Improving Collocated Robot Teleoperation with Augmented Reality

Hooman Hedayati
University of Colorado Boulder
Boulder, USA
Hooman.Hedayati@Colorado.edu

Michael Walker
University of Colorado Boulder
Boulder, USA
Michael.Walker-1@Colorado.edu

Daniel Szafir
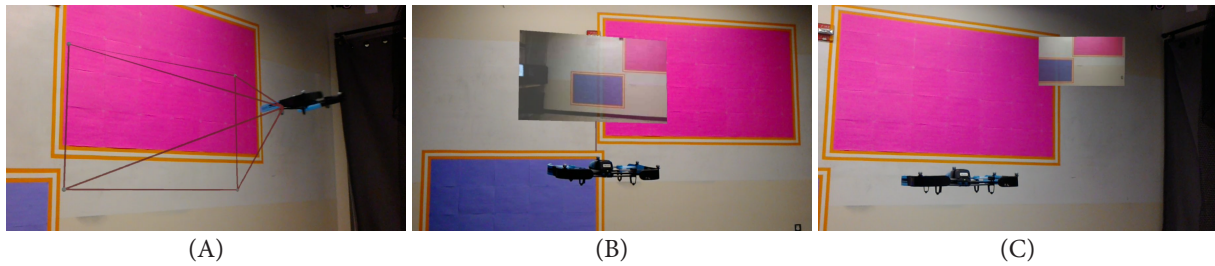University of Colorado Boulder
Boulder, USA
Daniel.Szafir@Colorado.edu

(A) (B) (C)

Figure 1: In this research, we explore how to leverage augmented reality (AR) to improve robot teleoperation. We present the design and evaluation of 3 AR design prototypes: (A) the Frustum design augments the environment giving users a clear view of what real-world objects are within the robot's FOV; (B) the Callout design augments the robot like a thought-bubble, attaching a panel with the live video feed above the robot; (C) the Peripheral design provides a window with the live video feed fixed in the user's periphery.

ABSTRACT

Robot teleoperation can be a challenging task, often requiring a great deal of user training and expertise, especially for platforms with high degrees-of-freedom (e.g., industrial manipulators and aerial robots). Users often struggle to synthesize information robots collect (e.g., a camera stream) with contextual knowledge of how the robot is moving in the environment. We explore how advances in augmented reality (AR) technologies are creating a new design space for mediating robot teleoperation by enabling novel forms of intuitive, visual feedback. We prototype several aerial robot teleoperation interfaces using AR, which we evaluate in a 48-participant user study where participants completed an environmental inspection task. Our new interface designs provided several objective and subjective performance benefits over existing systems, which often force users into an undesirable paradigm that divides user attention between monitoring the robot and monitoring the robot's camera feed(s).

CCS CONCEPTS

• **Computer systems organization** → **External interfaces for robotics**; • **Human-centered computing** → **Mixed / augmented reality**; **User interface design**; **Laboratory experiments**;

KEYWORDS

Aerial robots; drones; teleoperation; augmented reality; mixed reality; interface design; aerial photography; free-flying robot

1 INTRODUCTION

Robotic teleoperation, in which a user manually operates a robot, typically requires a high level of operator expertise and may impose a considerable cognitive burden. However, it also affords a high degree of precision and may require little autonomy on the part of the robot. As a result, teleoperation is still a dominant paradigm for human-robot interaction in many domains, including the operation of surgical robots for medical purposes [29, 30, 39], robotic manipulators in space exploration [22], and aerial robots for disaster response [32]. Even in future systems, where robots have achieved a greater degree of autonomy than in current human-robot teams, teleoperation may still have a role [3, 38]. For instance, in "shared control" and "user-directed guarded motion" paradigms, robots enable users to input direct teleoperation commands, but use these commands in an attempt to infer user intentions, rather than following received inputs exactly, particularly if received inputs might lead to unsafe operation [7, 26, 29, 30].

A substantial body of research has explored human performance issues in various forms of robotic teleoperation interfaces and mixed teleoperation/supervisory control systems (see [3] for a survey). In particular, prior work has highlighted the issue of *perspective-taking*—the notion that poor perceptions of the robot and its working environment may degrade situational awareness and thus have a

detrimental effect on operation effectiveness [3, 22]. This can be a problem for both remote teleoperation (as found in space exploration) and even when operator and robot are collocated (as may happen in search-and-rescue or building inspection scenarios).

In this work, we are interested in designing new interfaces that better support perspective-taking, particularly within the context of operating aerial robots. Aerial robots show significant promise in data-gathering, inspection, and recording tasks across a variety of domains (see [37] for a survey), although they are typically restricted to operating within line-of-sight range [32]. Such operations often require trained experts because users must rapidly and accurately synthesize information provide directly from the robot (e.g., via a live camera feed) with an understanding of where the robot is located within the larger context of the environment.

Current interface designs may exacerbate this problem as live robot camera feeds are typically presented in one of two ways: viewed directly in display glasses or on a traditional screen (e.g., a mobile device, tablet, or laptop computer). While video display glasses may help users achieve an egocentric understanding of what the robot can see, they may degrade overall situational awareness by removing a third-person perspective that can aid in understanding operating context, such as identifying obstacles and other surrounding objects that are not in direct view of the robot. On the other hand, routing robot camera feeds through traditional displays means that, at any point in time, the operator can only view the video stream on their display or the robot in physical space. As a result, operators must make constant context switches between monitoring the robot's video feed and monitoring the robot, leading to a divided attention paradigm that bears similarities to texting while driving.

We seek to address this challenge by designing new interface paradigms and interaction techniques that better support robot operation. Our approach is inspired by research exploring the integration of teleoperation and virtual reality [21, 25, 28] and prior work envisioning that augmented reality might someday afford new ways of facilitating information exchange between people and robots [6, 13, 14, 24, 31, 36, 40]. Recent advances in the development of consumer-grade, see-through augmented reality head-mounted displays (ARHMDs) are now making this vision possible, creating an exciting new design space for the field of human-robot interaction. We provide an in-depth examination of this design space within the context of supporting collocated robot teleoperation, specifically for aerial robots.

In this paper, we address the research question of: "how might augmented reality support collocated user teleoperation for collecting environmental data?" The answer to this question is important to many fields (e.g. search-and-rescue, inspecting construction sites, monitoring manufacturing logistics, etc). We outline relevant work that informed our design process and present a framework for designing augmented reality feedback to enhance HRI. We then describe the development of three new ARHMD interfaces that provide novel forms of visual feedback to operators aimed at reducing the need for context switching and better supporting user perspective-taking. Each design represents a unique fusion for embedding information about the robot's live camera feed directly into the user's view of the robot and environment. We evaluate these designs within a laboratory experiment with 48 participants. This study examined participant abilities at teleoperating a collocated aerial robot while completing a line-of-site environmental inspection task using our new designs compared with an industry standard interface in common use today. We conclude with a discussion of how augmented and mixed reality technologies may represent promising new tools for designing robot control interfaces.

2 BACKGROUND

In this work, we explore the design of augmented reality interfaces that enhance collocated robot teleoperation, specifically within the context of aerial robots. Below, we review current aerial robot control interfaces and provide a brief primer on augmented reality technologies.

2.1 Aerial Robot Interfaces

Currently, most interfaces for aerial robots take one of two forms: direct teleoperation interfaces, where robots are 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 outlining desired robot path [38]. In this work, we are primarily interested in designing new interface techniques to support the former paradigm.

Teleoperation interfaces for aerial robots often require that users have a great deal of skill to be able to pilot a robot with potentially unfamiliar degrees of freedom while monitoring a live robot video feed. As an example of the skill required, the Federal Aviation Administration (FAA) in the United States initially considered regulations that would have required commercial aerial robot operators to obtain a pilot's license [15] and still today requires that such robots be operated within line-of-sight range.

A great deal of research has sought to improve teleoperation paradigms. For example, certain interfaces provide the operator with a first-person view of the robot's video feed via display glasses. While this may help in certain tasks, it can also degrade overall situational awareness as the operator loses all contextual knowledge of the surrounding environment outside of the robot's immediate field of view. Other interfaces combine a live video display with virtual map data, often mixing teleoperation with forms of autonomous waypoint navigation [8, 12, 28]. Still other approaches have looked to develop control systems using multi-modal interfaces [35] including exotic designs where the user wears a head-mounted display (HMD) and the robot is controlled via a "floating head" metaphor, changing orientation in sync with the user [16]. Finally, other research has advanced the notion of "perceived first-order control," which allows users to "nudge" an aerial robot using gestures on a mobile touchscreen, supporting more precise and safe operation [33].

In this work, we seek to extend such prior research. One limitation of many existing systems is that they often focus on remote teleoperation, rather than collocated teleoperation, even though there are many deployments in which collocated control might be useful (e.g., monitoring construction sites, over-the-hill reconnaissance, factory logistics management, environmental surveys of mills, etc.) or even required due to line-of-sight regulations. In addition, we know of no current interface that enable users to view information collected by the robot (e.g., live video feed) while monitoring the robot itself directly within the operating environment. Instead, traditional interfaces present a robot video feed and other sensor information on a display (e.g., mobile device), requiring that users choose between

monitoring the robot or monitoring the robot's video feed, in a paradigm analogous to people who text while driving. Just as in texting while driving, monitoring a mobile device during teleoperation may distract users and degrade situational awareness, leading to poor piloting and even crashes. Our research explores how augmented reality might enable a new model of teleoperation interfaces that mitigates this issue, enabling users to synthesize the knowledge and benefits of both first- and third-person views.

2.2 Augmented Reality

Augmented reality (AR) technology overlays computer graphics onto real world environments in real time [13, 14]. AR interfaces have three main features: (1) users can view real and virtual objects in a combined scene, (2) users receive the impression that virtual objects are actually expressed and embedded directly in the real world, and (3) the virtual objects can be interacted with in real-time [1]. This contrasts with purely virtual environments or other parts of the mixed reality continuum, such as augmented virtuality in which real-world objects are mixed into a virtual environment [23].

Early AR systems were often custom-made in research laboratories and were quite limited in display fidelity, rendering speed, support for interaction, and generalizability. However, recent advancements in augmented reality head-mounted display (ARHMD) technology is creating an ecosystem of standardized, consumer-grade see-through ARHMDs. For example, the HoloLens and Meta 2 ARHMDs both afford high resolution stereographic virtual imagery displayed at a 60Hz refresh rate, built-in gesture tracking, depth sensing, and integration with standard development tools such as Unity and Unreal Engine. This advance in hardware accessibility is creating new opportunities for exploring AR as an interaction medium for enhancing HRI. Although we are not the first to recognize that AR holds potential for improving human-robot interactions (see [6, 13, 14, 24, 36]), we believe that this area is critically understudied and represents a fairly nascent research space, especially within the context of examining how the capabilities provided by modern ARHMD hardware might afford novel methods for teleoperation.

3 DESIGN

We undertook an iterative design process to explore the space of how ARHMD technology might mediate collocated human-robot interaction by providing feedback to support aerial robot teleoperation. Our design process began with an analysis of the potential of augmented and mixed reality technologies. Synthesizing information from past work in robot interface design with research across the mixed reality continuum, we developed a high-level framework for considering how ARHMD feedback might enhance HRI.

3.1 A Design Framework for AR-HRI

Our framework classifies potential interface designs for augmenting human-robot interactions with virtual imagery into three main categories, regarding whether an AR interface provides supplemental information in a manner (1) *augmenting the environment*, (2) *augmenting the robot*, or (3) *augmenting the user interface*.

Augmenting the environment: In this paradigm, interfaces can display information regarding robot operations and data collected by

the robot as virtual imagery that is directly embedded into the context of the operational environment, using an *environment-as-canvas* metaphor. For example, objects that the user/robot has inspected (or plans to inspect) might be highlighted, or information might be added to better indicate the robot's field-of-view. This paradigm affords a full three-dimensional "canvas" to utilize for virtual imagery, rather than a two-dimensional display typical in mobile, tablet, or laptop interfaces.

Augmenting the robot: In this archetype, virtual imagery may be attached directly to the robot platform in a *robot-as-canvas* metaphor. This technique may provide contextually relevant cues to an operator in a more fluid manner than traditional interfaces. For example, rather than displaying a battery indicator on a 2D display, requiring that the operator take their eyes off the robot to check status, virtual imagery might instead provide indication icons directly above the robot in physical space, enabling the operator to maintain awareness of both robot position and status.

Augmenting the user interface: In this paradigm, virtual imagery is displayed as an overlay directly in front of the user to provide an interface to the physical world, inspired by "window-on-the-world" AR applications [9] and heads-up display technologies used for pilots [11, 20, 27]. This *interface-as-canvas* metaphor may take a great deal of inspiration from traditional 2D interface designs, for instance providing supplemental information regarding robot bearing, attitude, GPS coordinates, or connection quality in the user's periphery while maintaining their view of the robot and preserving situational awareness of the environment.

These design paradigms may provide benefits over traditional interfaces. For example, ARHMDs support stereographic cues that can better leverage human depth perception, as opposed to monocular cues in traditional interfaces. Moreover, these paradigms enable interfaces that provide feedback directly in the context of where the robot is actually operating, reducing the need for context-switching between monitoring the robot and monitoring operations data and thus helping to resolving the perspective-taking problem.

3.2 Creating Reference Prototypes

Inspired by prior work in robot interface design, we used the framework described above to develop several interface prototypes that provide sample references within the overall space of how augmented reality might support collocated/line-of-sight robot teleoperation. Although ARHMD interfaces might provide feedback on many different aspects relevant to teleoperation, we focused our design exploration specifically on how to convey information about the robot's camera, as this is typically the most critical information for aerial robots operators. We developed three primary prototypes, each of which falls within one of the paradigms in our design framework.

We refer to these three design prototypes as *Frustum*, an example of augmenting the environment, *Callout*, an example of augmenting the robot, and *Peripherals*, an example of augmenting the user interface. These designs are each based on prior robot interface designs or other metaphors that may be common to user experiences, adjusted and extended to take advantage of ARHMD technology. Each offers potential tradeoffs in terms of how they support perspective-taking, the total information conveyed, potential scalability across

interaction distances, and possibility for user distraction and/or interface overdraw. While our work targets the context of collocated aerial robot teleoperation, we believe that our design framework and methodology, along with the main design metaphors and interface techniques we develop, may also provide utility for other robotic platforms (e.g., ground-mobile robots, underwater robots, or industrial robotic manipulators) and even higher-level supervisory or collaborative control systems.

Frustum: The Frustum design provides an example of augmenting the environment. This design has a spatial focus as it provides virtual imagery that displays the robot's camera frustum as a series of lines and points, similar to what might be seen emanating from a virtual camera in computer graphics and modeling applications (e.g., Maya, Unity, etc.). The virtual frustum provides information on the robot's aspect ratio, orientation, and position, while explicitly highlighting what objects in the environment are within the robot's field-of-view (Figure 1-A). An additional benefit of this design is the ability to see a robot's sensor frustum even if there is inherently no real-time feedback and/or if taking aggregate measurements (e.g. LiDAR, air quality sensor, spectrometer, etc.).

Callout: The Callout design represents an example of augmenting the robot itself with virtual imagery. It uses a metaphor inspired by callouts, speech balloons, and thought bubbles, along with prior work rendering real robot camera feeds within the context of virtual representations of the robot environment [28]. As in [28], the robot's live camera feed is displayed on a panel with an orientation corresponding to the orientation of the camera on the physical robot, enabling information in the video to be spatially similar to the corresponding physical environment (Figure 1-B). This feed also provides implicit information regarding the distance between the robot and operator for a perspective projection transformation is applied to the video callout panel based on the calculated offset between the user and the robot. As a result, panel size is proportional to operating distance, just as perceived sizes of real-world objects are proportional to viewer distance. While this design decision may impair long-range operations, we speculate that it may better support scalability and fan-out for operating multiple robots and provides a more realistic embedding of the information directly in the context of where the robot is collecting data at any given time. Alternate variations on this design might instead use billboarding, such that the callout panel always faces the operator. While this would ensure the operator always has a direct view of the robot's video stream, it may remove potentially useful orientation cues. Another design variation might use an orthographic projection, rather than perspective, such that the video panel always remains at a fixed size to the operator regardless of robot distance (this would create a similar effect to the Peripherals design, discussed below).

Peripherals: Our final design illustrates a potential method for an augmented-reality user interface that provides contextual information regarding the robot's camera feed in an egocentric manner. This design displays a live robot video feed within a fixed window within the user's view (Figure 1-C). Designers may specify fixed parameters regarding window size and location (e.g., directly in front of the user or in their peripheral view), or provide support for dynamic interaction that allows the user to customize window width, height, position, and even opacity. In this work, we placed the window in the user's periphery, fixed within the upper-right corner of the

user's view, in a manner inspired by ambient and peripheral displays [5, 17, 34].

4 EXPERIMENT

We conducted a 4×1 between-participants experiment to evaluate how our designs might affect user teleoperation of a collocated flying robot. The study tasked participants with operating a Parrot Bebop quadcopter to take several pictures in a laboratory environment as an analog to aerial robot inspection and survey tasks. The independent variable in this study corresponded to what type of teleoperation interface the participant used (four levels: Frustum, Callout, and Peripherals designs plus a baseline). In the baseline condition, participants still wore an ARHMD (to control for possible effects of simply wearing a HMD), but did not see any augmented reality imagery. Instead, participants used the Freeflight Pro application[1], the official piloting interface supplied by Parrot for the Bebop robot (the platform used in this experiment, detailed in §4.2). Dependent variables included objective measures of task completion and subjective ratings of operator comfort and confidence.

4.1 Task and Environment

Our overall experimental design was inspired by contexts where free-flying robots might assist with environmental inspections and surveys within human environments, as is already common practice among drone hobbyists and might soon be found within domains including disaster response [32], operations on-board the International Space Station [2, 10, 37], and journalism [4, 19]. In the study, participants operated an aerial robot in a shared environment. The environment measured $5\,m \times 5\,m \times 3\,m$ and contained motion tracking cameras that we utilized for precisely tracking the robot to ensure ARHMD visualizations were displayed at the appropriate locations for the Frustum and Callout designs (motion tracking was not needed for the baseline condition or the Peripherals design).

Two inspection targets adorned the walls of the experimental environment in the form of rectangular frames colored in pink and purple with an orange outline (targets can be seen in Figure 1). The larger pink target is $1.78\,m \times 1.0\,m$ and positioned $1.3\,m$ from the ground to its bottom edge. The smaller purple target is $1.35\,m \times 0.76\,m$ and positioned $0.34\,m$ from the ground to its bottom edge. The aspect ratio of the targets exactly matches that of the robot's camera. This gives participants the ability to capture exact photos of the targets. It is relatively easier to capture a perfect image of the pink target as it is larger and higher off the ground, while the purple target is more challenging due to its smaller size and shorter height from the ground (operating an aerial robot close to the ground is more challenging due to instabilities resulting up-drafts reflected off the ground). To capture a perfect image of the purple target requires the robot to fly in closer proximity to both the wall and ground, which increases chances of crashes due to operator error. Participants were tasked with piloting an aerial robot to take photos of the targets in a set order, first inspecting the larger pink target and then the smaller purple target. Participants were directed to prioritize capturing imagery as *fast* as possible, as *precisely* as possible, and with the *fewest number of total pictures* as possible. In terms of precision, participants were instructed to record a photo that captured the entirety of the pink or
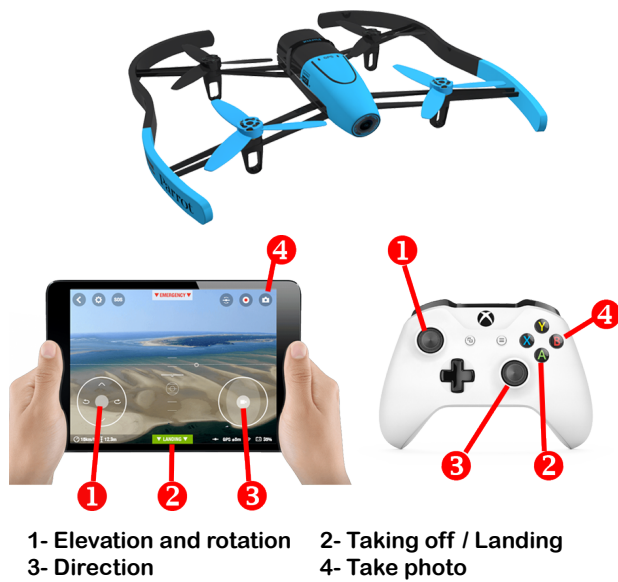
---

[1] https://itunes.apple.com/us/app/freeflight-pro/id889985763?mt=8

1- Elevation and rotation   **2- Taking off / Landing**
3- Direction   **4- Take photo**

Figure 2: Top: the Parrot Bebop robot used in this study. Bottom-left: the Freeflight Pro application used as our baseline condition. Bottom-right: the Xbox controller used in our AR conditions.

purple target region, with as little additional imagery (e.g., the orange target frame or other parts of the scene) in the photo as possible. Participants were allowed to determine when they had captured what they believed to be a suitable image such that they could then move on to the next target. Overall, this task mimicked an environmental inspection mission, requiring that participants operate a line-of-sight robot to take off from a set starting position, inspect a series of targets in order, and land the robot within six minutes. If participants crashed the robot it was reset to the starting position, after which they were allowed to continue performing the task as long as time remained.

4.2   Experimental Apparatus and Implementation

Robotic Platform: We used a Parrot Bebop quadcopter as our aerial robot for this experiment (Figure 2). The Bebop is a popular consumer-grade "drone" appropriate for flying indoors and outdoors with a digitally stabilized 14MPixel HD camera and autonomous hovering capabilities.

ARHMD Platform: We used a Microsoft HoloLens as our ARHMD. The HoloLens is a wireless, optical see-through stereographic augmented reality HMD with a $30° \times 17.5°$ FOV, an inertial measurement unit, depth sensor, ambient light sensor, and several cameras and microphones that support voice input, gesture recognition, and head tracking. The HoloLens was chosen due to its emerging popularity, ease of access, ability to support hands-free AR, and high potential as a model for future consumer ARHMD systems.

Implementing Teleoperation Interfaces: Participants in the baseline condition still wore the HoloLens ARHMD, but received no augmented reality visualizations. Instead, they controlled the robot via the "Free Flight Pro" application on an iPad. The Free Flight application represents the default control software for the Bebop robot

and is also developed by Parrot (the Bebop manufacturer). It is a popular application (with an average 3.8 rating from 24,654 reviews on the Android App Store) and is highly representative of modern aerial robot control interfaces used in practice today that aim to provide users with an intuitive control scheme. The application provides touchscreen controls for taking off/landing, positioning/orienting the robot, and recording pictures/video, all overlaid on the live video feed from the robot (Figure 2). In the absence of user input, the application ensures the robot continues to hover and automatically lands the robot upon detecting low battery.

Unfortunately, limitations in the robot platform and Freeflight application prevented us from using it as the control input in the other experimental conditions: if Freeflight is connected to the robot, it is not possible to also stream the robot's video feed to any other device. This makes it impossible to use the Freeflight interface with the Callout or Peripherals designs. Instead, in the AR conditions, participants received AR feedback while operating the robot using a wireless Xbox One controller. We set the button/joystick mappings on the Xbox controller to match the touchscreen controls in the Freeflight Pro application and calibrated Xbox controller sensitivity to make it as similar to the Freeflight controller as possible (Figure 2). As with the Freeflight Pro app, in the absence of user input the robot would continue to hover in place.

The augmented reality visuals for the Frustum, Callout, and Peripherals designs were implemented using the Unity game engine and deployed as an application that ran on the Microsoft HoloLens. The Frustum and Callout designs require a real-time understanding of robot position, such that virtual imagery can be correctly displayed in the environment relative to the robot (Frustum) or directly as an attachment to the robot itself (Callout). To accomplish this, we used motion tracking cameras to precisely localize the robot, feeding in robot position and orientation values to the HoloLens application. The Peripherals design did not rely on the motion tracking setup, however both the Callout and Peripherals designs displayed a live video feed from the robot's camera as a virtual object in augmented reality. To accomplish this, the robot's video stream was wirelessly broadcast to the HoloLens by routing it through a desktop computer. With this method, we were able to achieve an average framerate of 15 frames per second (FPS), slightly slower than the ~30FPS provided to participants in the baseline condition by the Freeflight Pro application.

Each of the three main designs we tested have several parameters that may be tuned by designers or users; however, we fixed each of these parameters to control for potential variance in our experiment. The Frustum was displayed as red lines in a wireframe view, as opposed to a shaded or highlighted region to minimize potential occlusion of the environment or task targets. The Callout was designed to emanate from the top of the robot, such that the video feed always appeared $13.5\,cm$ above the center of the Bebop as the operator flew it throughout the environment. The Peripherals design set the camera feed window to the upper-right corner of the user's view.

4.3   Participants

We recruited a total of 48 participants (28 males, 19 females, 1 self-reported non-binary) from the University of Colorado Boulder campus to participate in our study, which was approved by our

university IRB. Males and females were evenly distributed across conditions. Average participant age was 22.2 (SD = 7.2), with a range of 18 – 58. On a seven-point scale, participants reported a moderate prior familiarity with both aerial robots (M = 3.48, SD = 1.6) and ARHMDs (M = 3.38, SD = 1.75).

4.4 Experimental Procedure

The study took approximately 30 minutes and consisted of five main phases: (1) introduction, (2) calibration, (3) training, (4) task, and (5) conclusion.

(1) First, participants were given a high-level overview of the experiment and signed a consent form, then were led into the experimental space. (2) Participants were then fitted with the HoloLens and either given an iPad running the Freeflight Pro application or an Xbox controller depending on condition. At this point, the appropriate HoloLens application was also started based on condition (Frustum, Callout, or Peripherals, with no AR application for baseline participants). (3) Participants then received instructions on the controls (i.e., button-maps for the Freeflight and Xbox controllers) and had two minutes to practice piloting the robot. After two minutes were up, the robot landed and was placed in a fixed starting position for all participants. (4) Participants then completed the main task of inspecting the targets in order, as detailed in §4.1. Participants received six minutes to pilot the robot such that it took off from a fixed starting location, captured images of the targets, and then landed, simulating an environmental inspection mission. If participants crashed the robot, it was reset to the starting position and participants could continue the task if time still remained. (5) Once participants completed the task or the six minutes allotted to the task ran out, participants were given a post-survey on their experience and then debriefed.

4.5 Measures & Analysis

We used objective, behavioral, and subjective measurements to characterize the utility of our interface designs. We measured several objective aspects of task accuracy including: *accuracy*—measured by how well participant photographs captured the inspection target, each of which consisted of a visible, uniform grid of $297mm \times 210mm$ rectangles, which allowed us to measure the accuracy rate by comparing the rectangle grid in a perfect shot with the photos captured by participants; *completion time*—measured by the total flight time (lower times mean more efficient performance); and *operational errors*—the number of times they crashed the robot or otherwise caused it to land prematurely.

We also recorded first-person and third-person video to analyze behavioral patterns in participant actions. Two coders annotated video data from each interaction based on when participants were able to view the robot and when they were not. Data was divided evenly between coders, with an overlap of 15% of the data coded by both. Inter-rater reliability analysis revealed substantial agreement between raters (Cohen's κ = .92) [18]. This coding enabled us to calculate *distracted gaze shifts*—the number of times the participant was distracted looking away from the robot during the task; and *distraction time*—the total time spent not looking at the robot. We were interested in both measures as many small gaze shifts might be just as detrimental as fewer but longer periods of distraction.

We constructed several 7-point scales using Likert-type questionnaire items to capture subjective participant responses. These scales measured how the interface designs affected participant *comfort* (3 items, Cronbach's α = .86), *confidence* (5 items, Cronbach's α = .95), and perceptions of *task difficulty* (5 items, Cronbach's α = .93) while operating the robot. Participants also gave open-ended responses regarding their experiences.

Finally, we had all participants evaluate perceived 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.

We analyzed data from our objective and subjective measures using a one-way Analysis of Variance (ANOVA) with experimental condition (i.e., teleoperation interface) as a fixed effect. Post-hoc tests used Tukey's Honestly Significant Difference (HSD) to control for Type I errors in compared effectiveness across each interface.

5 RESULTS

Objective Results — We analyzed our objective task metrics to confirm that our designs were useful in allowing participants to teleoperate a collocated aerial robot more effectively. We found a significant main effect of design on task performance scores for accuracy, $F(3, 44)$ = 25.01, $p < .0001$. Tukey's HSD revealed that the Frustum (M = 63.2%) and Callout (M = 67.0%) interfaces significantly improved inspection performance over the baseline interface (M = 31.33%), with the Peripheral design (M = 81.1%) showing even further benefits by significantly outperforming both Frustum and Callout (all post-hoc results with $p < .0001$). We also found a significant main effect of design on task completion time, $F(3, 44)$ = 3.83, p = .016. Post-hoc comparisons against the baseline (M = 239.70s) revealed that participants were able to complete the task significantly faster using the Frustum (M = 140.69s), p = .017, and Peripherals (M = 154.44s), p = .050, but not the Callout (M = 191.09s), p = .434. Examining occurrences when users crashed the robot, we found a significant effect of interface design on operational errors, $F(3, 44)$ = 9.24, $p < .001$ with each of our AR designs significantly reducing the number of crashes compared to the baseline (Frustum: M = .250, $p < .0001$; Callout: M = .667, p = .003; Peripherals: M = .584, p = .001; Baseline: M = 2.17).

We next turned to our behavioral metrics to understand user distraction. We found a significant main effect of interface design on number of distracted gaze shifts, $F(3,44)$ = 40.28, $p < .001$, and on total distraction time, $F(3, 44)$ = 48.72, $p < .001$. Post-hoc tests showed that all three AR designs significantly decreased both the number and length of distractions compared to the baseline (all comparisons at $p < .0001$).

Subjective Results — Participants rated the several facets regarding their experiences teleoperating an aerial robot. We found a significant effect of interface design on user comfort working with the robot, $F(3, 44)$ = 8.12, $p < .001$, with users more comfortable using the Frustum (M = 5.58), p = .002, Callout (M = 5.87), $p < .001$, and Peripherals (M = 5.14), p = .019, designs than the current Freeflight interface (M = 3.64). We also found a significant effect of design on confidence operating the robot $F(3, 44)$ = 7.93, $p < .001$. Post-hoc comparisons against the baseline (M = 3.0) revealed that users
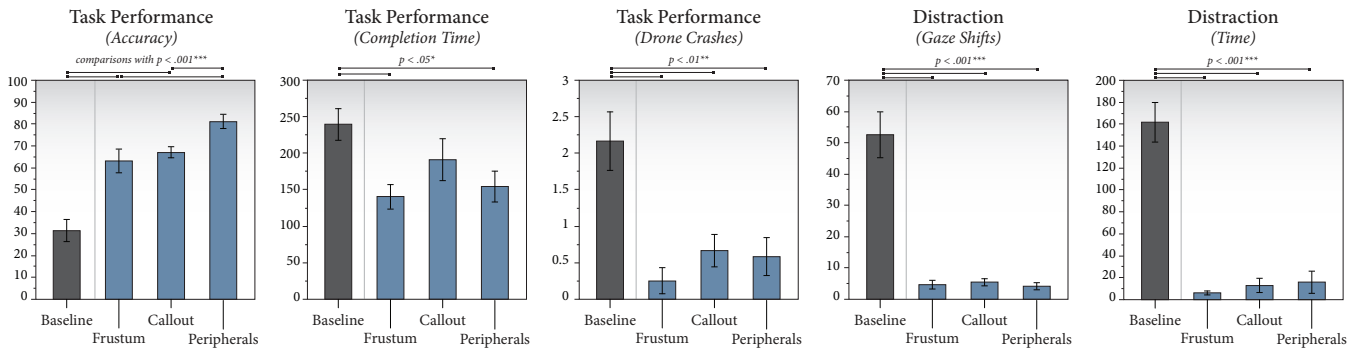
Figure 3: Objective results show that the augmented reality interface designs improved task performance in terms of accuracy and number of crashes, while minimizing distractions in terms of number of gaze shifts and total time distracted. (*), (**), and (***), denote comparisons with $p < .05$, $p < .01$, and $p < .001$ respectively.

were significantly more confident using the Frustum ($M = 5.02$) and Callout ($M = 5.0$) design, both $p < .001$, but not the Peripherals ($M = 4.0$), $p = .179$. Finally, we found a significant effect of design on perceived task difficulty, $F(3, 44) = 4.17$, $p = .011$, with participants finding the task significantly easier using the Frustum ($M = 4.03$), $p = .027$, Callout ($M = 4.13$), $p = .016$, designs and marginally easier using Peripherals ($M = 3.78$), $p = .082$, designs than the baseline ($M = 2.5$). Perceived interface usability was evaluated with the SUS. We found a significant effect of interface design on SUS total score, $F(3, 44) = 7.38$, p < .001. Tukey's HSD revealed each of our AR designs had significantly higher usability ratings over the baseline (Frustum: $M = 80.21$, $p < .001$; Callout: $M = 76.88$, $p = .003$; Peripherals: $M = 71.04$, $p = .041$; Baseline: $M = 55.0$).

6   DISCUSSION

Our new interface designs that provided users with augmented reality feedback while teleoperating an aerial robot demonstrated significant improvements over a modern interface that is representative of popular designs currently in use. We found that each of our designs enabled users to complete an inspection task faster and more accurately than with the robot's default control interface, while also leading to safer operation with fewer crashes. Overall, users rated our designs as largely more favorable in terms of usefulness and their own comfort and confidence when operating the robot. It is possible that differences in using a physical Xbox controller rather than touchscreen controls could be partially responsible for our results. However, we do not believe this to be the case, as we took great care to match the Xbox control mappings and sensitivities to the touchscreen controls provided by the Freeflight Pro application. Moreover, we are unaware of any prior research that suggests simply altering whether a controller is physical or touch-based might lead to results as extreme as we found.

Instead, we believe the reason behind the success of our designs can principally be found within our behavioral measures: our interface designs enabled users to get live video feedback without taking their eyes off the robot, whereas current designs force users to make context switches that sacrifice either situational awareness of the robot in the environment or their ability to closely monitor the robot's camera feed at any given time.

In our experiment, as in real-world deployments, users needed an understanding of both of these aspects to successfully complete their task. This was especially important as participants attempted to inspect the smaller, pink target that was closer to the ground. The size and placement of this target required that participants navigate fairly close to both the wall and ground, which can lead to difficulties for the robot's internal stabilization mechanism, potentially causing the robot to drift while hovering. Observing our experimental recordings, we see that participants using the Frustum, Callout, and Peripherals designs were able to quickly notice drift and re-stabilize the robot. However, participants in the baseline condition often took much longer to realize the robot was drifting as they were staring at the tablet which provided the robot video feed, rather than monitoring the robot itself. By the time these participants noticed the drift, it was often too late to correct, leading to crashes, landings, or loops of overcorrection, giving participants the impression that the robot was difficult to control:

P8[Baseline]: *"I found it fairly difficult and had a hard time getting precise shots with the robot. At times, it seemed that the drone was moving out of my control and I was forced to land it"*
P15[Baseline]: *"The environment interfering with how the robot could move was the biggest challenge; when it got too close to walls it would spin or move unpredictably. The precise task was hard due to this interference."*
P16[Baseline]: *"It was difficult to make minor adjustments and get close to the wall"*

On the other hand, participants in the AR conditions, operating the exact same robot with the same flight capabilities and stabilization limitations, found operation much easier:

P12[Frustum]: *"It was easier than I had thought..."*
P20[Callout]: *"That wasn't so hard..."*
P40[Peripherals]: *"Flying was very easy...I felt very comfortable."*

In addition to this improvement in overall utility, our results also revealed an interesting tradeoff between objective performance and subjective user preference. Overall, participants used the Peripherals design to achieve significantly better task performance than in any other condition. However, users felt no more comfortable

using Peripherals than they did using the baseline interface, felt that Peripherals only made the task marginally easier, and gave the Peripherals the lowest SUS score of the three AR designs. On the other hand, the Frustum and Callout design still offered significant improvements over the baseline in terms of task performance, but were rated much better in terms of subjective preferences. This result might inform future designs: for non-critical use cases (e.g., hobby flying, entertainment purposes, etc.), a Frustum or Callout design might improve both objective and subjective usability; however, for mission-critical applications, a Peripheral design might optimize performance at a small cost to perceived ease-of-use.

6.1 Limitations and Future Work

While our results are promising, we recognize that this study represents only an initial exploration into the design of AR interfaces for robot teleoperation. Our participants gave us several suggestions to further improve our designs, often involving the lack of a visual indication for when an image had been captured. Our systems could easily be modified to provide such feedback, possibly leveraging the idiom now common in mobile devices where images briefly freeze to indicate a photo has been taken. Many participants found this missing feature to be a downside, indicating that teleoperation interface designers should take special care in including this functionality in future systems (current interfaces, including the Freeflight Pro used as our baseline, also lack this feature):

P24[Callout]: *"...the lack of a visual indicator when a picture was taken meant I didn't know if I'd done so or not."*
P27[Peripherals]: *"It was fairly easy to line up the rectangles for a brief moment, but found the lack of feedback after taking a photo difficult; I didn't know whether the photo had actually been captured upon pressing 'B' on the remote."*
P30[Frustum]: *"I felt like I was blind with respect to whether I was capturing a good image with the appropriate field of view."*

Aside from such improvements to our current designs, we also recognize that our models represent only a small subset of the rich design space regarding how interface designers might leverage AR feedback to improve robot teleoperation. We found our framework, which divides models across augmenting the environment, the robot, or the user interface useful in inspiring and structuring our design choices, however this framework may need further extensions and development to truly serve as an optimal taxonomy or predictive model for the AR-HRI design space. Other perspectives may also prove useful, for example thinking about "what" feedback may be presented to the user, compared with "how" such information is rendered. In this work, we focused on better ways of transmitting information about what is in the robot's field of view, as this will often be the most critical information for robot operators, and examined these effects in isolation to determine potential tradeoffs that might inspire the design of future systems. However, there is no reason our designs couldn't be combined (e.g., provide a Frustum and Peripheral feedback). Moreover, there is a variety of other information that might be important to convey, such as robot battery life, GPS coordinates, or readings from onboard sensor payloads.

As an additional limitation, we explicitly scoped this research within examining collocated aerial robot teleoperation. This decision

was based on three considerations: line-of-site is currently mandated for aerial robots, interactions involving collocated flight are already common among hobbyists, and our collocated inspection task mimics the context of many envisioned aerial robot deployments, such as free-flyers operating within factories [37, 38] or the International Space Station [2]. Although we used an aerial robot in this experiment, we believe that our results could generalize to many other robots (e.g. ground robots, robotic arms, etc.). Future systems might explore the broader generalizability of our designs, for instance examining collocated interaction at longer distances, exploring how to convey additional robot data besides camera information, investigating augmented reality feedback for remote teleoperation, or studying scalability for operating multiple robots. Finally, future research might improve certain limitations of our current implementation. For instance, future work might remove our reliance on motion capture for precise robot localization by integrating our designs with simultaneous localization and mapping (SLAM) software. Alternatively, studies might improve upon our networking solution to increase the framerate of the Callout and Peripheral designs while presenting higher resolution images.

7 CONCLUSIONS

In this work, we explored the design of augmented reality interfaces that support collocated robot teleoperation of aerial robots. We propose a framework for structuring the design of AR interfaces that support HRI and use this framework to develop three new teleoperation models that leverage ARHMD technology to provide visual feedback on robot camera capabilities. We conducted a 48-participant user study comparing these designs against a baseline teleoperation system without AR feedback that is currently in common use. We found that our designs significantly improved objective measures of teleoperation performance and speed while reducing crashes. We attribute this enhanced performance to the support our systems provide for synthesizing information about the physical location of the robot with robot camera information, compared to traditional interfaces whose use leads to divided attention interactions. We believe that our designs represent a new paradigm that can reduce the detrimental effects of such constant context switching. Our subjective results revealed an interesting tradeoff, where participants preferred designs that moderately improved performance over the best-performing design. Overall, our results show that AR technologies hold great promise for mediating human-robot interactions and we hope our research can serve as a catalyst for additional explorations within this space.

8 ACKNOWLEDGMENT

REFERENCES

[1] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47.
[2] Maria Bualat, Jonathan Barlow, Terrence Fong, Christopher Provencher, Trey Smith, and Allison Zuniga. 2015. Astrobee: Developing a Free-flying Robot for

the International Space Station. In *AIAA SPACE 2015 Conference and Exposition*. 4643.

[3] Jessie YC Chen, Ellen C Haas, and Michael J Barnes. 2007. Human Performance Issues and User Interface Design for Teleoperated Robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 6 (2007), 1231–1245.

[4] Mark Corcoran. 2014. Drone Journalism Takes Off. http://www.abc.net.au/news/2012-02-21/drone-journalism-takes-off/3840616. (February 2014). Accessed: 2018-01-05.

[5] Enrico Costanza, Samuel A Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. 2006. Eye-q: Eyeglass Peripheral Display for Subtle Intimate Notifications. In *Proceedings of the ACM Conference on Human-Computer Interaction with Mobile Devices and Services*. 211–218.

[6] Mike Daily, Youngkwan Cho, Kevin Martin, and Dave Payton. 2003. World embedded interfaces for human-robot interaction. In *System Sciences (HICSS'03), 2003. Proceedings of the 36th Annual Hawaii International Conference on*. IEEE, 6–pp.

[7] Anca D Dragan and Siddhartha S Srinivasa. 2013. A Policy-blending Formalism for Shared Control. *The International Journal of Robotics Research* 32, 7 (2013), 790–805.

[8] Jill L Drury, Justin Richer, Nathan Rackliffe, and Michael A Goodrich. 2006. *Comparing Situation Awareness for Two Unmanned Aerial Vehicle Human Interface Approaches*. Technical Report. DTIC Document.

[9] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the World: 2D Windows for 3D Augmented Reality. In *Proceedings of the 6th annual ACM Symposium on User Interface Software and Technology (UIST'93)*. 145–155.

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

[11] Foyle, David C and Andre, Anthony D and Hooey, Becky L. 2005. Situation Awareness in an Augmented Reality Cockpit: Design, Viewpoints and Cognitive Glue. In *Proceedings of the 11th International Conference on Human Computer Interaction*. 3–9.

[12] Michael A Goodrich, Bryan S Morse, Damon Gerhardt, Joseph L Cooper, Morgan Quigley, Julie A Adams, and Curtis Humphrey. 2008. Supporting Wilderness Search and Rescue using a Camera-equipped Mini UAV. *Journal of Field Robotics* 25, 1-2 (2008), 89–110.

[13] Scott A Green, Mark Billinghurst, XiaoQi Chen, and J Geoffrey Chase. 2008. Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. *International Journal of Advanced Robotic Systems* 5, 1 (2008), 1.

[14] Scott A Green, J Geoffrey Chase, XiaoQi Chen, and Mark Billinghurst. 2009. Evaluating the Augmented Reality Human-Robot Collaboration System. *International Journal of Intelligent Systems Technologies and Applications* 8, 1-4 (2009), 130–143.

[15] Drew Harwell. 2014. 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).

[16] Keita Higuchi, Katsuya Fujii, and Jun Rekimoto. 2013. Flying Head: A Head-synchronization Mechanism for Flying Telepresence. In *International Conference on Artificial Reality and Telexistence (ICAT'13)*. 28–34.

[17] Yoshio Ishiguro and Jun Rekimoto. 2011. Peripheral Vision Annotation: Noninterference Information Presentation Method for Mobile Augmented Reality. In *Proceedings of the ACM Augmented Human International Conference*. 8.

[18] J Richard Landis and Gary G Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1977), 159–174. Issue 1.

[19] Paula Lavigne. 2014. Eyes in the Sports Sky. ESPN website. (2014). Accessed: 2018-01-05.

[20] Li, Hua and Zhang, Xin and Shi, Guangwei and Qu, Hemeng and Wu, Yanxiong and Zhang, Jianping. 2013. Review and Analysis of Avionic Helmet-Mounted Displays. *Optical Engineering* 52, 11 (2013), 110901–110901.

[21] Jeffrey I Lipton, Aidan J Fay, and Daniela Rus. 2018. Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing. *IEEE Robotics and Automation Letters* 3, 1 (2018), 179–186.

[22] M Alejandra Menchaca-Brandan, Andrew M Liu, Charles M Oman, and Alan Natapoff. 2007. Influence of Perspective-taking and Mental Rotation Abilities in Space Teleoperation. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI'07)*. 271–278.

[23] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems* 77, 12 (1994), 1321–1329.

[24] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. 1993. Applications of Augmented Reality for Human-Robot Communication. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*, Vol. 3. 1467–1472.

[25] Masmoudi Mostefa, L Kaddour El Boudadi, A Loukil, Khelf Mohamed, and Dahane Amine. 2015. Design of Mobile Robot Teleoperation System based on Virtual Reality. In *IEEE International Conference on Control, Engineering & Information Technology (CEIT'15)*. 1–6.

[26] Robin R Murphy. 2014. *Disaster Robotics*. MIT Press.

[27] Nicholl, Ryan. 2014. Airline Head-up Display Systems: Human Factors Considerations. *Available at SSRN:https://ssrn.com/abstract=2384101* (2014).

[28] Curtis W Nielsen, Michael A Goodrich, and Robert W Ricks. 2007. Ecological Interfaces for Improving Mobile Robot Teleoperation. *IEEE Transactions on Robotics* 23, 5 (2007), 927–941.

[29] Shahin S Nudehi, Rohan Mukherjee, and Moji Ghodoussi. 2003. A Haptic Interface Design for Minimally Invasive Telesurgical Training and Collaboration in the Presence of Time Delay. In *Proceedings. 42nd IEEE Conference on Decision and Control*, Vol. 5. 4563–4568.

[30] Allison M Okamura. 2004. Methods for Haptic Feedback in Teleoperated Robot-assisted Surgery. *Industrial Robot: An International Journal* 31, 6 (2004), 499–508.

[31] Christos Papachristos and Kostas Alexis. 2016. Augmented Reality-Enhanced Structural Inspection Using Aerial Robots. In *IEEE International Symposium on Intelligent Control (ISIC'16)*. IEEE, 1–6.

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

[33] David Pitman and Mary L Cummings. 2012. 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* 2012 (2012), 18.

[34] Benjamin Poppinga, Niels Henze, Jutta Fortmann, Wilko Heuten, and Susanne Boll. 2012. AmbiGlasses-Information in the Periphery of the Visual Field. In *Mensch & Computer*. 153–162.

[35] Morgan Quigley, Michael A Goodrich, and Randal W Beard. 2004. Semi-autonomous human-UAV interfaces for fixed-wing mini-UAVs. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04)*, Vol. 3. IEEE, 2457–2462.

[36] Emanuele Ruffaldi, Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli. 2016. Third Point of View Augmented Reality for Robot Intentions Visualization. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. 471–478.

[37] Daniel Szafir. 2015. *Human Interaction with Assistive Free-Flying Robots*. Ph.D. Dissertation. The University of Wisconsin-Madison.

[38] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2017. Designing Planning and Control Interfaces to Support User Collaboration with Flying Robots. *The International Journal of Robotics Research* 36, 5–7, 514–542.

[39] Russell H Taylor, Arianna Menciassi, Gabor Fichtinger, Paolo Fiorini, and Paolo Dario. 2016. Medical Robotics and Computer-Integrated Surgery. In *Springer Handbook of Robotics*. 1657–1684.

[40] John Thomason, Photchara Ratsamee, Kiyoshi Kiyokawa, Pakpoom Kriangkomol, Jason Orlosky, Tomohiro Mashita, Yuki Uranishi, and Haruo Takemura. 2017. Adaptive View Management for Drone Teleoperation in Complex 3D Structures. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI'17)*. ACM, 419–426.