The virtual mirror: a new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging* 28, 9 (2009), 1498–1510.

[4] Ian M Bullock and Aaron M Dollar. 2011. Classifying human manipulation behavior. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*. IEEE, 1–6.

[5] Ashley Colley, Olli Koskenranta, Jani Väyrynen, Leena Ventä-Olkkonen, and Jonna Häkkilä. 2014. Windows to other places: exploring solutions for seeing through walls using handheld projection. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 127–136.

[6] Mustafa Tolga Eren and Selim Balcisoy. 2018. Evaluation of X-ray visualization techniques for vertical depth judgments in underground exploration. *The Visual Computer* 34, 3 (2018), 405–416.

[7] Thomas Feix, Ian M Bullock, and Aaron M Dollar. 2014. Analysis of human grasping behavior: Object characteristics and grasp type. *IEEE transactions on haptics* 7, 3 (2014), 311–323.

[8] Richard Held, Aglaia Efstathiou, and Martha Greene. 1966. Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology* 72, 6 (1966), 887.

[9] Samantha Horvath, John Galeotti, Bing Wu, Roberta Klatzky, Mel Siegel, and George Stetten. 2014. FingerSight: Fingertip Haptic Sensing of the Visual Environment. *IEEE Journal of Translational Engineering in Health and Medicine* 2 (2014), 1–9. https://doi.org/10.1109/JTEHM.2014.2309343

[10] Daisuke Iwai and Kosuke Sato. 2011. Document search support by making physical documents transparent in projection-based mixed reality. *Virtual reality* 15, 2-3 (2011), 147–160.

[11] LS Jakobson and Melvyn A Goodale. 1989. Trajectories of reaches to prismatically-displaced targets: evidence for "automatic" visuomotor recalibration. *Experimental Brain Research* 78, 3 (1989), 575–587.

[12] Bernhard Kainz, Stefan Hauswiesner, Gerhard Reitmayr, Markus Steinberger, Raphael Grasset, Lukas Gruber, Eduardo Veas, Denis Kalkofen, Hartmut Seichter, and Dieter Schmalstieg. 2012. OmniKinect: Real-time Dense Volumetric Data Acquisition and Applications. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology (VRST '12)*. ACM, New York, NY, USA, 25–32. https://doi.org/10.1145/2407336.2407342

[13] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 167–176. https://doi.org/10.1145/2380116.2380139

[14] Robert Krempien, Harald Hoppe, Lüder Kahrs, Sascha Daeuber, Oliver Schorr, Georg Eggers, Marc Bischof, Marc W. Munter, Juergen Debus, and Wolfgang Harms. 2008. Projector-Based Augmented Reality for Intuitive Intraoperative Guidance in Image-Guided 3D Interstitial Brachytherapy. *International Journal of Radiation Oncology*Biology*Physics* 70, 3 (mar 2008), 944–952. https://doi.org/10.1016/j.ijrobp.2007.10.048

[15] Takeshi Kurata, Nobuchika Sakata, Masakatsu Kourogi, Hideaki Kuzuoka, and Mark Billinghurst. 2004. Remote collaboration using a shoulder-worn active camera/laser. In *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, Vol. 1. IEEE, 62–69.

[16] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 129:1–129:13. https://doi.org/10.1145/3173574.3173703

[17] Fei Liu and Stefan Seipel. 2018. Precision study on augmented reality-based visual guidance for facility management tasks. *Automation in Construction* 90 (2018), 79–90.

[18] J. Liu, F. Feng, Y. C. Nakamura, and N. S. Pollard. 2014. A taxonomy of everyday grasps in action. (Nov 2014), 573–580. https://doi.org/10.1109/HUMANOIDS.2014.7041420

[19] Walterio W Mayol-Cuevas, Ben J Tordoff, and David W Murray. 2009. On the choice and placement of wearable vision sensors. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 39, 2 (2009), 414–425.

[20] James Mccrae, Niloy J Mitra, and Karan Singh. 2013. Surface perception of planar abstractions. *ACM Transactions on Applied Perception (TAP)* 10, 3 (2013), 14.

[21] Shohei Mori, Sei Ikeda, and Hideo Saito. 2017. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications* 9, 1 (dec 2017), 17. https://doi.org/10.1186/s41074-017-0028-1

[22] Shohei Mori, Momoko Maezawa, and Hideo Saito. 2017. A work area visualization by multi-view camera-based diminished reality. *Multimodal Technologies and Interaction* 1, 3 (2017), 18.

[23] Nassir Navab, Joerg Traub, Tobias Sielhorst, Marco Feuerstein, and Christoph Bichlmeier. 2007. Action-and workflow-driven augmented reality for computer-aided medical procedures. *IEEE Computer Graphics and Applications* 27, 5 (2007), 10–14.

[24] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2016. 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 - CSCW '16*. ACM Press, New York, New York, USA, 1714–1723. https://doi.org/10.1145/2818048.2819965

[25] Rob Reilink, Gart de Bruin, Michel Franken, Massimo A Mariani, Sarthak Misra, and Stefano Stramigioli. 2010. Endoscopic camera control by head movements for thoracic surgery. In *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE, 510–515.

[26] Alberto Romay, Stefan Kohlbrecher, David C Conner, and Oskar Von Stryk. 2015. Achieving versatile manipulation tasks with unknown objects by supervised humanoid robots based on object templates.. In *Humanoids*. 249–255.

[27] Yves Rossetti, Kazuo Koga, and Tadaaki Mano. 1993. Prismatic displacement of vision induces transient changes in the timing of eye-hand coordination. *Perception & Psychophysics* 54, 3 (1993), 355–364.

[28] Kaoru Sekiyama, Satoru Miyauchi, Toshihide Imaruoka, Hiroyuki Egusa, and Takara Tashiro. 2000. Body Image as a Visuomotor Transformation Device Revealed in Adaptation to Reversed Vision. *Nature* 407, 6802 (sep 2000), 374–377. https://doi.org/10.1038/35030096

[29] Roy Shilkrot, Jochen Huber, Jürgen Steimle, Suranga Nanayakkara, and Pattie Maes. 2015. Digital Digits: A Comprehensive Survey of Finger Augmentation Devices. *ACM Comput. Surv.* 48, 2, Article 30 (Nov. 2015), 29 pages. https://doi.org/10.1145/2828993

[30] Lee Stearns, Victor DeSouza, Jessica Yin, Leah Findlater, and Jon E. Froehlich. 2017. Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17*. ACM Press, New York, New York, USA, 361–362. https://doi.org/10.1145/3132525.3134812

[31] Kazuya Sugimoto, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. 2014. Half-diminished reality image using three rgb-d sensors for remote control robots. In *Safety, Security, and Rescue Robotics (SSRR), 2014 IEEE International Symposium on*. IEEE, 1–6.

[32] Robert J. Teather and Wolfgang Stuerzlinger. 2008. Exaggerated Head Motions for Game Viewpoint Control. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share (Future Play '08)*. ACM, New York, NY, USA, 240–243. https://doi.org/10.1145/1496984.1497034

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

[34] Colin Ware and Jeff Rose. 1999. Rotating virtual objects with real handles. *ACM Transactions on Computer-Human Interaction (TOCHI)* 6, 2 (1999), 162–180.

[35] Robert B Welch and Gerald Goldstein. 1972. Prism adaptation and brain damage. *Neuropsychologia* 10, 4 (1972), 387–394.

[36] Mark Wentink, Paul Breedveld, Dirk W Meijer, and Henk G Stassen. 2000. Endoscopic camera rotation: a conceptual solution to improve hand-eye coordination in minimally-invasive surgery. *Minimally Invasive Therapy & Allied Technologies* 9, 2 (2000), 125–131.

[37] Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2012. Magic Finger: Always-available Input Through Finger Instrumentation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 147–156. https://doi.org/10.1145/2380116.2380137

[38] Tsuneo Yoshikawa. 2000. Force control of robot manipulators. 1 (2000), 220–226.

[39] Stefanie Zollmann, Raphael Grasset, Gerhard Reitmayr, and Tobias Langlotz. 2014. Image-based X-ray visualization techniques for spatial understanding in Outdoor Augmented Reality. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*. ACM, 194–203.