2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Diminished Reality for Close Quarters Robotic Telemanipulation

Ada V. Taylor, Ayaka Matsumoto, Elizabeth J. Carter, Alexander Plopski, and Henny Admoni

Abstract— **In robot telemanipulation tasks, the robot can sometimes occlude a target object from the user's view. We investigate the potential of diminished reality to address this problem. Our method uses an optical see-through head-mounted display to create a diminished reality illusion that the robot is transparent, allowing users to see occluded areas behind the robot. To investigate benefits and drawbacks of robot transparency, we conducted a user study that examined diminished reality in a simple telemanipulation task involving both *occluded* and *unoccluded* targets. We discovered that while these visualizations show promise for reducing user effort, there are drawbacks in terms of task efficiency and user preference. We identified several friction points in user experiences with diminished reality interfaces. Finally, we describe several design trade-offs among different visualization options.**

I. Introduction

Robotic telemanipulation provides the opportunity for users to perform tasks with greater strength, reach, and repeated accuracy than they can achieve alone. However, when operating robots, users often face occlusion problems [2] due to the layout of the environment, other objects within the task space, or the robot body occluding their view (Figure 1a). To fix this issue, users must either move the robot or change their own perspectives of the scene. This may be difficult for a variety of reasons, especially in situations where the viewpoint cannot easily be changed, such as remote teleoperation or for users with motor impairments.

Diminished Reality (DR) provides an approach to mitigate occlusions. By overlaying computer generated views of the scene onto the occluding object, DR effectively renders the object transparent, enabling a direct view of the working environment. While DR has been explored on Video See-Through Head-Mounted Displays (VST-HMDs) and mobile phones [5], these are impractical for telemanipulation. Mobile phones must be held in the user's hands while manipulating the robot. VST-HMDs must re-create the user's view of the scene and thus are not fail-safe. On the other hand, Optical See-Through Head-Mounted Displays (OST-HMDs) do not modify the user's view of the real world because the user can still see the environment. We leverage recent advances in OST-HMDs to explore the viability of using DR in close-quarters robotic telemanipulation tasks.

In this work, we investigate two distinct DR representations. Our first DR interface creates the illusion that users are seeing through the robot by overlaying background information onto the robot itself (Fig. 1b). Our second interface adds an outline around the transparent robot to enable users to keep track of the arm's position better (Fig. 1c).

To investigate the potential of these two DR interfaces, we conducted a within-subjects user study with 36 participants teleoperating a Kinova MICO arm and wearing a Microsoft HoloLens for DR visualizations. We measured effort, efficiency, preference, and performance in three different visualization modes (*normal*, *invisible*, and *outlined*) when the location of the target was both *occluded* and *unoccluded* by the robot arm. We recorded the movements of users' heads, time required to execute each task, and path of the robot arm. We conducted surveys during and after the experiment.

Our results show that using a DR interface on an OST-HMD could indeed help users address occlusion problems in everyday telemanipulation tasks without requiring them to expend as much effort changing viewpoints. However, although our DR interfaces reduced users' average head movement during the task, it came at the cost of increased mental demand. This effect was particularly strong when the information provided did not align well with the task requirements, such as when the arm was unnecessarily invisible, and when the outlined mode provided visual clutter.

II. Related Work

Handling Occlusions in Telemanipuation A variety of strategies have been deployed to address the problem of occlusion in telemanipulation. Some designs attach forward-facing cameras to the robot's manipulator to provide a first-person view on a screen [23] or in augmented reality (AR) [11], while others provide a third-person view using an additional actuated camera that automatically moves to an unoccluded view of the manipulation being conducted by the primary arm [21]. While these approaches mitigate occlusion issues, they require the viewer to re-map the additional camera perspective to their teleoperation frame of reference, which can be challenging. In contrast, our technique provides the user an extension of their existing third-person view of the robot, rather than additional perspectives.

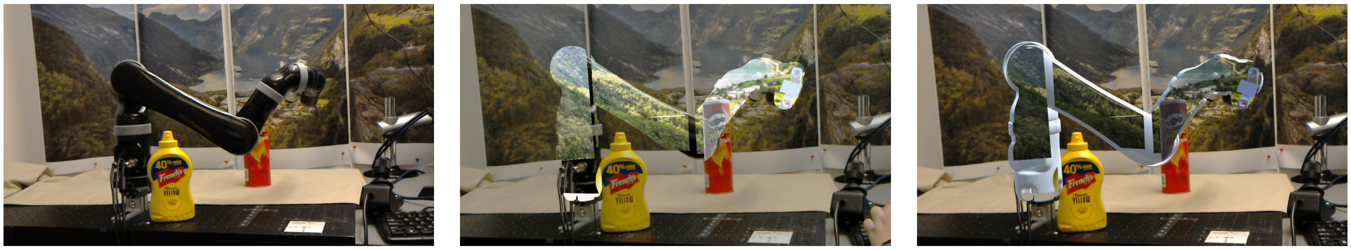
AR in HRI First concepts of AR-based robot manipulation were presented more than 20 years ago [18], and the development of more user-friendly AR devices in recent years has led to more research on AR for robot control [17], [25], [2].

AR approaches in HRI have primarily focused on adding information to the scene, often by physically projecting relevant information about a task onto the task space [8],

Ada V. Taylor, Elizabeth J. Carter, and Henny Admoni are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. {adat+iros, ejcarter, hadmoni}@andrew.cmu.edu Ayaka Matsumoto did this work while at the Nara Institute of Science and Technology and Alexander Plopski is with the University of Otago. alexander.plopski@otago.ac.nz.

This work was supported in part by JSPS Program for Fostering Globally Talented Researchers, the NAIST Bilateral research grant, the National Science Foundation (#IIS1755823, #DGE1745016), Paralyzed Veterans of America Research Foundation, and Sony Corporation.

978-1-7281-6211-9/20/$31.00 ©2020 IEEE



(a) Our control, *normal* mode

(b) *invisible* mode

(c) *outlined* mode

Fig. 1: We investigate a diminished reality illusion of transparency designed to aid users in robotic telemanipulation during tasks where the grasping target is occluded by the robot arm. In a user study, we assessed the efficacy of two types of transparency, (b) *invisible* and (c) *outlined*, versus (a) a non-transparent *normal* control.

[27], [6], [13]. Our use of a head-mounted display (HMD) presents a more consolidated hardware platform for adding information to the scene that can be used as a user travels between multiple scenes or workspaces, rather than requiring multiple projector installations.

The potential for AR to communicate alternate viewpoints in an intuitive manner has been investigated in remote teleoperation of a humanoid robot from a first person perspective [1] and in virtual first-person robot control [15]. AR can also be used to help users "imagine" the path or route that they will take in advance of a task [26]. These approaches can leave the operator unaware of events happening in their proximity. Our extension of the user's natural perspective does not disconnect users from their surroundings.

Head-mounted displays and sensors also present an area of high potential for supporting users with different informational needs or methods of physical interaction with the world. Grice et al. [9] used a head-mounted sensor to sense winces as cue to act as an e-stop for a robot arm. Munroe et al.'s [20] AR eyeglasses for rehabilitation at home demonstrate how an AR system can be applied in a home environment to enhance quality of life or provide additional feedback to differently abled users performing a daily task. Similarly, Aruanno et al. [3] use AR as a therapeutic tool for users with Alzheimer's Disease, conducting a user study with thirty elderly participants and receiving positive feedback about the viability of using the HoloLens in these populations. Cheung et al. [7] used an HMD to allow users to experience tourism in different locales.

Diminished reality DR, a subcategory of AR, has previously been considered for a variety of tasks, such as X-Ray vision [22], manufacturing [19], and teleoperation [24]. It has demonstrated the potential to communicate spatial information more effectively than video feeds of multiple secondary viewpoints, aiding users in quickly and accurately performing scene understanding tasks [4]. However, many of these systems were designed for VST systems and are not viable for in-situ robot control due to physical constraints of mobile devices and safety concerns of VST-HMDs.

III. Diminished Reality System

To explore the usability of DR in everyday applications we developed our system on hardware that can be deployed in a user's home or workplace. Our system consists of three elements: a system to track the occluded portion of the environment, a robot to interact with the environment, and an OST-HMD to present the occluded environment to users.

A. Environment Sensing

To render background information in the DR visualization, that information must first be recovered from sensors, most often cameras, in the scene [19], [24]. One approach is to warp images captured by additional cameras into the user's view to approximate what users would see. Although image warping is computationally inexpensive, it usually results in artifacts due to missing depth information when transforming from one viewpoint to another. Another approach is to reconstruct a model of the occluded environment and project this model into the user's view. This approach is more accurate, but more computationally expensive, as the scene must be recovered and updated in real time. This is not possible on the Microsoft HoloLens due to its limited computational capabilities. To reduce the computational complexity of recovering a background model, we used a precaptured environment model that is tracked with an OptiTrack [14] system. This provided consistency and ensured that all participants had a similar visual experience.

Our environment was composed of a single Pringles can that users could pick up and move around, and several walls in the background. We attached fiducial markers to the walls to create a non-uniform background, and to the table to calibrate the transformations between all elements of our setup. To track the pose of the Pringles can via OptiTrack, we attached a fiducial marker and several OptiTrack markers to the can and calibrated the transformation between these via hand-eye calibration. Finally, we manually aligned the origin of the reconstructed model with the center of the fiducial marker. We also confirmed that the virtual model correctly overlays over the Pringles can in our OST-HMD.

B. Robot

Our robot is a Kinova MICO 2 arm that is rigidly mounted to the table. The robot is controlled by a standard control scheme using a 2-DOF joystick for end effector control, which cycles between three modes to control the xyz position, angle, and gripper position of the end effector, respectively. The current mode and orientation of all joints is sent to the HMD via WiFi.

	Unoccluded			Occluded		
	N	I	O	N	I	O
Total Number	35	36	35	34	36	35
Number (time)	34	35	34	33	35	34
Number (head path length)	33	35	33	33	35	34
Number (average head deviation)	34	35	34	33	35	34
Number (robot path length)	35	35	35	33	35	34

TABLE I: Number of users evaluated in each condition.

We used a combination of hand-eye calibration and manual tuning to calibrate the robot to the fiducial marker on the table. We verified the alignment by placing the robot in a variety of expected poses and visually confirming that all joints of the virtual HMD model correctly overlay their counterparts of the actual robot.

C. Rendering

We investigated two DR visualization modes. *Invisible* mode is a minimalistic DR representation that aims to remove the robot entirely from the user's awareness (Fig. 1b). This mode overlays the background image onto the robot and renders a virtual representation of the robot's gripper to allow users to grasp occluded objects. However, this diminished view makes it difficult to maintain a mental model of the robot's overall pose. Users in pilot trials requested indicators of the robot's pose, which is especially important when the robot is operated in tight environments or its movement is constrained by its joint limits. We thus developed a second visualization, *outlined*, that augments the illusion of invisibility with a white outline of the robot's shape and semi-transparent overlay of the robot's form (Fig. 1c).

We chose to present the generated virtual content on an OST-HMD because it is failsafe and does not falsify the user's view of the world, making it ideal for everyday deployment. We chose a Microsoft HoloLens because it is untethered and features state-of-the-art visual tracking for re-localization and pose estimation.

The HoloLens can recognize a previously-seen environment and place virtual objects at predefined locations. We use Vuforia to detect a marker placed on the table and save this as the origin of the environment (a "WorldAnchor" object placed at the pose of the detected fiducial marker in Unity). We also localize the fiducial markers on the walls with the HoloLens and store their location relative to the marker on the table. During runtime, we load the stored information and place all objects (background, Pringles can, robot model and outline) in the scene relative to it.

The different components of our system communicate via local WiFi with a latency of approximately 50ms. We found that although this delay is noticeable, it does not significantly affect the overall experience, as users tend to reduce the speed at which they control the robot when they need to perform minor adjustments.

IV. User Study

We designed a 36-person within-subjects user study to investigate the effect of our diminished reality human-robot interface on robot control efficiency, user effort, overall preference, and task performance. Users operated in each of the three visualization modes (*normal*, *invisible*, and *outlined*), across two conditions where the operation target was either *occluded* or *unoccluded*.

A. Hypotheses

H1 (Effort): The *invisible* and *outlined* modes were designed to facilitate teleoperation when the robot occludes the target. We expect that if the target is *occluded*, users will move their heads less and report lower difficulty when the robot is either *invisible* or *outlined* relative to *normal* mode.

H2 (Efficiency): We expect the ability to look through the robot will help users better plan the movement of the robot, resulting in simpler paths and faster operation. We thus hypothesize that in the *occluded* condition, users will complete the task faster and with less robot movement when the robot is either *invisible* or *outlined*. There will be fewer "knockovers" of the can by the robot arm during operation.

H3 (Preference): We anticipate users will prefer modes that give them the most information, ranking their favorite as *outlined*, then *invisible*, and least favoring *normal* mode.

H4 (Relative Performance): We expect that the *occluded* condition will require more user effort and the task will be performed less efficiently than in the *unoccluded* position when users experience the *normal* visualization. By mitigating the issue of occlusion with the *invisible* and *outlined* visualizations, we expect that the effort and efficiency of users in these trials will not be significantly different from *normal* visualization in the *unoccluded* case.

B. Participants

Forty-three participants (17 male, 25 female, 1 Other) were recruited from the general community through an online recruitment tool or word of mouth. The age of participants ranged from 17 to 62 years (M = 26.69, SD = 8.74). All reported normal or corrected-to-normal vision. Six participants had to be excluded for unsafe handling of the robot and/or for being unable to complete the tutorial by halfway through the testing period (45 minutes). These exclusions were based on operation of the robot arm prior to experiencing any of the diminished reality visualizations. Another user was removed due to calibration errors of more than 2cm, which made the illusion untenable. This brought our final participant pool to 36 people. Our research was approved by our institutional review board, and participants were compensated 15 USD.

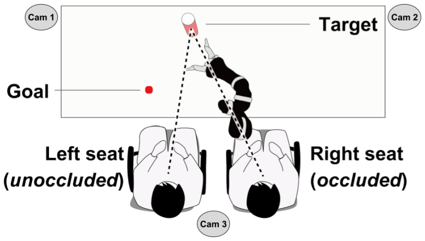


Fig. 2: Top view of experimental setup. Three representative OptiTrack camera locations are shown, additional were used.

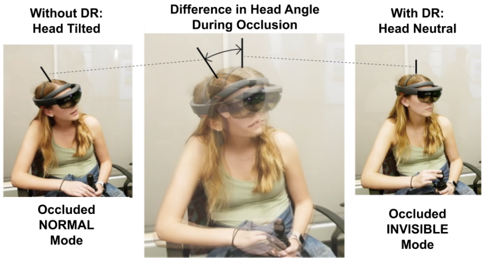


Fig. 3: Example of increased head angle in the *occluded* chair when the robot is *invisible* vs *normal*. Without DR, the user must resort to a non-ergonomic head angle (farther from a neutral position) to gain the information needed for the task.

C. Experimental Procedure

To evaluate the effects of our interface on the user's ability to manipulate the robot, we created a simple pick-and-place telemanipulation task. We asked users to pick up a Pringles can using a joystick-controlled Kinova MICO robot arm, and place it on a target position. The workspace (Fig. 2) was set up so that the robot arm directly occludes the view of the Pringles can from the user when in the right seat (*occluded*), but not in the left seat (*unoccluded*).

The experimental setup consisting of the HoloLens, OptiTrack, and robot arm was set up and calibrated at the beginning of each day. After providing informed consent, each user was led through a calibration of the HoloLens to their specific pupillary distance.

Our study was a 2 (occlusion) X 3 (visualization) within-subjects experiment. Thus, the task of picking up and moving the Pringles can is repeated under six conditions with one trial per condition: *occluded* or *unoccluded* positioning, and either no illusion (*normal*), invisibility (*invisible*), or invisibility with the robot's outline (*outlined*). We used a balanced Latin Squares design to order these six trials to compensate for the learning curve of participants.

As mentioned in Sec. III-B, users operated the robot arm by cycling through three control modes with the joystick. Though this is a potentially taxing way to use the robot [12], we chose this control method because it is typical for this commercial robot. Due to the sharp learning curve of operating the robot arm, users were given a tutorial lasting up to 45 minutes before the experimental trials. In the tutorial, they received coaching on how to operate the robot arm using the joystick. The tutorial concluded when users successfully moved the Pringles can from the start position to the end position by themselves in the *unoccluded* condition.

User head positions and the locations of all objects in the scene were recorded through the HoloLens. During trials, users were not allowed to talk with facilitators in order to avoid unnecessary head movements. Robot joint positions were also logged. After each trial, users were asked about task difficulty and completed the six subjective subscales of the NASA-TLX survey [10]. After all trials were completed, users were given a final survey asking them to rank the visualizations in order of preference and difficulty and to identify the most difficult aspect of using each visualization. The experiment lasted up to 90 minutes.

D. Metrics

For each hypothesis, we quantify user performance and experience for every combination of conditions as follows:

H1: EFFORT A standard response to occlusions is for the user to change their viewpoint. However, this approach requires extra energy and inconvenience, which is undesirable and may not be tenable for some users. Quantifying effort is therefore important in assessing DR techniques. To do this, we tracked the head angle and location of users, and calculated both total and average path length of head movements during the task. We also calculated average head offset from the user's start position during the task, to assess the average deviation from an ergonomically neutral position.

We used the unweighted/raw version of the NASA-Task Load Index (NASA-TLX) to assess the relative workload of the task, and asked users how "cumbersome", "hard to use", and "difficult to learn" the trial was on 7-pt Likert scales. We asked users to rate their "confidence" on the same scale.

H2: EFFICIENCY To understand user efficiency, we measured how long it took users to perform the task overall and the length of the path taken by the robot. We also measured the number of times users unintentionally knocked over the can when attempting to perform the task.

H3: PREFERENCE To capture direct user feedback, we asked users to rank the modes in order of preference in each condition (*occluded* and *unoccluded*) and to provide feedback on their experience after each trial. User preferences likely encompass aspects of both efficiency and physical effort, as well as additional aspects such as cognitive load or intuitiveness. To further clarify and support other quantitative findings, we asked users to identify the most difficult aspect of using each mode in a freeform response.

H4: RELATIVE PERFORMANCE Our goal in providing this interface was to enable users to perform the task in the *occluded* condition with similar effort and efficiency as the *unoccluded* case, and perform no worse. Therefore, in addition to comparing performance between the three visualizations, we also compare performance across the *occluded* and *unoccluded* conditions.

V. Results

We segmented the grasping task into subsections to focus our analysis on the portion of the task most affected by occlusion, which occurred when the user attempted to approach and grasp the can. A trial began (time = 0) with the first movement of the arm. The user approached the can with the arm, until the grippers first began to open and grasping began. The grasping task was completed when the user began the xyz movement of lifting the can. This segmentation provided us a greater focus on the occluded task and a clearly defined end point. We considered using proximity to the end target as our end point, but participants often performed a highly variable series of unoccluded fine adjustments during the "touchdown" period of the task with

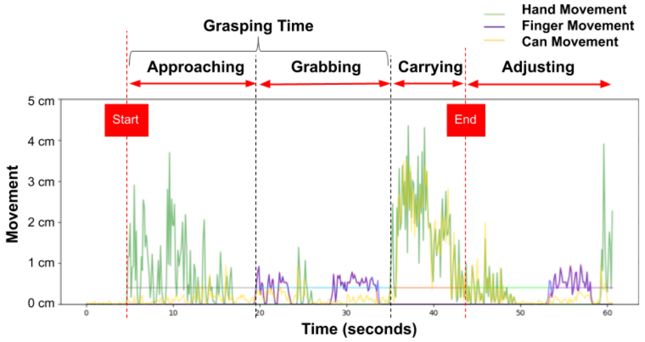


Fig. 4: Phases of the grasping task. Each segment was automatically labeled based on the robot arm's motion.

a greater potential for getting into complex or over-extended manipulator configurations during this period. This set of delineations can be seen on a sample user's data in Fig. 4.

A. Data

In reviewing footage of user views, we discovered that five users had a HoloLens menu open during one trial each (randomly distributed). We excluded these trials from consideration. We also removed trials that were more than three standard deviations from the mean: six for grasping time, seven for robot path length, eight for average head path length and six for average head deviation. The final number of trials per condition is shown in Table. I, with a maximum exclusion rate of 6.0%. In our analysis we assumed statistical significance if the results showed that $p < 0.05$.

H1: EFFORT We compared average head path length in each visual mode in the *occluded* case using an ANOVA and found no significant difference between visual modes ($F(1.7, 56.7) = 1.23,\ p = n.s.$). Additionally, we calculated the average head deviation from its starting position to understand how much users moved their heads during the task. The results of an ANOVA showed visual mode significantly affected the average offset ($F(1.72, 60.36) = 4.89$, $p < 0.05$). A post-hoc test using the Holm-Bonferroni method revealed that on average users move their head in *normal* mode more than in *invisible* mode and *outlined* mode ($t(35) = 3.15$, $p < 0.01$; $t(34) = 3.08,\ p < 0.01$), as seen in Fig. 5. This trend can also be observed in the heatmap of user head movements in the X-Y plane shown in Fig. 6. This data supports our hypothesis that the *invisible* and *outlined* modes reduce the head movement necessary to perform the manipulation task when the robot occluded the scene.

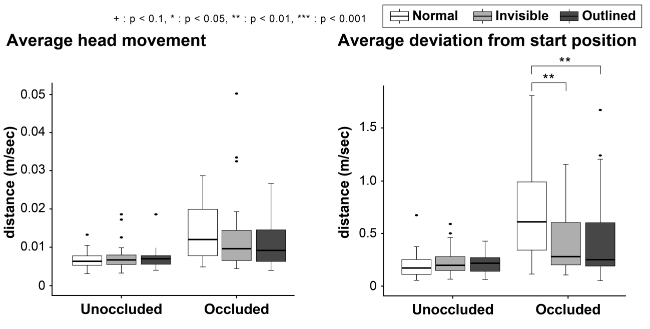


Fig. 5: Average head path length and average deviation from starting position in each mode.

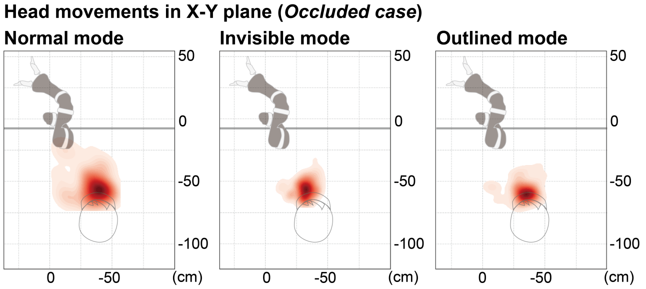


Fig. 6: Heat map of user head locations in the X-Y plane, shown relative to the robot and chair.

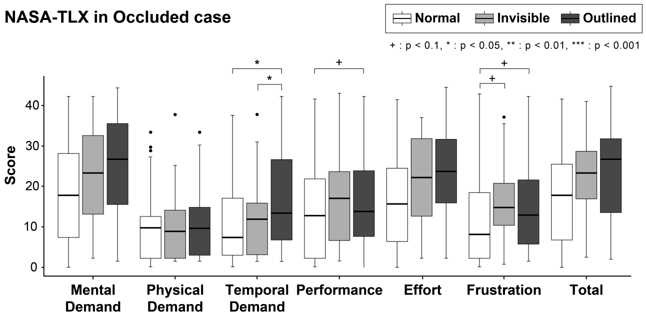


Fig. 7: NASA-TLX Survey for the *occluded* condition.

We conducted the Friedman test to compare NASA-TLX subscales and workload scores (shown in Fig. 7) across the three modes in the *occluded* case. We found significant differences in temporal demand ($\chi^2(2) = 11.25$, $p < 0.01$) and frustration ($\chi^2(2) = 6.64,\ p < 0.05$). Post-hoc analysis with the Wilcoxon signed-rank test revealed the score of temporal demand in *outlined* mode was higher than both *normal* mode and *invisible* mode ($p < 0.01$; $p < 0.05$). Frustration did not show significant differences between any pair of visual modes, but the difference between *normal* mode versus *invisible* mode and *normal* mode versus *outlined* mode did approach significance. Additionally, the difference between reported performance in the *normal* mode and *outlined* mode pair approached significance.

We also investigated differences in user perception of how cumbersome, hard to use, and difficult to learn each visual mode was in the *occluded* case, as well as how confident users felt while controlling the robot in each mode. The different visual modes did not significantly affect these perceptions, as seen in Fig. 8.

H2: EFFICIENCY We show the total task time and total robot path in Fig. 9. Results of a one-way ANOVA in the *occluded* case show the visual mode significantly affected grasping time ($F(2, 70) = 5.34$, $p < 0.01$). A post-hoc test using the Holm-Bonferroni method found users in

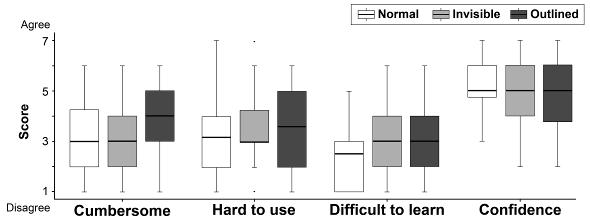


Fig. 8: User ratings after each trial for how ”cumbersome” the mode felt, how ”hard to use” and ”difficult to learn” it was, and how ”confident using the robot” they were.

normal mode completed the task faster than in *outlined* mode ($t(35) = 3.25, p < 0.01$). We found no significant difference in robot path length across visual modes in the *occluded* case ($F(2,70) = 1.42,\ p = n.s.$). We found no significant difference in the rate of can knockovers across modes.

H3: PREFERENCE Overall user preference and perceived difficulty based on a rank ordering are shown in Fig. 10. We compare the reported ranking of the different modes in each condition with the Bradley-Terry model. The results show that in the *unoccluded* condition users preferred using *normal* mode compared to *invisible* ($z = -3.8$, $p < 0.001$) and to *outlined* ($z = -4.817$, $p < 0.001$). The test did not reveal any difference in preference between *invisible* and *outlined*. We also did not find any mode to be preferred in the *occluded* condition.

When we compared the ranked difficulties of the different modes in each condition we found that the *normal* mode was perceived as less difficult than *invisible* ($z = -2.19$, $p = 0.0285$) and *outlined* ($z = -4.348$, $p < 0.001$). We also found that *outlined* was perceived to be easier than *invisible* ($z = 5.566$, $p < 0.001$). We did not find any significant differences in the difficulty rankings for the *occluded* case.

H4: RELATIVE PERFORMANCE When analyzing relative performance across the *occluded* and *unoccluded* conditions with each of the different visualizations using two-way ANOVA, we found position and visual mode did not have any interaction effect on grasping time ($F(2,70) = 2.08$, $p = n.s.$), as seen in Fig, 9. As mentioned in Sec. V-A, the main effects of position and visual mode were significant ($F(1,35) = 37.28,\ p < 0.001;\ F(2,70) = 5.39,\ p < 0.001$). A post-hoc test using the Holm-Bonferroni method revealed users in *normal* mode completed the task faster than users in *outlined*, ($t(35) = 3.25,\ p < 0.01$).

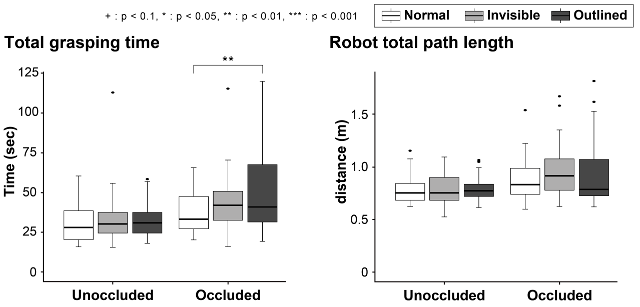


Fig. 9: Total task time (left) and total robot path length (right) during grasping.

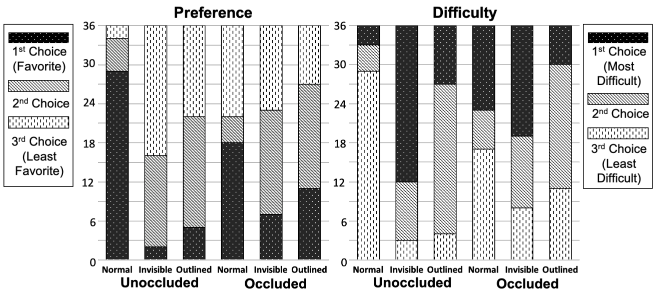


Fig. 10: Ranking of difficulty and preference in *occluded* and *unoccluded* positions. Darker shades indicate higher preference and greater difficulty. Note clear preference and ease of use for *normal* mode in the *unoccluded* case.

We also found no interaction effect for robot path length ($F(2,70) = 0.62$, $p = n.s.$). The main effects of position were significant ($F(1,35) = 16.17,\ p < 0.001$). See Fig. 9.

B. Qualitative Feedback

Table II presents affinity diagramming feedback about the most difficult aspect of using each visualization mode. Depth perception was consistently labeled as a point of difficulty even in the absence of a DR visualization. Issues in *invisible* mode tended to focus on smaller errors that caused users to be unable to fully rely on the illusion. In *outlined* mode, the outline actually distracted from users’ view or focus.

VI. Discussion

Our results support hypothesis **H1** in terms of physical effort, but fail to support it in terms of mental effort. Though users kept their heads on average closer to a neutral position in the *invisible* and *outlined* modes, users also perceived these modes as more demanding and frustrating. This suggests that our interface did provide an improved understanding of the occluded space and therefore helped users avoid having to change viewpoints to complete the task, but this new visualization added complexity for the user.

Our results contradicted **H2** for a few potential reasons. First, we did not require users to keep their viewpoint static. As shown in Fig. 6, participants changed their viewpoint in *normal* mode to simplify the task and acquire the information required for efficient pathing. Similar robot movement paths across modes could also arise from participants figuring out a suitable path during early trials. It would be interesting to explore a scenario with an initially unoccluded target where movement of the robot would result in occlusion.

Our results did not support **H3**. When the robot was not occluding the target area, participants preferred using the normal mode and judged it to be simpler than other modes. In this condition, the *invisible* and *outlined* modes did not provide additional useful information to the user and the added graphics confused the users rather than helping them complete the task. On the other hand, we did not find a preference for any of the modes in the occluded condition. Participants also perceived the task to be similarly

None	#	Invisible Mode	#	Outline Mode	#
Arm occluding task	11	Arm proprioception	7	Outline cluttering/distracting	9
Grasping occlusion	9	Confused by visualization, Illusion errors: imprecision	5	Outline obstructing	8
Depth perception	8	Depth perception, Establishing grip, Finding correct approach angle	4	Depth perception	6
Establishing grip	5	Illusion errors: grasping, Illusion errors: flatness, Illusion errors: Lag	3	No problems, Illusion: Double vision, Positive feedback	4
Control mechanisms, No problems	2	No problems	2	Illusion: imprecision	3
Wanted to see target in AR	1	AR learning curve, Not fully invisible, Need visualization of target	1	Illusion: lag, Making grip	2
				Not fully invisible, Learning curve, Attack angle	1

TABLE II: Frequency of participants' comments on the most difficult part of using each visualization mode. Some users provided multiple categories of feedback, so totals do not sum to 36.

difficult to perform in all modes. This result is encouraging in that for the occluded case, the visualization modes were providing enough utility to compete with the reflex to change viewpoints. A more complicated scenario without the option of resolution via head movement might show further value.

Our results also did not support **H4**. We found that *normal* mode in fact had the best performance as far as task time in both positions. This was most likely because once the occlusion issue has been surmounted by a change of viewpoint, this condition does not require additional mental load for learning how to use a new tool. Furthermore, *normal* mode does not remove information the user may want, as in the *invisible-unoccluded* case, nor add potentially distracting additional elements, as in the *outlined-occluded* case.

A. Experimental Design Limitations

In assessing the pragmatics of this scenario, we presume that the AR headset is tolerable for long periods of time. Current AR hardware can sometimes induce nausea or simply discomfort from the weight of the headset [16], though our users did not report experiencing these. We asked users to wear the headset in every trial, even the mode with no DR visualization, so that user perception responses would not be affected by hardware. Our balanced Latin square design should also mitigate the issue of fatigue affecting our results.

We assumed in several of our key metrics that users would feel comfortable moving their heads over the course of the experiment. To enable this, we encouraged users to move their heads around during the calibration period and adjust it to their satisfaction. When participants asked about moving, we told participants they could do whatever they wanted as long as they sat in the correct chair. However, despite being asked to tilt and move their heads during the calibration process, some participants noted after the study was complete that they thought they ought to keep their heads still during the study due to wearing the HMD. Informal comments about their hesitation indicated users did not want to break the device or interfere with its effectiveness. When asked how they completed the task in the occluded no-visualization with no head movement, users indicated that they executed the grasping task from memory or educated guesswork.

B. Considerations for DR Interfaces

As noted in Table II, details are important when implementing a DR interface. Small errors in precision and lag can disrupt the illusion for users. If the user cannot trust the illusion or is misled into creating errors, the experiment provides us with feedback not on the potential of DR for addressing manipulation-occlusion problems, but instead on user frustration with illusion inaccuracy.

Specific to the DR illusions, users indicated they found the visualizations somewhat "flat" or "missing depth cues". This is likely due to the fact that while background images were "projected" in space to the correct locations, they all had full brightness and none of the shadowing expected of a physical object. Possible solutions are capturing ambient lighting or providing uniform lighting. Additional overlays on the arm would not address this issue because these projections would also be perfectly well-lit and flat unless a specific shadowing scheme was added. Given the importance of depth perception in all modes, as well as the fact that knockovers occurred equally in all three conditions, explicit indicators of relative distances between objects or depth would be a valuable addition to a system of this kind. Additional visualizations also may increase the risk of users noticing minor errors in the DR illusion because the edges of the outline make it easier to identify misalignments (see Table II).

User feedback indicated that issue of "flatness" might be partly due to the HMD's design. First, the single focal plane that did not match the focal distance of the robot or the background effectively forced users to continuously refocus between the real and the virtual content. Furthermore, although we tried to adjust the lighting of the virtual environment to match the environment, it did not perfectly replicate the real world. Another concern is the tracking accuracy of the HMD. Although we carefully aligned the virtual content with the real world, the spatial tracking of the system produced small misalignments. These imprecisions could have affected users' confidence in the illusion. This problem could be overcome by a video see-through HMD or more sophisticated future hardware.

C. Design Tradeoffs

Perceived Difficulty vs. Physical Difficulty Despite similar task performance in terms of path length (Fig. 5), total time (Fig. 9), and number of knockovers, users found the *invisible* mode more difficult than either the *normal* mode or the *outlined* mode in the *unoccluded* condition (Fig. 10). Because this increase in difficulty was not reflected in the physical measures of the task, it likely reflects a larger cognitive load for the user. To explain this, we observe that the illusion of invisibility does not contribute to task completion in the *unoccluded* case, but erasing the pose of the robot does require the user to put in more mental effort to remember the location of the arm. This was supported by comments on the *invisible* mode: participants said it was "disorienting that [they] couldn't see the robot arm at all, although [they] could see the Pringles can", "hard to understand where the arm was", and "get[ing] the background confused with what I was seeing through the lenses... I didn't know where to look." Therefore, there are scenarios or specific users who did want the additional pose information provided by the *outlined* mode or the simplicity of the *normal* mode.

Information vs. Saliency At the same time, there is a tradeoff in presenting task-relevant information without overloading the user. In the *occluded* case, users did not express as clear of a preference for the *normal* mode (Fig. 10) as in the *unoccluded* case. In the *outlined* visualization, visual clutter may add to task completion time, a possibility supported by comments (Table II) such as "seeing too many lines, overwhelming", "the outline added more visual clutter", and "watching out for both... the arm and object and the overlap was confusing". It is important to strike a balance between too little and too much information in DR interfaces. It may be useful for visualizations to be switched on and off directly by users when needed or for the system's visuals to dynamically adjust to the situation at hand. Given our findings, it is likely that an increased number of objects in a scene with transparency illusions would need careful handling to remain clear and concise for the user.

D. Future Work

A longer study with a greater number of repeated tasks might increase user trust and efficiency using novel DR visualizations. While our study focused on assessing individual DR visualizations, allowing the user to manually toggle modes as needed and/or a system where modes are dynamically engaged on their behalf could potentially enhance value for users. Finally, it would be valuable to conduct this study with users who have upper body motor impairments or who use this robot on a daily basis.

References

[1] J. Allspaw, J. Roche, N. Lemiesz, M. Yannuzzi, and H. A. Yanco. Remotely teleoperating a humanoid robot to perform fine motor tasks with virtual reality–. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[2] R. M. Aronson, T. Santini, T. C. Kübler, E. Kasneci, S. Srinivasa, and H. Admoni. Eye-hand behavior in human-robot shared manipulation. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 4–13. ACM, 2018.

[3] B. Aruanno, F. Garzotto, and M. C. Rodriguez. Hololens-based mixed reality experiences for subjects with alzheimer's disease. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, CHItaly '17, pages 15:1–15:9, New York, NY, USA, 2017. ACM.

[4] B. Avery, B. H. Thomas, and W. Piekarski. User evaluation of see-through vision for mobile outdoor augmented reality. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 69–72. IEEE, 2008.

[5] V. Buchmann, T. Nilsen, and M. Billinghurst. Interaction with partially transparent hands and objects. In *Proceedings of the Sixth Australasian Conference on User Interface - Volume 40*, AUIC '05, page 17–20, AUS, 2005. Australian Computer Society, Inc.

[6] J. R. Cauchard, A. Tamkin, C. Y. Wang, L. Vink, M. Park, T. Fang, and J. A. Landay. Drone. io: A gestural and visual interface for human-drone interaction. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 153–162. IEEE, 2019.

[7] C. W. Cheung, I. Tsang, and K. H. Wong. Robot avatar: A virtual tourism robot for people with disabilities. *International Journal of Computer Theory And Engineering, Singapore*, 9(3):229–234, 2017.

[8] R. K. Ganesan, Y. K. Rathore, H. M. Ross, and H. B. Amor. Better teaming through visual cues: how projecting imagery in a workspace can improve human-robot collaboration. *IEEE Robotics & Automation Magazine*, 25(2):59–71, 2018.

[9] P. M. Grice, A. Lee, H. Evans, and C. C. Kemp. The wouse: A wearable wince detector to stop assistive robots. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 165–172. IEEE, 2012.

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

[11] H. Hedayati, M. Walker, and D. Szafir. Improving collocated robot teleoperation with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 78–86. ACM, 2018.

[12] L. V. Herlant, R. M. Holladay, and S. S. Srinivasa. Assistive teleoperation of robot arms via automatic time-optimal mode switching. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 35–42, 2016.

[13] M. Inami, N. Kawakami, and S. Tachi. Optical Camouflage Using Retro-Reflective Projection Technology. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 348–349, 2003.

[14] N. Inc. Optitrack. `https://optitrack.com/`. Last accessed: 2020-3-1.

[15] T. Kot, P. Novák, and J. Bajak. Using hololens to create a virtual operator station for mobile robots. In *2018 19th International Carpathian Control Conference (ICCC)*, pages 422–427. IEEE, 2018.

[16] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1):47–56, 2000.

[17] H. Liu, Y. Zhang, W. Si, X. Xie, Y. Zhu, and S.-C. Zhu. Interactive Robot Knowledge Patching Using Augmented Reality. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1947–1954. IEEE, 2018.

[18] P. Milgram, S. Zhai, D. Drascic, and J. Grodski. Applications of Augmented Reality for Human-Robot Communication. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 1467–1472, 1993.

[19] S. Mori, M. Maezawa, and H. Saito. A Work Area Visualization by Multi-View Camera-based Diminished Reality. *Multimodal Technologies and Interaction*, 1(3):18, 2017.

[20] C. Munroe, Y. Meng, H. Yanco, and M. Begum. Augmented reality eyeglasses for promoting home-based rehabilitation for children with cerebral palsy. In *The eleventh ACM/IEEE international conference on human robot interaction*, pages 565–565. IEEE Press, 2016.

[21] D. Rakita, B. Mutlu, and M. Gleicher. An autonomous dynamic camera method for effective remote teleoperation. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 325–333. ACM, 2018.

[22] C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila. An Augmented Reality X-Ray System Based on Visual Saliency. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pages 27–36, 2010.

[23] J. Shaw and K. Cheng. Object identification and 3-d position calculation using eye-in-hand single camera for robot gripper. In *2016 IEEE International Conference on Industrial Technology (ICIT)*, pages 1622–1625. IEEE, 2016.

[24] K. Sugimoto, H. Fujii, A. Yamashita, and H. Asama. Half-Diminished Reality Image Using Three RGB-D Sensors for Remote Control Robots. In *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 1–6, 2014.

[25] M. Walker, H. Hedayati, J. Lee, and D. Szafir. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 316–324. ACM, 2018.

[26] M. E. Walker, H. Hedayati, and D. Szafir. Robot teleoperation with augmented reality virtual surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 202–210. IEEE, 2019.

[27] T. Weng, L. Perlmutter, S. Nikolaidis, S. Srinivasa, and M. Cakmak. Robot object referencing through legible situated projections. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8004–8010. IEEE, 2019.