Robot Teleoperation with Augmented Reality Virtual Surrogates*

Michael E. Walker
Dept. of Computer Science
University of Colorado Boulder
Boulder, USA
michael.walker-1@colorado.edu

Hooman Hedayati
Dept. of Computer Science
University of Colorado Boulder
Boulder, USA
hooman.hedayati@colorado.edu

Daniel Szafir
Dept. of Computer Science & ATLAS Institute
University of Colorado Boulder
Boulder, USA
daniel.szafir@colorado.edu

***Abstract*—Teleoperation remains a dominant control paradigm for human interaction with robotic systems. However, teleoperation can be quite challenging, especially for novice users. Even experienced users may face difficulties or inefficiencies when operating a robot with unfamiliar and/or complex dynamics, such as industrial manipulators or aerial robots, as teleoperation forces users to focus on low-level aspects of robot control, rather than higher level goals regarding task completion, data analysis, and problem solving. We explore how advances in augmented reality (AR) may enable the design of novel teleoperation interfaces that increase operation effectiveness, support the user in conducting concurrent work, and decrease stress. Our key insight is that AR may be used in conjunction with prior work on predictive graphical interfaces such that a teleoperator controls a *virtual robot surrogate*, rather than directly operating the robot itself, providing the user with foresight regarding where the physical robot will end up and how it will get there. We present the design of two AR interfaces using such a surrogate: one focused on real-time control and one inspired by waypoint delegation. We compare these designs against a baseline teleoperation system in a laboratory experiment in which novice and expert users piloted an aerial robot to inspect an environment and analyze data. Our results revealed that the augmented reality prototypes provided several objective and subjective improvements, demonstrating the promise of leveraging AR to improve human-robot interactions.**

***Index Terms*—Robots, teleoperation; augmented reality; mixed reality; ARHMD; interface design; aerial robots; drones;**

I. Introduction

Teleoperation, in which a user manually controls a robot, represents a well known and popular paradigm in the field of robotics. Although advances in robot perception are increasing autonomous capabilities, teleoperation provides the benefit of tightly coupling user input with robot actions and is still used in wide variety of domains, including surgical robots [1]–[3], robotic manipulators in space exploration [4], and aerial robots for disaster response [5].

However, teleoperaton represents a challenging task, often requiring a high degree of operator focus and skill. Prior work has identified a number of human performance issues that may arise when utilizing teleoperation interfaces or mixed teleoperation/supervisory control systems (see [6] and [7] for reviews). For example, researchers have noted difficulties in distance estimation, understanding robot orientation and attitude, predicting robot motion, and accomplishing any concurrent task work. We believe a number of these issues can be explained by the concept of a "Gulf of Execution," [8] a common concept in the cognitive engineering and human-computer interaction (HCI) communities that details how users experience difficulties interacting with complex systems when they are unable to translate high level goals (i.e., working with a robot to achieve some task) into inputs the system can understand (i.e., low-level teleoperation commands). This "gulf" represents one part of the *Human Action Cycle*, a proposed model describing human interaction with complex systems that would suggest robot teleoperators complete a predictable sequence of (1) forming mental goals, (2) specifying and executing intentions and actions, (3) perceiving how their actions altered the state of the system, and (4) evaluating whether the outcome has achieved the original goal (or brought the goals closer). Operators might then repeat the sequence until all goals are achieved, or until the user simply gives up [9]. This model highlights three potential issues in existing teleoperation paradigms: (1) it may be difficult for operators to understand what commands might achieve their goals, (2) operators may have difficulties predicting how a given command will affect system state, and (3) operators must actually move the robot to complete the evaluation step, which may lead to unexpected and/or dangerous situations for the robot or nearby users. Each of these issues arise because teleoperation hides the mapping between operator input and robot dynamics from the user. This mapping is implicitly encoded into the system and requires users to learn it indirectly through experience (e.g., learn the relationship between joystick angle and motor torque).

In this work, we are interested in designing new interfaces that can help operators better bridge this gulf. In particular, we aim to provide more direct feedback on this mapping in an intuitive, visual manner to enable users to better predict how their actions will affect a robotic system. We provide this feedback by leveraging advances in see-through augmented reality head-mounted display (ARHMD) technology. Inspired by past work examining interaction with virtual surrogates of physical objects (e.g., toy cars, planes, etc.) [10], our key insight is that we can use modern ARHMDs to provide users with an immersive *virtual surrogate robot*—virtual imagery in

*Link to the paper's supplementary video: https://youtu.be/7vy8krhT3Ys

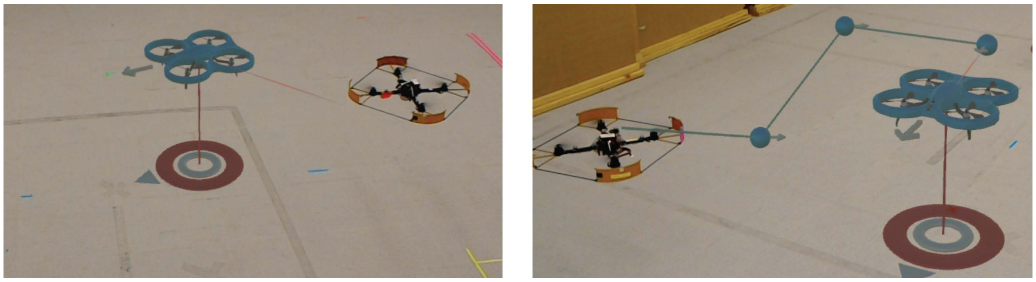

Fig. 1. Two AR teleoperation interfaces designed in this work: Left - Realtime Virtual Surrogate (RVS), Right - Waypoint Virtual Surrogate (WVS).

the form of a "ghost" of the real robot that is embedded within the same operational environment, with accurate stereoscopic depth cues and a matching dynamics model, but that cannot physically interact with the environment (i.e., cannot be damaged or present a hazard to other physical objects or users). This surrogate can enable teleoperators to more rapidly iterate through goal/action/evaluation phases without the potential of an action that leads to a negative consequence that is realized only after evaluation (e.g., a robot colliding with an obstacle or person). Although certain planning systems from prior work have explored related ideas (e.g., waypoint delegation systems that provide a virtual model of the robot to indicate desired position/attitude at each waypoint [7]), such systems typically rely on interaction mediated via a 2D screen (e.g., tablet or laptop), which removes the control system from the context of the physical environment and only provides limited monocular cues for depth estimation.

Through our exploration of virtual surrogate robots, we hope to contribute to answering a higher level research question: "how might augmented reality improve collocated user teleoperation?" We believe that answering this question may have a positive impact for many fields that currently rely on collocated teleoperation (e.g. disaster response, environment or warehouse inspection, space exploration, etc.). Below, we outline relevant work that informed our design process and present a framework for designing augmented reality feedback to enhance human-robot interactions. We then describe the development of two new interfaces that leverage ARHMDs to provide novel ways of teleoperating a robot by directing its virtual surrogate. We evaluate these designs using a within-participants laboratory experiment with 18 participants, comprising both novice and expert users. We conclude with a discussion of how AR technology represents a promising new tool for designing robot control interfaces.

II. Background

In this research, we explore the design of augmented reality interfaces that enhance collocated robot teleoperation by enabling users to manipulate an immersive virtual robot surrogate that foreshadows the actions a physical robot will take. Below, we review relevant control interfaces and provide a brief primer on augmented reality technologies.

Teleoperation interfaces for robots often require a great deal of skill to manually control a robot. This problem becomes increasingly challenging as robots increase in degrees of freedom or dynamic complexity of their motor capabilities. As an example of the skill required for an aerial robot with 6 degrees of freedom, the Federal Aviation Administration (FAA) in the United States considered requiring commercial aerial robot operators to obtain a pilot's license [11] and still today requires such robots be operated only within line-of-sight range.

Most current human-in-the-loop robot control interfaces take one of two forms: direct teleoperation interfaces, where robots are manually controlled by a user via joysticks (or similar device) with a video display, or a higher-level supervisory interface that allows users to plot waypoints or similar temporal/spatial indicators that outline a desired robot path or actions [7]. Researchers have explored a variety of methods to improve user effectiveness within these two major paradigms, including providing haptic feedback [12], mediating direct user input to achieve shared control and "safeguarded" motion [3], [13], [14], or improving precision via "nudging" as part of "perceived first-order control" [15]. Other work offers hybrid interfaces, often by combining a live video display with virtual map data and providing support for mixing teleoperation with forms of autonomous waypoint navigation [16]–[18]. Recent systems in aerial robot photography and cinematography take advantage of increasing autonomous capabilities, enabling users to focus more on their task, rather than piloting the robot [19]–[22]. However, traditionally all of these systems present the control interface to users via traditional 2D displays (e.g., mobile devices, tablets, or laptop base stations).

Alternatively, researchers are now beginning to explore how augmented reality (AR) might support human-robot interactions (see [23] for a survey). Augmented reality (AR) technologies overlay virtual graphics onto real environments [23], [24]. AR interfaces may take different forms, such as 2D screens (such as a smartphone or tablet) or see-through ARHMDs, such as the HoloLens and Meta 2. For example, Kasahara et al. implemented a novel embodied spatially-aware interface to manipulate actuated objects in which the user can move an object using AR provided through an iPad [25]. Hashimoto et al. proposed a touch-based interface for controlling aspects of robot actuation [26]. Both solutions have similarities in that they enable a user to manipulate a virtual representation of the robot leading to actions by the physical robot. However, both systems rely on a "window-on-the-world" paradigm, where users interact with a touchscreen that provides 2D AR feedback and affords input to control the robot, as opposed to more immersive, 3D hands-free augmented reality via head-mounted displays.

Robot interfaces using ARHMDs represent a different type of interaction than either traditional 2D displays, static/remote AR displays, or AR displays on mobile devices/tablets, as ARHMD designs provide stereoscopic depth, do not require gaze shifts between the device and the real world/robot [27], enable free user movement in the environment, and can support hands-free operation. As a result, ARHMD interfaces can easily be integrated with existing control systems that users may already be familiar with (e.g., traditional teleoperation via joysticks) in a manner that tablet-based systems cannot and may allow the preservation of teleoperation precision.

More recently, research has examined how ARHMDs might support human-robot interaction, building off of prior systems that demonstrated how AR might overlay robot planning and world model data on top of a user's field-of-view [28], [29]. One of our previous studies demonstrated that ARHMD feedback of aerial robot flight intent could significantly improve task efficiency for users working alongside a robot [30]. Additionally, our prior work presented one of the first studies to investigate the use of modern ARHMD technologies for robot teleoperation, developing an immersive system for visualizing robot-collected data and camera feeds [27]. Our work here builds on such prior research to further explore the potential of AR for improving human-robot interactions.

III. System Design and Implementation

A. Design

In this research, we sought to explore how AR might improve robot teleoperation by providing users with an immersive virtual robot surrogate to control, rather than directly controlling a physical robot. We hypothesize that controlling a surrogate might enable users to better predict how their actions will affect the system and offer foresight into the eventual pose and position of the physical robot as it mimics the actions taken by the surrogate. Although we believe such a system might be beneficial for operating a wide variety of robots (e.g., manipulators, underwater robots, etc.), in this work we specifically examined interfaces for aerial robot teleoperation. Although prior work has implemented interface systems using a surrogate-style concept (input arbitration, control proxies, etc.) [26], [31]–[36] we wish to extend this research into the domain of realtime ARHMDs interfaces. This introduces different design challenges and opportunities compared with other AR technologies, such as leveraging its 3D capabilities by integrating depth cues or taking advantage of its hands-free nature and integrating within preexisting control paradigms.

Robot actions might be tied to actions of a virtual surrogate in a variety of ways. For instance, mappings between the physical and virtual robot might range from instantaneous duplication (which in effect would exactly match standard teleoperation), to delayed duplication (the physical robot moves to match the virtual surrogate position/attitude after some period of time), to confirmed duplication (the physical robot matches the virtual surrogate when triggered by the user), to more planning-oriented systems (e.g., using the surrogate to denote waypoints for future execution by the physical robot).

To guide our design decisions, we explored two control paradigms: (1) real-time operation where the robot responds instantaneously to user input (matching standard forms of teleoperation) and (2) a delayed form of control that leverages AR's ability to place virtual information and objects within the user's environment (see Figure 1).

Realtime Virtual Surrogate (RVS): This design presents users with a virtual aerial robot that shares a physical environment with a physically-embodied aerial robot. The virtual surrogate's appearance is modeled after the physical robot and a virtual "fishing line" connects the surrogate to the physical robot. In addition, a virtual line is rendered directly below the robot connecting it to a double ring visualization laid flat on the ground that serves as a depth indicator (mimicking drop shadow depth cues, which have shown effectiveness in communicating depth in AR applications [37]).

In this design, user teleoperation commands are intercepted and directed to the virtual surrogate, rather than directly controlling the physical robot as in traditional teleoperation. The virtual surrogate is used as a setpoint or goal state for a planning algorithm running on the physical robot that causes the physical robot to constantly "chase" the surrogate, stopping only when the virtual and physical robot share the same position. Any desired planning algorithm might be used, although for our implementation we opted for a simple PID controller. This design offers several tunable parameters, including the control speed, "chase speed," and delay time of the virtual robot and potentially any additional constraints to be enforced by the planner (e.g., motion smoothness). However, from an implementation standpoint this system only requires that the surrogate's 6DOF pose be converted from the augmented reality coordinate system that displays the surrogate to the planning coordinate system as a desired goal state for the onboard robot autonomy. The details of this implementation and conversion are provided in §III-B. We hypothesize that this design will help decrease the gulf of execution by allowing users to preview and better understand how a physical robot might react to user-issued commands in real-time, giving the user a chance to detect, evaluate, and fix erroneous commands before the physical robot actually executes them.

Waypoint Virtual Surrogate (WVS): This design extends the RVS model to provide better support for longer term planning, potentially at the cost of immediate precision control. As in the RVS model, the teleoperator controls a virtual robot surrogate rather than directly operating a physical robot. However, in the WVS paradigm, the physical robot stays in place while the user operates the virtual surrogate to create a plan by adding/deleting/editing virtual waypoints. At any point, the user can signal to the physical robot to start executing the planned path, which is defined by a series of 6DOF goal poses specified by the waypoints. The user can edit waypoint positions, delete any planned waypoints, and add additional waypoints on-the-fly. This interface is inspired by recent work in higher-level aerial robot interfaces (e.g., [7], [21]) and allows us to examine trade-offs between the RVS system and a more supervisory control scheme.

B. Implementation

In addition to these two designs, we also implemented a baseline teleoperation system that enables users to directly pilot the physical robot rather than control a virtual surrogate. This system was based on modern aerial robot teleoperation interfaces, requiring the user to use joysticks to move the robot in 3D space, although in the absence of user input the robot autonomously maintains a stable hover.

Coordinator: Our designs are built on top of a backend coordination system that we developed as a custom application within the Unity game engine. The basis of this system is a *virtual aerial robot* object within the Unity application, which is run on a Microsoft HoloLens. Our application translates user input from a physical, wireless, handheld controller (Xbox) to a segmented list of desired poses for the virtual robot, which then navigates the scene per the specifications of the waypoint list. For every iteration of the application engine's update loop, we send the virtual robot's 3D pose from the HoloLens to an onboard system controlling the physical robot via UDP. Virtual robot pose values are transformed from Unity coordinates to real-world coordinates using a transformation matrix calculated *a priori* in an initial calibration procedure that calculates the relative origins and basis vectors for each coordinate system. After this transformation, pose values in the Unity scene correspond to the same positions in the user's real environment, enabling the physical robot to fly through the study space in a manner that matches the virtual drone.

Controller: For the robot's position to correspond with the virtual imagery, a PID controller is continuously receiving 3D goal positions from the ARHMD application that the PID controller then navigates the robot to. The PID controller runs at 20Hz and precisely controls the altitude, position, and orientation of the aerial robot. The robotic platform utilized in this study lacks onboard sensing capabilities sufficient for accurate localization, thus the PID controller currently uses motion tracking cameras embedded in the environment to track the physical robot. User control inputs were designed to match the default AscTec control scheme that is representative of modern teleoperation systems (Figure 2-D). The handheld controller sensitivity was calibrated to make it as similar as possible to the commercial teleoperation system and kept constant across all interface designs.

IV. Experiment

We conducted a 3×1 within-participants experiment to evaluate, relative to a baseline, how our RVS and WVS designs might affect user experiences when teleoperating a collocated flying robot. The study protocol had participants teleoperate a physical quadcopter through a laboratory environment that simulates the real-time collection and analysis of data. The independent variable in this study corresponds to what type of teleoperation interface the participant used: a baseine teleoperation system in which the handheld controller input directly controlled the physical robot, the realtime virtual surrogate design, or the waypoint virtual surrogate system, as described in §III. The same handheld controller and control mapping were utilized across all conditions. Dependent variables included objective measures of completion time, response time, and interface usage, as well as subjective rankings directly comparing each interface and their perceived multitasking ability, stress, and ease of use.

A. Task, Environment, & Apparatus

Our experiment represents a scenario in which a user teleoperates a line-of-sight aerial robot to collect and analyze environmental data, inspired by analogous use cases for aerial robots in disaster response [5], construction [38], and space exploration [39]–[41].

The experimental task required the completion of two subtasks: (1) piloting an aerial robot such that it visited and "collected data" at a series of six points of interest (POI) and (2) periodically completing a quiz to represent the notion of analyzing data that the robot had just collected. The POIs were designated by six stools placed within the environment arranged in two rows of three (Figure 2-B). Participants were instructed to pilot the robot to maintain a stable hover over each POI (requiring both horizontal and vertical alignment) for five seconds to simulate collecting environmental data.

To examine user experiences operating the robot over a variety of distances, we divided the environment with tape into two zones: a *user-allowed* area and *user-restricted* area (constant across conditions). As a result, the user would operate the robot in close proximity (when visiting stools in the user-allowed area) or be forced to operate the robot at a distance (when visiting POIs further away in the user-restricted area). The experimental environment in which participants operated the aerial robot measured $6m \times 10m \times 6m$.

In all conditions participants wore an ARHMD while teleoperating the robot. The ARHMD presented the order in which to visit the POIs by displaying virtual order numbers (1–6) above each stool (order held constant across all participants and conditions). In all conditions, participants also received AR feedback in the form of a virtual progress bar that would fill up as participants hovered on top of POIs indicating that the robot was "collecting data." If participants left the POI early (designated by a virtual cylinder outlining POI area), their progress was lost and they would have to re-position the robot such that it was within the POI radius and the "data collection" (i.e., progress bar) would restart.

After completing the scan of each POI, participants were presented with a data analysis subtask, in which they were asked to answer two multiple choice quiz questions. These questions simulated the notion of participants analyzing the "data" the robot had just collected at the POI. The quiz questions appeared on a smartphone mounted to the user's arm (users selected preferred arm), again simulating the idea

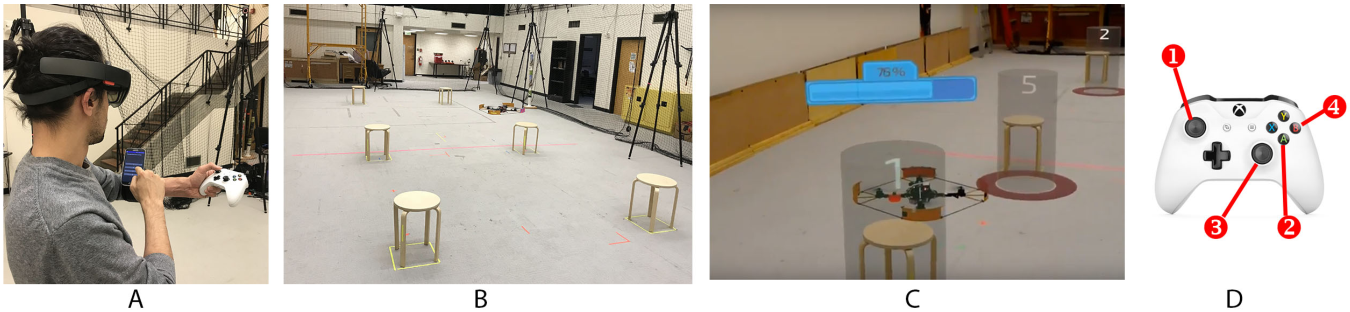

Fig. 2. (A) Participant performing data analysis. (B) Experiment environment. (C) POI target cylinder and scanning progress bar. (D) Teleoperation handheld controller and control mapping: (1) rotation and vertical translation; (2) confirm waypoint placement; (3) horizontal translation; (4) toggle pause/resume.

of using a robot to collect and analyze data in the field. Each quiz question presented users with a sentence of roughly 25 characters and required users to select the answer corresponding with the number of vowels the sentence contained from four options. The questions mimic a real data analysis task's requirement to multitask and were designed to force participants to perform contextual shifts between focusing on robot operation and focusing on a secondary interface for information analysis. Participants were not forced to complete the quiz immediately, instead they were free to continue piloting the robot to a new POI and complete the "data analysis" whenever they chose. However, each completed POI would add another two questions to the quiz queue. Successful completion of the overall task required that participants collect data from all POIs and answer all quiz questions.

We used an AscTec Hummingbird quadcopter as our aerial robot (a popular research "drone") for this experiment (Figure 2 B & C). Additionally, a Microsoft HoloLens ARHMD and Unity game engine were used to develop and deploy our applications. The HoloLens is a wireless, optical see-through stereographic augmented reality HMD with a $30° \times 17.5°$ FOV. The HoloLens was chosen due to its emerging popularity, ease of access, hands-free AR support, and high potential as a model for future consumer ARHMD systems.

B. Participants

We recruited a total of 18 participants (11 males, 7 females) from the University of Colorado Boulder campus. We worked to ensure our population sample contained both novice users and users experienced at piloting aerial robots. In total, 7 of our participants represented expert users who were recruited from a local "Drone Club," 8 participants reported moderate familiarity with aerial robots, while 3 participants had little to no experience operating flying robots. Average participant age was 20.7 ($SD = 3.59$), with a range of 18–27.

C. Experimental Procedure

Each participant completed the full task (visiting POIs to "collect data" and answering quiz questions) three times, once using each interface (baseline, realtime virtual surrogate, or waypoint virtual surrogate). Interface order was counterbalanced across participants to mitigate potential transfer effects such as learning or fatigue that might have been introduced due to our within-subjects design. Questionnaires were administered after the completion of each task.

Prior to each task, participants watched a short $60s$ tutorial video that presented the interface design they were going to use, covering both the controls and what the visual feedback looked like (if any). Next, the ARHMD application was then started, calibrated, and fitted on each participant, with a researcher receiving verbal confirmation that participants were able to see the augmented reality imagery as intended. The POI visitation order appeared over each POI as virtual imagery prior and during each trial to eliminate the potential advantage of order memorization. Finally, before each main task began, participants were then given two minutes to test the interface, giving them time to become familiar with the controller, augmented reality imagery, and the robot. Participants were instructed to complete the full task (visiting all POIs and completing all quizzes) as quickly as possible.

After completing the task with each individual interface, participants completed the full task one final time as part of a summary evaluation. During this phase, participants were told they could freely switch between any of the interface designs at any point, as many times as they desired. This enabled us to record objective data regarding user preferences regarding which interface they preferred to use. Prior to starting the summary evaluation, participants practiced switching between interfaces on-the-fly and were instructed to select the interface they wished to start with as active and reminded they can switch to any other design at any point.

D. Measures & Analysis

We used both objective and subjective measurements to evaluate our designs. We measured several objective aspects of task performance including: *completion time*—measured by time elapsed in seconds between the task start and finishing the last data analysis quiz; *response time*—the average time elapsed in seconds between scanning a POI and completing its associated data analysis quiz for all six points of interest; and *design usage*—measured by the percent of total task time that participants used each interface design during the summary evaluation phase in which they completed the task while free to switch between designs.

We also collected data from a number of subjective measurements. After using each interface, participants evaluated perceived interface usability using the System Usability Scale (SUS), an industry standard ten-item attitude survey. SUS scores below 68 are considered below average, scores above 68 are considered above average, and scores above 80.3 are considered in the top 10th percentile.

In addition to the SUS, we constructed a number of scales from 7-point Likert-style questionnaire items we created to measure participant perceptions and preferences. Scales rated perceived *ease of distal operation* (2 items, Cronbach's α = .77), *ease of precise positioning* (2 items, Cronbach's α = .77), *ability to multitask* (3 items, Cronbach's α = .94), and *stress* (5 items, Cronbach's α = .91).

Following the summary evaluation phase, participants directly compared the three designs, ranking them relative to one other (1 (best) – 3 (worst)). Participants ranked the designs in terms of *easy to learn* and *would want to use in the future*.

Qualitative feedback was obtained through open-ended questions posed to each participant. Questions included (but not limited to) "what made performing the task easier," "what made performing the task harder," and "how did this design impact your ability to control the drone."

We analyzed the objective measures, SUS, and constructed rating scales using a repeated-measures analysis of variance with experimental condition (i.e., interface design) as a fixed effect, with condition order included as a covariate to control for potential variance that might arise from ordering effects. Post-hoc tests used Tukey's Honestly Significant Difference (HSD) to control for Type I errors in comparing effectiveness across each interface.

Participant rankings of each interface were analyzed with a nonparametric Kruskal-Wallis Test with experimental condition as a fixed effect. Post-hoc comparisons used Dunn's Test for analyzing design sample pairs for stochastic dominance.

V. RESULTS

A. Objective Results

We analyzed our task performance metrics to determine if our AR surrogate designs helped participants teleoperate a collocated aerial robot more effectively. We found a significant main effect of robot interface design on task completion time, F(2, 45) = 13.65, $p < 0.001$. Using Tukey's HSD, we found the Realtime Virtual Surrogate (M = 186.39s), p = .001 and Waypoint Virtual Surrogate (M = 184.39s), p = .001 designs significantly improved completion time over the baseline interface (M = 260.11s).

We also found a significant main effect of design on response time, F(2, 45) = 8.43, $p < .001$. Post-hoc comparisons against the baseline (M = 90.56s) revealed that participants were able to respond to the data analysis quizzes significantly faster using the RVS (M = 47.44s), p = .004, and WVS (M = 44.61s), p = .002, designs.

Finally, we found a significant main effect in regard to design usage during the final trial summary evaluation where participants could switch between designs at will, F(2, 51) = 34.92, $p < .001$. Tukey's HSD revealed participants used WVS (M = 81.94%) significantly more than the Virtual Surrogate (M = 18.06%) and Baseline (M = 0%) designs (all comparisons at $p < .001$), with not a single participant ever using the baseline design at any point. See Figure 3 for objective results.

B. Subjective Results

Perceived interface usability was evaluated with the SUS. We found a significant main effect of interface design on SUS total score, F(2, 45) = 5.91, p = .005. Tukey's HSD revealed both of our AR designs had significantly higher usability ratings over the baseline (RVS: M = 86, p = .008; WVS: M = 83.6, p = .022; Baseline: M = 66.8).

Participants rated the designs in terms of ease of distal and precise teleoperation. A significant main effect was found in terms of ease of distal operation, F(2, 45) = 5.56, p = .007. Tukey's HSD confirmed both RVS (M = 6.25), p = .016, and WVS (M = 6.25), p = .016 were rated significantly higher than the baseline (M = 4.61). Finally, there was a significant main effect in regard to design and ease of precise positioning, F(2, 45) = 4.9, p = .012. Post-hoc analysis revealed that RVS (M = 6.31), p = .012 was rated significantly higher than the baseline (M = 4.72), while WVS (M = 5.89), p = .075 was not found to be significant at the $\alpha = .05$ level.

Participants also rated their perceived ability to multitask using each interface. We found a significant main effect of interface design on these ratings, F(2, 45) = 22.93, $p < .001$. Tukey's HSD exhibited that both RVS (M = 5.42), $p < .001$ and WVS (M = 6.10), $p < .001$ were rated significantly higher than the baseline (M = 3.00).

Our post-task surveys also collected data on perceived stress. We found a significant main effect of interface on stress ratings, F(2, 45) = 6.87, p = .003, with Tukey's HSD showing the baseline (M = 3.27) design elicited more perceived stress than either of the AR designs (RVS: M = 2.14, p = .013; WVS: M = 1.99, p = .004).

Finally, following the summary evaluation phase, participants were asked to compare the designs directly to one another, ranking them from 1 (best) to 3 (worst) across how easy they found the interfaces to learn and which design they would like to use in the future. The Kruskal-Wallis Test did not find a significant main effect in participant rankings of how easy the designs were to learn, H = 5.07, p = .079. However, we did find a significant main effect regarding which design participants would want to use in the future, H = 16.85, $p < .001$. Post-hoc analysis with Dunn's Test found that participants consistently ranked both RVS (M = 1.89), p = .017, and WVS (M = 1.5), $p < .001$ higher than the baseline (M = 2.61). No significant difference was observed in comparing the relative rankings of RVS and WVS. See Figure 4 for subjective results.

VI. DISCUSSION

We found that our surrogate AR designs allowed users to complete the task and respond to the data analysis quizzes faster than a baseline teleoperation interface modeled after existing systems in common use today. We attribute this gain

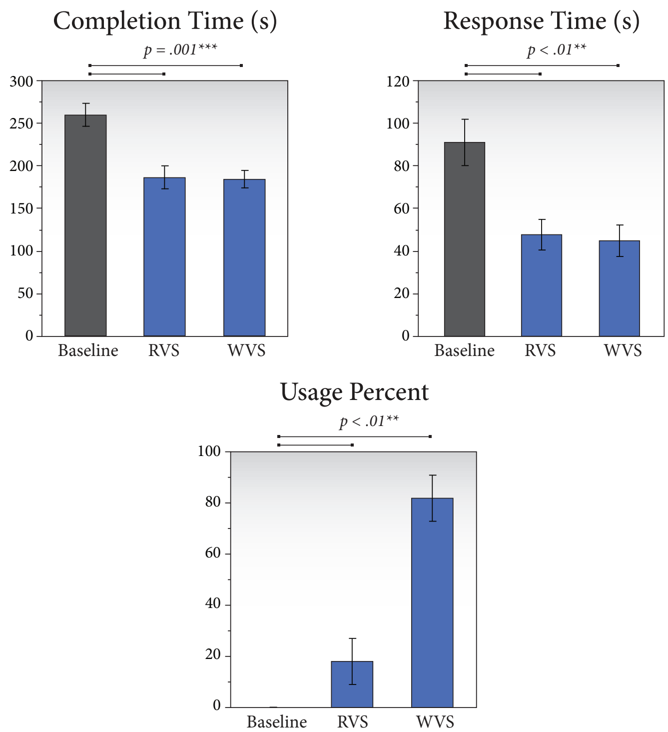

Fig. 3. The RVS and WVS systems showed improvement over the baseline along all objective measures (error bars encode standard error).

in performance to the support both the RVS and WVS designs provide for previewing robot actions, which has the added benefit that if users are happy with the preview, their time is freed to supervise the robot while completing other concurrent tasks. This improvement in multitasking did not appear to reduce participant operational accuracy in terms of navigating the robot to particular POIs, a promising development as typically we see a tradeoff space between interfaces that provide high accuracy (e.g., low level teleoperation) and those that support multitasking (e.g., supervisory control systems).

Additionally, our subjective results show the surrogate designs outperformed the baseline in terms of perceived ability to multitask and wanting to use the surrogate interfaces in the future. It is also important to note that not a single expert user used the baseline interface in the final summary evaluation, which is the control paradigm with which they had expertise in. Participant responses to open-ended questions provide further evidence to this effect; below we highlight representative participant quotes to provide additional context (note participant numbers are followed by an "E" or "N" label that indicate whether the participant was an expert or non-expert user respectively):

P4 - E [RVS]: *"After moving the hologram to the next target, one could focus on the quizzes until the drone reached the next target and completed it."*
P15 - N [WVS]: *"This design requires direction at the beginning and pretty much it was on automation after that. That allows me room to focus my mental capacity elsewhere."*
P6 - E [Baseline]: *"I had to keep my eyes on the drone which prevented me from multitasking."*

Overall, both novice and expert participants appreciated that the surrogate designs provided a preview of robot motions and found the previews to be helpful when operating the robot. These findings help support our hypothesis that providing support for users crossing the gulf of execution via the goal/action/evaluation cycle can improve teleoperation.

P9 - N [RVS]: *"Moving the hologram before the drone gave me greater confidence over what the drone was going to do."*
P8 - E [WVS]: *"It was much easier to control the drone using the waypoint method than a tradition sense of moving the drone itself. It allowed me to be more confident on where the drone was going to be and what it was going to do because I could see the hologram do it before the physical drone did."*
P16 - N [Baseline]: *"There were no guide points to go off of. As a new user, I found it helpful to have a drone hologram to guide me so I could see where the real drone was going."*

While most of our measures indicated support for the surrogate designs over the baseline, we also found a sub-theme among certain users that preferred how the baseline interface provided a feeling of "being in control." This subset of participants responded negatively to the lack of direct control provided by the WVS design.

P6 - E [Baseline]: *"I like to know what I'm making the drone do as it's happening, it prevents accidents for unaccounted obstacles or people passing through the drone's path."*
P12 - N [Baseline]: *"Performing the task was easier because I felt like I was actually in control of the drone."*
P5 - E [WVS]: *"This was my least favorite design. Less control, not as responsive."*

This disconnect between the increased efficiency provided by the AR interface and user detachment from the robot, observed with both novice and expert users, may be important for interface designers to consider when designing future AR teleoperation interfaces. However, many participants felt the RVS design struck a balance between automation and manual control and provided an effective, yet enjoyable experience.

P3 - N [RVS]: *"I liked this control scheme. I felt like I had more control over the drone, but it was still able to move without me always looking at it."*
P6 - E [RVS]: *"It offers the fine adjustment capabilities of the classic control scheme, in conjunction with the ability to automate the drone with the waypoint control scheme."*
P8 - E [RVS]: *"It was a great mix of the waypoints and traditional flying"*

Finally, some participants voiced concerns about the safety of the WVS interface, as they found themselves paying less attention to the robot.

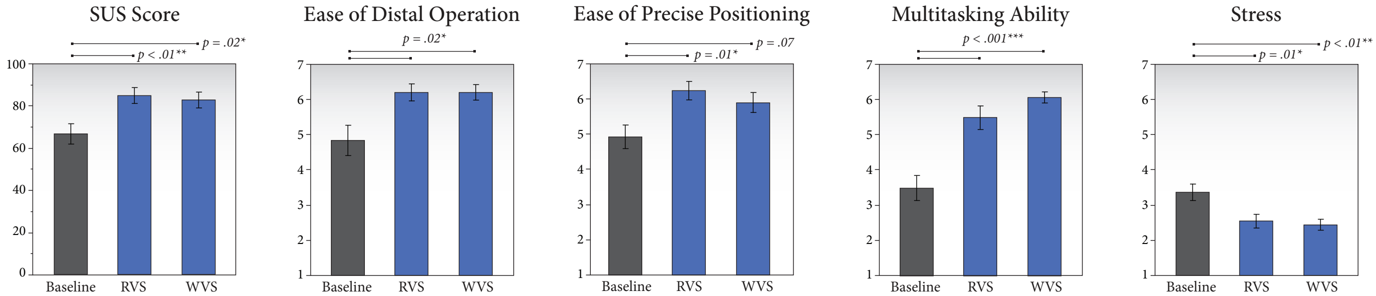

Fig. 4. The RVS and WVS design demonstrated improvements in usability and several additional subjective measurements (error bars encode standard error).

P1 - N [WVS]: *"This scheme made the task more efficient but might lull operators into a false sense of complacency."*

This concern seems valid as many participants happily reported they were able to "ignore" or "forget" about the drone after it was set in motion.

P13 - N [WVS]: *"This controller design made it easier because once I told the drone all the places I wanted it to go I could ignore it for most of the rest of the tasks."*
P8 - E [WVS]: *"It was very easy to set the waypoints and forget about the drone."*

Although the WVS system received mixed feedback in these open-ended responses, it was consistently ranked highest in terms of the system users most wanted to use again and by a wide margin the most-used system during the summary evaluation in which users were free to use whichever interface they preferred. Synthesizing the objective, subjective, and open-ended feedback suggests that the RVS system may be most appropriate for hobby use, non-critical tasks, or when users prefer more direct control as it struck a balance between being an enjoyable, responsive, and effective system, while the WVS system may be more useful in professional or multitasking applications where performance trumps user preference.

A. Limitations and Future Work

Although our work has shown promising results in regards to robotic teleoperation, it is not without limitations. Our current implementation relies on motion capture cameras for precise robot localization and navigation. Future work might merge our approach with more advanced robotic platforms and research in robotic perception and simultaneous localization and mapping (SLAM) to remove this constraint. Additionally, the field-of-view of modern ARHMDs is much smaller than that of the human eye making viewing the virtual imagery in its entirety difficult for users and was a common concern voiced by participants. Finally, the task's quizzes did not closely mimic real-life teleoperator analysis of collected data, which could have implications for completion time and participant mental load, however; this impact was mitigated to a degree as the quiz format was held constant across conditions.

Aside from these improvements to the current system, there are many ways to further explore this nascent design space. For example, our baseline emulates that of typical consumer teleoperation interfaces, but it does not take into account state-of-the-art research systems that have also proven to be more effective than consumer models. Future work might compare or integrate our approach with other newly developed teleoperation interfaces, such as XPose [21], Airways [19], or Szafir et al.'s "collaborative interface" [7]. Additionally, we believe that user comments regarding potential safety issues due to decreased attention with the WVS interface present an interesting area for followup work examining interactions between autonomy, trust, and safety. Finally, a follow-up study that compares our AR designs with 2D-waypoint or other graphical predictive interfaces already in existence would allow us to better understand the impact of immersive 3D AR compared to interfaces designed for traditional displays.

VII. Conclusion

In this work, we explore how modern AR technologies might improve robot teleoperation. Based on the idea of reducing the gulf of execution, we redirect user input to an AR virtual surrogate robot that foreshadows physical robot actions rather than have users control a physical robot directly.

As mixed, collocated teams of humans and robots become increasingly prevalent in our society, we envision interfaces, such as those evaluated in this study, assisting in fields where direct teleoperation of robots is necessary and range from robot inspection of equipment and structures, manual override teleoperation on factory floors, and all the way to space exploration where astronauts will be in direct contact and control of ground and aerial robots.

Through this work we have extended our knowledge of graphical predictive interfaces for robotic control (e.g., with an experiment directly comparing a real-time preview design (RVS) against a waypoint-style design (WVS) with traditional teleoperation as a baseline), demonstrated interface flexibility to switch between different robot control paradigms on-the-fly, and built upon existing knowledge of how ARHMD interfaces might mediate HRI. Our results show that emerging AR technologies hold great promise for the field of human-robot interaction and we hope our work spurs further research examining how AR and robotics might be successfully integrated.

Acknowledgments

This work was supported by an Early Career Faculty grant from NASA's Space Technology Research Grants Program under award NNX16AR58G and the NSF under award 1764092.

References

[1] R. H. Taylor, A. Menciassi, G. Fichtinger, P. Fiorini, and P. Dario, "Medical Robotics and Computer-Integrated Surgery," in *Springer Handbook of Robotics*, 2016, pp. 1657–1684.

[2] S. S. Nudehi, R. Mukherjee, and M. Ghodoussi, "A Haptic Interface Design for Minimally Invasive Telesurgical Training and Collaboration in the Presence of Time Delay," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 5, 2003, pp. 4563–4568.

[3] A. M. Okamura, "Methods for Haptic Feedback in Teleoperated Robot-assisted Surgery," *Industrial Robot: An International Journal*, vol. 31, no. 6, pp. 499–508, 2004.

[4] M. A. Menchaca-Brandan, A. M. Liu, C. M. Oman, and A. Natapoff, "Influence of Perspective-taking and Mental Rotation Abilities in Space Teleoperation," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI'07)*, 2007, pp. 271–278.

[5] F. Perez-Grau, R. Ragel, F. Caballero, A. Viguria, and A. Ollero, "Semi-autonomous Teleoperation of UAVs in Search and Rescue Scenarios," in *International Conference on Unmanned Aircraft Systems (ICUAS'17)*, 2017, pp. 1066–1074.

[6] J. Y. Chen, E. C. Haas, and M. J. Barnes, "Human Performance Issues and User Interface Design for Teleoperated Robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1231–1245, 2007.

[7] D. Szafir, B. Mutlu, and T. Fong, "Designing Planning and Control Interfaces to Support User Collaboration with Flying Robots," vol. 36, no. 5–7, 2017, pp. 514–542.

[8] D. A. Norman and S. W. Draper, *User centered system design; new perspectives on human-computer interaction*. L. Erlbaum Associates Inc., 1986.

[9] J. Scholtz, "Theory and evaluation of human robot interactions," in *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*. IEEE, 2003, pp. 10–pp.

[10] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst, "Grasp-Shell vs Gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR'14)*. IEEE, 2014, pp. 73–82.

[11] D. Harwell, "This government rule could cripple commercial drone flight," https://www.washingtonpost.com/blogs/the-switch/wp/2014/11/25/this-government-rule-could-cripple-commercial-drone-flight/, 2014.

[12] J. Glassmire, M. O'Malley, W. Bluethmann, and R. Ambrose, "Cooperative manipulation between humans and teleoperated agents," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS'04. Proceedings. 12th International Symposium on*. IEEE, 2004, pp. 114–120.

[13] A. D. Dragan and S. S. Srinivasa, "A Policy-blending Formalism for Shared Control," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 790–805, 2013.

[14] R. R. Murphy, *Disaster Robotics*. MIT Press, 2014.

[15] D. Pitman and M. L. Cummings, "Collaborative exploration with a Micro Aerial Vehicle: a novel interaction method for controlling a mav with a hand-held device," *Advances in Human-Computer Interaction*, vol. 2012, p. 18, 2012.

[16] J. L. Drury, J. Richer, N. Rackliffe, and M. A. Goodrich, "Comparing Situation Awareness for Two Unmanned Aerial Vehicle Human Interface Approaches," DTIC Document, Tech. Rep., 2006.

[17] M. A. Goodrich, B. S. Morse, D. Gerhardt, J. L. Cooper, M. Quigley, J. A. Adams, and C. Humphrey, "Supporting Wilderness Search and Rescue using a Camera-equipped Mini UAV," *Journal of Field Robotics*, vol. 25, no. 1-2, pp. 89–110, 2008.

[18] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, "Ecological Interfaces for Improving Mobile Robot Teleoperation," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 927–941, 2007.

[19] C. Gebhardt, B. Hepp, T. Nägeli, S. Stevšić, and O. Hilliges, "Airways: Optimization-based Planning of Quadrotor Trajectories According to High-level User Goals," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'16)*, 2016, pp. 2508–2519.

[20] N. Joubert, M. Roberts, A. Truong, F. Berthouzoz, and P. Hanrahan, "An Interactive Tool for Designing Quadrotor Camera Shots," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 238, 2015.

[21] Z. Lan, M. Shridhar, D. Hsu, and S. Zhao, "XPose: Reinventing User Interaction with Flying Cameras," 2014.

[22] T. Nägeli, L. Meier, A. Domahidi, J. Alonso-Mora, and O. Hilliges, "Real-time Planning for Automated Multi-view Drone Cinematography," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 132, 2017.

[23] S. A. Green, M. Billinghurst, X. Chen, and J. G. Chase, "Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design," *International Journal of Advanced Robotic Systems*, vol. 5, no. 1, p. 1, 2008.

[24] S. A. Green, J. G. Chase, X. Chen, and M. Billinghurst, "Evaluating the Augmented Reality Human-Robot Collaboration System," *International Journal of Intelligent Systems Technologies and Applications*, vol. 8, no. 1-4, pp. 130–143, 2009.

[25] S. Kasahara, R. Niiyama, V. Heun, and H. Ishii, "extouch: spatially-aware embodied manipulation of actuated objects mediated by augmented reality," in *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*. ACM, 2013, pp. 223–228.

[26] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "Touchme: An augmented reality based remote robot manipulation," in *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, vol. 2, 2011.

[27] H. Hedayati, M. Walker, and D. Szafir, "Improving Collocated Robot Teleoperation with Augmented Reality," in *In Press, ACM/IEEE International Conference on Human-Robot Interaction (HRI'18)*, 2018.

[28] B. Giesler, T. Salb, P. Steinhaus, and R. Dillmann, "Using augmented reality to interact with an autonomous mobile platform," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'04)*, vol. 1. IEEE, 2004, pp. 1009–1014.

[29] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating robot arm motion intent through mixed reality head-mounted displays," *arXiv preprint arXiv:1708.03655*, 2017.

[30] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating Robot Motion Intent with Augmented Reality," in *In Press, ACM/IEEE International Conference on Human-Robot Interaction (HRI'18)*, 2018.

[31] A. D. Dragan and S. S. Srinivasa, *Formalizing assistive teleoperation*. MIT Press, July, 2012.

[32] M. Sauer, M. Hess, and K. Schilling, "Towards a predictive mixed reality user interface for mobile robot teleoperation," *IFAC Proceedings Volumes*, vol. 42, no. 22, pp. 91–96, 2009.

[33] C. Guo, J. E. Young, and E. Sharlin, "Touch and toys: new techniques for interaction with a remote group of robots," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 491–500.

[34] K. Ishii, Y. Takeoka, M. Inami, and T. Igarashi, "Drag-and-drop interface for registration-free object delivery." in *RO-MAN*. Citeseer, 2010, pp. 228–233.

[35] A. Singh, S. H. Seo, Y. Hashish, M. Nakane, J. E. Young, and A. Bunt, "An interface for remote robotic manipulator control that reduces task load and fatigue," in *RO-MAN, 2013 IEEE*. IEEE, 2013, pp. 738–743.

[36] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual reality on a wim: interactive worlds in miniature," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 1995, pp. 265–272.

[37] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir, "Designing for depth perceptions in augmented reality," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2017, pp. 111–122.

[38] J. Irizarry, M. Gheisari, and B. N. Walker, "Usability assessment of drone technology as safety inspection tools," *Journal of Information Technology in Construction (ITcon)*, vol. 17, no. 12, pp. 194–212, 2012.

[39] M. Bualat, J. Barlow, T. Fong, C. Provencher, T. Smith, and A. Zuniga, "Astrobee: Developing a Free-flying Robot for the International Space Station," in *AIAA SPACE 2015 Conference and Exposition*, 2015, p. 4643.

[40] T. Fong, M. Micire, T. Morse, E. Park, C. Provencher, V. To, D. Wheeler, D. Mittman, R. J. Torres, and E. Smith, "Smart SPHERES: A Telerobotic Free-Flyer for Intravehicular Activities in Space," in *Proc. AIAA Space'13*, 2013.

[41] D. Szafir, "Human interaction with assistive free-flying robots," Ph.D. dissertation, The University of Wisconsin-Madison, 2015.