



Virtual, Augmented, and Mixed Reality for Human-robot Interaction: A Survey and Virtual Design Element Taxonomy

MICHAEL WALKER, University of North Carolina at Chapel Hill, United States

THAO PHUNG, Colorado School of Mines, United States

TATHAGATA CHAKRABORTI, IBM Research AI, United States

TOM WILLIAMS, Colorado School of Mines, United States

DANIEL SZAFIR, University of North Carolina at Chapel Hill, United States

Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAM-HRI) has been gaining considerable attention in HRI research in recent years. However, the HRI community lacks a set of shared terminology and framework for characterizing aspects of mixed reality interfaces, presenting serious problems for future research. Therefore, it is important to have a common set of terms and concepts that can be used to precisely describe and organize the diverse array of work being done within the field. In this article, we present a novel taxonomic framework for different types of VAM-HRI interfaces, composed of four main categories of virtual design elements (VDEs). We present and justify our taxonomy and explain how its elements have been developed over the past 30 years as well as the current directions VAM-HRI is headed in the coming decade.

CCS Concepts: • **Human-centered computing** → *Virtual reality; Mixed/augmented reality; User interface design*; • **Computer systems organization** → **External interfaces for robotics**;

Additional Key Words and Phrases: Human-robot interaction, human-computer interaction, robots, virtual reality, mixed reality, augmented reality, interface design

ACM Reference format:

Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafrir. 2023. Virtual, Augmented, and Mixed Reality for Human-robot Interaction: A Survey and Virtual Design Element Taxonomy. *ACM Trans. Hum.-Robot Interact.* 12, 4, Article 43 (July 2023), 39 pages.

<https://doi.org/10.1145/3597623>

1 VIRTUAL, AUGMENTED, AND MIXED REALITY FOR HUMAN-ROBOT INTERACTION

Although robots are poised to increasingly support human society across a multitude of critical industries (e.g., healthcare, manufacturing, space exploration, agricultural), robot usage has remained limited due to the difficulties of robot control, supervision, and collaboration. A major source of this limitation arises from the bi-directional challenge of human-robot communication. Robots are often found to be incomprehensible, and humans struggle to predict robot capabilities

This work was funded in part by NSF grant IIS-1909864 and NSF award 1764092.

Authors' addresses: M. Walker and D. Szafrir, 232 S Columbia St, 209 Sitterson Hall, Computer Science Department, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514; emails: michael.e.walker@unc.edu, daniel.szafrir@cs.unc.edu; T. Phung and T. Williams, 1600 Illinois Street, Brown Building 280-J, Colorado School of Mines, Golden, CO 80401; emails: {thaophung, twilliams}@mines.edu; T. Chakraborti, IBM, AI Composition Lab, Cambridge, MA USA; email: tchakra2@ibm.com.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2023 Copyright held by the owner/author(s).

2573-9522/2023/07-ART43

<https://doi.org/10.1145/3597623>

ACM Transactions on Human-Robot Interaction, Vol. 12, No. 4, Article 43. Publication date: July 2023.

or intentions. Simultaneously, robots lack the ability to reason about complex human behaviors: a skill inherently required for effective collaboration with humans.

At the heart of this problem lies the issue of poor information exchange between humans and robots, where neither can understand what the other is explicitly or implicitly conveying. This is analogous to the *Gulf of Execution* and *Gulf of Evaluation* concepts within the *Human Action Cycle*, a proposed model describing human interactions with complex systems from the cognitive engineering and human-computer interaction communities [66]. Humans regularly have issues conveying their high-level goals as inputs a robot can understand (e.g., gulf of execution), while robots often provide ineffective or no feedback to allow humans to assess the robotic system's state (e.g., gulf of evaluation).

An example gulf of evaluation found in human-robot interactions is the motion inference problem, where robot users find the task of predicting when, where, and how a robot teammate will move to be difficult due to lack of information communicated from the robot. Information regarding a robot's planned movement is often invisible to human users, and even in circumstances where a robot is made to communicate its movement intent, the robot may lack the ability to share its motion plans as a human teammate would. A large amount of work in **human-robot interaction (HRI)** has looked at addressing this gulf of evaluation, such as by having robots use human-inspired social cues (e.g., gaze, gestures) to communicate their intentions [78], altering robot trajectories to be more legible or expressive [24, 90], adding light-based signals [7, 26, 91], or using various other means such as auditory indicators [93]. Although such techniques have shown effectiveness in reducing this gulf of evaluation by making robot motion more predictable, common constraints (e.g., computational, platform, environmental) may limit their feasibility when deployed in the real world. For example, an aerial robot's morphology would prevent it from performing hand gestures or gaze, a dynamic or cluttered environment may restrict robots from altering from optimal trajectories to be more legible, and robot auditory indicators may be rendered ineffective if deployed by a noisy robot (e.g., aerial robots) or in a noisy environment.

To mitigate these issues, new methods of human-robot communication are being explored by HRI researchers that leverage more than the verbal or non-verbal cues seen in traditional human communication. New forms of visual communication have shown great promise in enhancing human-robot interaction, ranging from enhanced graphical displays to improve robot control that reduce gulfs of execution to LED lights that communicate various robot signals [91] that reduce gulfs of evaluation. Recently, the rise of consumer-grade, standardized **virtual, augmented, and mixed reality (VAM)** technologies (including the iPad, Microsoft HoloLens, Meta 2, Magic Leap, Oculus Rift/Quest, HTC Vive, etc.) has created a promising new medium for information exchange between users and robots and is well suited to enhance human-robot interactions in a variety of ways. VAM interfaces allow users to see 3D virtual imagery in a virtual space or contextually embedded within their environment. Up to this point, robot users have been forced to use traditional 2D screens to analyze the rich 3D data a robot often collects about its environment. VAM technology can also be used hands-free when in the form of a **head-mounted display (HMD)** that allows for more fluid and natural interactions with robots in a shared physical environment. Users can also be immersed in purely virtual worlds and interact with virtual robots, which allows HRI researchers to evaluate interactions that would otherwise be impossible to observe either due to safety concerns or lack of access to an expensive physical robot(s). Finally, VAM interfaces allow HRI researchers to record and analyze human-robot interactions unlike ever before by leveraging the body, head, and gaze tracking inherent with VAM HMDs.

This article traces the development of early work merging HRI and VAM technology (which, while promising, was often hampered by limitations in underlying VAM technologies) and highlights more recent work that leverages modern systems. While this interdisciplinary surge of

research is exciting and valuable, in research fields it can raise disciplinary challenges. For example, many researchers are likely (unbeknownst to each other) to be simultaneously working on similar problems or using similar techniques, while potentially using significantly different terminology and conceptual frameworks to ground and disseminate their work. Such a lack of shared awareness and shared terminology may introduce problems for future research. Even the most basic terms used within this new wave of VAM-HRI work, that might naturally be assumed to have obvious, commonly agreed upon meanings—words such as “virtual reality,” “augmented reality,” “mixed reality,” “waypoint,” and “digital twin”—are rendered imprecise by the multifarious uses of mixed reality visualizations and of virtual and mixed reality systems. It is thus critical for the research community to have a common set of terms and concepts that can be used to accurately and precisely describe and organize the wide array of work being done within this field.

In this article, we present a taxonomy for VAM-HRI 3D command sequencing paradigms and *Virtual Design Elements (VDEs)* to provide researchers with a shared, descriptive basis for characterizing the types of systems HRI researchers are developing and deploying in both mixed and virtual reality systems. This taxonomy is the result of surveying 175 papers published in 99 conferences and journals spanning over the years 1991 to 2019 and is informed by recent (but nascent) attempts [101] to grapple with the breadth and complexity of this field through the series of *Virtual, Augmented, and Mixed-Reality for Human Robot-Interaction (VAM-HRI)* workshops held in conjunction with the ACM/IEEE International Conference on Human-Robot Interaction [103]. A complete categorized listing of the papers surveyed can be found at our VAM-HRI taxonomy companion website and is accessible at: vamhri.com. We collected papers through Google Scholar using combinations of the following set of search terms:

- Human Robot Interaction/HRI/VAM-HRI/Robotics
- Robot(s)/Robot Arm(s)/Mobile Robot(s)/Ground Robot(s)/Aerial Robot(s)/UAV(s)/Autonomous Vehicle(s)
- Augmented Reality/Mixed Reality/Augmented Virtuality/Virtual Reality/Diminished Reality
- Head-Mounted Display/HMD/CAVE/Projector/HoloLens/Oculus/Vive/Tablet

We supplemented the search results with papers from the VAM-HRI workshop series, which is where late-breaking work on this topic has been published for the past six years (since 2018). For all papers in this collection, we performed citation tracing, considering all papers citing or cited by these papers. We used citation tracing to perform a depth-limited breadth-first search for additional relevant papers, stopping at a depth of 2. We additionally limited our survey to include papers that implemented functional systems while excluding papers that only proposed or planned VAM-HRI interface designs.

To generate our classification frameworks a thematic analysis was performed on all 175 papers included in our survey. Annotators coded each paper (with 100% overlap) based on the VDEs instance characteristics. These codes were then analyzed to reduce redundancies and extract hierarchical structures (for Category/Class/VDE identification). This process continued in an iterative fashion until the final taxonomy stabilized (i.e., no changes after re-coding all papers while yielding 100% annotator agreement).

Each category, class, and VDE within our proposed taxonomy aims to provide HRI researchers (both those working within VAM-HRI and otherwise) with the shared language necessary to advance this subfield along a productive and coherent path. Additionally, our goal in creating this taxonomy is not only to provide a shared language for researchers to use to describe and disseminate their work, but to also aid researchers in connecting with the host of complementary work being performed in parallel to their own. This taxonomy may enable researchers to better

understand the benefits of different types of virtual imagery, more quickly identify promising graphical representations across different domains and contexts of use, and build knowledge regarding what types of representations may currently be under- (or over-) explored.

2 VAM-HRI ADVANCEMENT OVER TIME

To begin to understand how the field of VAM-HRI reached its current state, one can trace the development of VAM technologies back to Sutherland's vision of "The Ultimate Display" (itself influenced by Vannevar Bush's conception of the Memex) [88] and later developments with the Sword of Damocles system [89]. The earliest major work leveraging VAM for HRI appears to date back to a push in the late 1980s and early 1990s with various work exploring robot teleoperation systems [9, 48]. Perhaps the most fully developed instance of these early systems was the ARGOS interface for augmented reality robot teleoperation [61]. While the ARGOS interface used a stereo monitor, rather than the head-mounted displays in vogue today, the system introduced several design elements for displaying graphical information to improve human-robot communication and introduced concepts such as virtual pointers, tape measures, tethers, landmarks, and object overlays that would influence many subsequent designs. Later developments throughout the 1990s introduced several other important concepts, such as the use of virtual reality for both actual robot control and teleoperator training [39], the integration of HMDs (including the first use of an HMD to control an aerial robot) [22], projective virtual reality where user "reaches through" a VR system to control a robot that manipulates objects in the real world [29], the rise of VAM applications for robotics in medicine and surgery [13], and continued work on ARGOS and ARGOS-like systems [60]. At a high level, major themes appear that focus on using VR for simulation or training purposes, VR and/or AR as new forms of information displays (e.g., for data from robot sensors), and VAM-based robotic control interfaces. While many of these developments appear initially promising, it is interesting to note that following an initial period of intense early research on HRI and VAM, later growth throughout the 1990s appears to have happened at a relatively stable rate, rather than rapidly expanding. In addition, efforts to take research developments beyond laboratory environments into commercial/industrial systems appear to have been largely unsuccessful (indeed, even today robot teleoperation interfaces are still typically based on standard 2D displays rather than leveraging VAM).

In recent years, research in the field of VAM-HRI has seen explosive growth. This recent explosion is due in part to the emergence of commercial head-mounted displays (HoloLens, Vive, Oculus, etc.) as well as enhanced computer performance, which together have shifted the field from one requiring specialized hardware developed in research labs to one in which significant advances can be made immediately using standardized, inexpensive hardware systems that provide expansive software development libraries.

The rise of new VAM-HRI research has demonstrated the potential for **virtual reality (VR)**, **augmented reality (AR)**, and **mixed reality (MR)** across a number of application domains that are of significant historical interest to the field of HRI, including education [17, 32], social support [37, 72], and task-based collaboration [16, 27], while illustrating how MR and VR can be used as new tools in a HRI researcher's toolkit for robot design [14], human-subject experimentation [98, 100], and robot programming and debugging [40].

AR visualization techniques offer significant promise in HRI due to their capability to effortlessly communicate robot intent and behavior. For example, researchers have shown how AR can be used to visualize various aspects of robot navigational state, including heading, destination, and intended trajectory [95]. Similarly, others have shown how AR can be used to visualize a robot's beliefs and perceptions by displaying both exteroceptive (e.g., laser range finder data) and proprioceptive (e.g., battery status) sensor data [6, 14]. Such virtual imagery can improve the situational

awareness of human teammates during human-robot collaboration and teleoperation, making interaction more fluid, intuitive, safe, and enjoyable.

VR techniques, however, provide unique opportunities for safe, flexible, and novel environments in which to explore HRI. For instance, researchers have explored how virtual environments can help humans learn how to work with new robots [70] and help robots learn new skills [25, 44]. Using VR interfaces for these purposes reduces risks, addresses spatial and monetary limitations, and promotes new opportunities to experiment with new (or not yet physically feasible) robots and environments. Moreover, VR interfaces can be used to visualize the real world in new ways, allowing teleoperators and supervisors to view robots' real environments through video streams or point cloud sensor displays in a more immersive and intuitive manner than traditional 2D displays [12, 87]. In addition, VR systems may enable users to exert more fine-grained control over teleoperated robots, including unmanned aerial and nautical vehicles, by leveraging related technologies such as head tracking, haptic controllers, and tactile gloves [42].

3 VAM INTERFACES FOR ROBOTICS

The advancements in hardware accessibility have created a host of new opportunities for exploring VAM technology as an interaction medium for enhancing various aspects of HRI. Various VAM displays, degrees of reality, and coordinate system calibration techniques, and VAM-HRI interface paradigms have all been used in different combinations to successfully apply VAM interfaces to robotics.

3.1 The Reality-Virtuality Continuum

VAM displays are computer displays that either immerse users in an entirely synthetic world or merge both the real world and a synthetic world. Note that, although VAM refers to more than just visual senses (i.e., haptic, auditory, etc.), for the duration of this article, VAM will refer only to synthetic imagery.

VAM technology is capable of mixing varying degrees of reality and virtuality. As observed by Milgram et al. in 1994, all VAM displays fall upon a "Reality-Virtuality Continuum" (See Figure 1) [59]. This taxonomy has served as a useful tool for classifying VAM interfaces as well as coined the term "Mixed Reality." Per the Reality-Virtuality Continuum, interfaces that place users in environments consisting of only synthetic imagery are considered "**Virtual Reality**" (VR), accordingly interfaces that only consist of real imagery are considered based in reality. This leaves a middle ground, where synthetic and real imagery are combined to form the space of "**Mixed Reality**" (MR). Within the design space of MR lie two sub-categories of "**Augmented Reality**" (AR), where synthetic imagery is added to a real environment, and "**Augmented Virtuality**" (AV), where real imagery is added to a synthetic environment. Together with VR, MR, consisting of AR and AV, one can categorize VAM interfaces and provide more specific design guidelines for HCI interfaces.

3.2 Display Hardware

VAM interfaces can be implemented by various display hardware. The following is a list of common VAM display types:

2D Monitor Video Displays: 2D monitor video displays provide a means of overlaying AR synthetic imagery on real images or video feeds, as "window-on-the-world" displays. Due to the inherent 2D nature of traditional monitors, users of these displays are not able to experience stereoscopic depth perception when viewing interfaces on these displays. Additionally, these displays tether users to computer terminals and do not allow for a hands-free, free-roam of environments in which AR imagery are being added.



Fig. 1. Milgram's reality-virtuality continuum is a continuous scale ranging between *reality* (no synthetic imagery) and *virtual reality* (only synthetic imagery). Between the two ends of the continuum is *mixed reality*, where real and synthetic imagery are combined with either synthetic imagery added to a real environment as *augmented reality* or real imagery added to a synthetic environment as *augmented virtuality* [59].

3D Monitor Video Displays: 3D monitor video displays are similar to the above AR 2D monitor video displays, with the difference being that users can see interfaces and data with stereoscopic depth perception. Users can experience this depth perception with either glasses-free 3D monitors or with 3D glasses (i.e., anaglyph, polarized, or active shutter).

Projectors: Projector-based displays provide AR synthetic imagery to a user's environment, that users can see without the need of special equipment, such as glasses or **head-mounted displays (HMDs)**, and also allow users to freely roam their environment. Drawbacks to this display type include: The projected imagery is inherently 2D; occlusions (from the environment, users, or robots) disrupt and/or block the AR imagery; and the projected imagery can get washed out in environments that are too bright.

Tablet Displays: Tablet displays are also similar to AR 2D monitor video displays in that they can add AR imagery to a live camera feed; however, a major difference between the two displays is that tablets are highly mobile and allow users to freely explore their surrounding environment. Tablets provide users with touchscreen-based interactions and, due to the portable nature of these displays, allow users to visualize virtual objects and data contextually within the user's and/or robot's environment.

Optical See-through HMDs: Optical see-through HMDs present AR through transparent lenses, allowing users to see virtual imagery contextually overlaid on the real world. Similarly to tablet displays, the highly mobile optical see-through HMDs allow users to freely navigate their environment; however, unlike tablets, this exploration can be performed as a hands-free experience. Unfortunately, current optical see-through HMD technology restricts the viewing of virtual imagery to a narrow field-of-view. Additionally, the imagery provided by these HMDs is easily washed out in bright lights, including outdoor sunlight.

Video Pass-through HMDs: Video pass-through HMDs are unique in that they can provide AR, AV, and VR experiences to users. By mounting a stereo cameras to the front of a VR HMD, video pass-through HMD's can pipe video imagery from the real outside world to the dual lens within the HMD. The video feed can be intercepted between camera and lens, allowing for virtual imagery being added to the captured image frames. This process of displaying AR imagery allows these HMD displays to show AR imagery in environments with bright light, such as the outdoors.

VR HMDs: VR HMDs are 3D stereoscopic displays worn on users' heads to fully immerse users in VR or AV environments. Recent advances in VR HMD technology have allowed users to not only

control the display with head motion, but with hand and body motion as well, permitting users to be hands-free while able to be free to walk around synthetic environments naturally with their own body. A downside to these displays is that they close off users from the real world without an option of seeing reality from the display while worn.

CAVES: CAVES, or **Cave Automatic Virtual Environments**, are VR displays where users are placed in a three- to six-walled environment. The walls of the environment display 3D imagery when paired with 3D glasses, allowing users to be immersed in AV or VR environments. CAVES allow for multiple users to experience the VR or AV environments, but unfortunately are expensive, confining, and immovable.

3.3 VAM Interface Coordinate Frame Rectification

Finally, to successfully merge both reality and virtual reality in a single interface and to have VAM imagery appear in its appropriate position and orientation within the world, a singular coordinate frame must be maintained to manage the positions of the various virtual imagery. However, VAM interfaces for HRI present the unique challenge of multiple physical agents (users and robots) each having their own unique perspectives and coordinate frames. These agents' frames are also often moving and changing at any given times, requiring real-time tracking of users, robots, and objects within an environment. Research has explored the following various methods for extracting and unifying coordinate frames:

Fiducial Markers: Fiducial markers are images that have their optical properties known by a VAM interface beforehand and act as visual reference points for the systems. When a VAM interface's camera detects a marker, it can determine the marker's relative pose to that of the camera. These markers are often placed in a robot's environment, on robots, or on objects within a robots environment. This method of frame rectification is popular due to its high portability and low cost [11, 28, 36, 49].

Motion Capture Cameras: Motion capture cameras can obtain poses of robots, users, and objects in an environment with high precision by tracking patterns of infrared reflecting markers. Unfortunately, these systems are expensive and immovable, often making interfaces that use this method of frame rectification constrained to laboratory environments [37, 95, 96].

Odometry: Another method of unifying coordinate frames between user and robot is by using odometry (including visual odometry with computer vision algorithms such as SLAM [63]). If agent coordinate frames are initially synced at the initialization of a VAM-HRI interface, then odometry can track the relative pose changes agents have undergone over time, which can be used to maintain a rectified coordinate system [34, 73].

Virtual Object Manual Alignment: An additional method of coordinate frame rectification leverages the intelligence and capabilities of the robot operator through a process of manual alignment. Interfaces utilizing this technique require users to manually place a virtual object (controlling the object's position and rotation) so it overlaps with a physical target object that has a pose is already known to the robot. This method can be used to provide a rough estimate of the reference-frame transformation and can be further refined using additional frame rectification methods (such as Iterative Closest Point) [34, 73].

Iterative Closest Point Alignment: Coordinate frame rectification is also made possible by using point cloud alignment algorithms, such as **Iterative Closest Point (ICP)** [79], which can take

in separately collected point clouds and output the transformation between the two overlapping perspectives. This method requires all agents to have some overlap between their collected point clouds, which makes this method most suitable for initial frame synchronization between multiple agents that is followed by odometry-based frame rectification methods above [34, 73].

Machine Learning Image-based Pose Estimation: Relative poses between agents can also be found through machine learning algorithms, trained to estimate an object's pose through image frames. Current technology limits this method to known, simple objects; however, as machine learning algorithms strengthen [104] it is feasible to imagine this method becoming more popular for agent frame rectification, especially if it means fiducial markers or motion capture cameras can be eliminated [10].

4 DESIGN PARADIGMS OF VAM-HRI INTERFACES

Although VAM-HRI is still a relatively nascent field, recurring higher-level implementation paradigms have emerged for designing robot interfaces that utilize VAM technologies. This can be seen especially in the cases of HMD teleoperation interfaces for remote robots (as classified by Lipton et al. [55]) and robot 3D command sequencing, a classification scheme that we propose in this work.

4.1 Remote Robot Teleoperation HMD Interfaces

The recent advent of mass-produced HMD technology has seen a rise in HMD interfaces that mediate remote robot teleoperation, for either navigational tasks or manipulation tasks. As noted by Lipton et al. [55], these interfaces can fall into three classes: Direct Interfaces, Virtual Control Room Interfaces, and Cyber-physical Interfaces.

Direct HMD Teleoperation Interfaces: Direct HMD teleoperation interfaces live-stream stereo video feeds from remote robots to the HMD's lenses. This process allows users to see from the robot's "eyes" with immersive 3D stereoscopic vision as if they were embodying the remote robot, especially if user head motions control robot head motions with a one-to-one mapping. These HMD-based interfaces have shown to significantly improve robot teleoperation tasks. Additionally, if virtual imagery is overlaid on the video stream, then the interfaces become a AR interface [38]. However, there is a significant drawback associated with Direct HMD Interfaces that stems from both miscues from the user's proprioceptive system and the communication delays inherently found with current network technology. When a remote robot moves and the local user does not move, the user's proprioceptive system receives conflicting cues (visual cues of movement vs. no body motion detected) causing nausea. The same happens in reverse when a local user turns their head and robot's head does not immediately turn to match the movement (due to mechanical limitations or communication delays), which creates conflicting proprioceptive cues (no visual cues of movement vs. body motion detected) that also cause nausea.

AV HMD Teleoperation Interfaces: To mitigate the nauseating effects of the Direct HMD Interfaces, two AV HMD Teleoperation Interface paradigms have arisen from research in the VAM-HRI field: Virtual Control Room and Cyber-Physical Interfaces [55]. In both interface styles, the state of the user's eyes are decoupled from the robotic systems state to remove the conflicting proprioceptive system cues. By placing the user in a AV environment, the users' eyes are represented by virtual cameras in the virtual space that move freely with the user's head and body movements. This decoupling method helps mitigate nausea caused by communications/hardware delays and/or imperfect mappings between user head motion and robot head motion.

Virtual Control Room Model: In the Virtual Control Room Model, the user is placed in a virtual room that serves as a supervisory command and control center of a remote robot. Within the control room, the user is able to interact with displays and objects in the virtual space itself and can view 3D stereo video streams projected on the rooms walls, thus still allowing the user to still see from the robot's perspective with stereoscopic depth perception [50, 55] (see Figures 3(A) and 4(C)).

Cyber-physical Model: In the Cyber-physical Model, a shared AV virtual space is created (typically with a one-to-one mapping) between: (1) a remote robot and a virtual environment; and (2) a human operator(s) and a virtual environment. Additionally, a 3D reconstruction of the robot's remote environment is rendered (typically with dense RGB point clouds) within the virtual environment to provide situational context and awareness to the human operator. A virtual robot replica of the remote physical robot is also added to the virtual environment in the same relative location within the virtual environment as in the real remote environment. This virtual robot mimics the remote real robot's state and actions. The user can also use the virtual robot to send commands to the remote real robot or visualize the current state or actions being undertaken by the physical robot. A benefit of this interface paradigm over that of the Virtual Control Room is that the user can freely change their viewpoint within the remote environment by walking around the AV environment 3D reconstruction, as they are not only provided the view from the robot's camera(s). However, the sense of immersion from virtually embodying the remote robot is lost [1, 76, 87] (see Figure 3(B)).

4.2 VAM-HRI Interface 3D Command Sequencing

Our VAM-HRI literature review revealed recurring high-level themes for robot 3D command sequencing. We propose the following three paradigms that capture these methods of controlling robots with VAM interfaces: *Direct Manipulation*, *Environment Markup*, and *Digital Twins*. Note that these methods are not mutually exclusive and can be used in isolation or conjunction (such as in Cao et al.'s AR-based robot prototyping and control system [14], which concurrently utilizes both Environment Markup and Digital Twins).

Direct Manipulation: The paradigm of Direct Manipulation leverages more traditional methods to directly control robots and utilizes 3D translation and/or rotation input from either physical or virtual source, commonly to send teleoperation commands to end effectors or navigational systems. Direct manipulation from physical inputs include body/head tracking and VAM-based 3D controllers [97]. Virtual controllers can act as metaphors for existing physical control input devices (i.e., levers, handles, joysticks, etc.) [36] or utilize novel designs unconstrained by physics such as floating control spheres [51].

Environment Markup: Under the Environment Markup paradigm, users send commands to robots by adding virtual annotations to a robot's environment. Examples of such environmentally anchored annotations include waypoints, trajectories, and planned future poses of manipulable objects. These annotations can take the form of a simple single command or can be chained/combined together to form a series of commands or a singular complex command [15, 43].

Digital Twins: In contrast to commands sent under the Environment Markup paradigm, Digital Twin command sequencing does not add virtual annotations to a robot's environment. Instead, command sequences are generated by the manipulation of a *digital twin*, which is a virtual replica (or representation) of a robot, object, or environment. For example, in the case of a robot with a digital twin, a user could press a virtual button on the robot's digital twin, at which point, the real

robot would respond as if its own physical button was pressed. Additionally, robot teleoperation can be achieved by directly manipulating the robot digital twin (arm, body, etc.), after which, the real robot imitates the action taken by the digital twin (e.g., the robot moves itself in the exact trajectory taken by the digital twin or moves its end effector to the final position taken by the digital twin). Real object manipulation is performed in a similar manner, but in this case, a robot mimics a user's object digital twin manipulations on the associated real object [28, 36, 52, 87].

5 VAM-HRI VIRTUAL DESIGN ELEMENT TAXONOMY

In this article, we propose a novel taxonomy for identifying and categorizing VAM-HRI **Virtual Design Elements (VDEs)** [101]. VAM-HRI VDEs are VAM-based visualizations that impact robot *interactivity* by providing new or alternate means of interacting with robots. VDEs can appear in two ways: (1) *user-anchored* – attached to points in the user's camera's coordinate system, unchanging as the user changes their field of view (see Figure 4(D)); or (2) *environment-anchored* – attached to points in the coordinate system of a robot or some other element of the environment, rather than the interface itself [101] (see Figure 3(C)).

Over the course of the following sections, we will detail each category and class of VDE within the taxonomy, which we organize into the following categories: *Virtual Entities*, *Virtual Alterations*, *Robot Status Visualizations*, and *Robot Comprehension Visualizations* (see Figure 2). Each VDE category was identified after surveying the aforementioned 175 VAM-HRI research papers and acts as a high-level categorical grouping of VDEs that share common purposes for enhancing and/or manipulating human-robot interactions with virtual or mixed reality. Note that VDEs are *instantiations* of the taxonomic sub-classes and can be used both in isolation and in synergistic conjunction by combining two or more VDEs (e.g., using a *Cosmetic Alteration* to communicate the *Internal Readiness* of a robot joint [6]).

5.1 Virtual Entities

The first category of VDEs we will examine are *Virtual Entities*: visualizations in which virtual entities, such as robots or objects, are added to real or virtual user environments. These VDEs can be a standalone VAM-based graphical elements that act as a visualization aid, input device, or simulation of an entity found in reality; however, they can also take the form of digital twins that are directly associated with a physical entity in the user and/or robot environment. We detail each of the three types of Virtual Entity classes below.

Virtual Entities – Robots

The *Robots* class of the *Virtual Entity* VDE category encompasses visualizations of robots that can be provided to users as either a visual tool for inspection or have full kinematic models and allow for complex interactions with users. These VDEs provide a level of immersion (a requirement for HRI simulations) unparalleled by simulations viewed on traditional 2D displays (such as Gazebo [57]), since the Functional Virtual Robots is visually integrated within the user's environment. We identify three sub-classes of such virtual robot entities:

Visualization Robots improve user understandings of a robot's current state or future actions but do not afford any two-way interactions with users (i.e., users cannot direct input to the graphical representation). These VDEs typically model a robot's 3D morphology, in whole or in part. Uses include providing users with a means of understanding a robot's current state in which case a virtual 3D model of a real robot mimics a physical robot's joint configurations in real time. Visualization Robots are particularly useful in situations of limited situational awareness when the user cannot directly see the physical robot (or portions of the robot), such as in remote

VAM-HRI Virtual Design Elements (VDEs)		
Category	Class	VDE
Virtual Entities	Robots	Visualization Robots
		Simulated Robots
		Robot Digital Twins
	Control Objects	Panels & Buttons
		Controllers
		Simulated Agents
	Environmental	Simulated Objects
		Simulated Environments
		Object Digital Twins
		Environment Digital Twins
Virtual Alterations	Superficial	Cosmetic Alterations
		Special Effect Alterations
	Morphological	Body Extensions
		Body Diminishments
		Form Transformations
	Robot Status Visualizations	Internal
Internal Readiness		
External		Robot Poses
		Robot Locations
Robot Comprehension Visualizations	Environment	External Sensor Purviews
		External Sensor Numerical Readings
		External Sensor Images & Videos
		External Sensor 3D Data
		Sensed Spatial Regions
		Robot Inherent Spatial Regions
		User-Defined Spatial Regions
	Entity	Entity Labels
		Entity Attributes
		Entity Locations
		Entity Appearances
	Task	Headings
		Waypoints
		Callouts
		Spatial Previews
		Trajectories
Alteration Previews		
Command Options		
Task Statuses		
Task Instructions		

Fig. 2. The VAM-HRI virtual design element (VDE) taxonomy table displays the complete listing of all VDEs in addition to the VDE category and class hierarchical structure. VDEs are VAM-based visualizations that enhance robot interactivity that can be implemented in isolation or conjunction with one another.

teleoperation tasks [50] (see Figure 3(A)). Visualization Robot VDEs can also provide users with a preview of proposed or planned robot motion by overlaying a 3D robot model onto the environment to allow for better means to assess how a robot will navigate through an environment and whether it will successfully travel to a desired location without collisions. Finally, Visualization Robots can also show the locations of robots that are partially or fully occluded by the environment, such as behind a wall or door [76] (see Figures 3(B) and 5(D)).

Simulated Robots are not linked to a specific instance of a physical robot and are instead independent robot simulations used to evaluate robots in fully virtual settings when using a real robot is not ideal, such as when robot hardware is unavailable, unsafe, has limited battery life, and/or faces physical depreciation when operated repeatedly (as required for sample-inefficient learning algorithms and/or interaction studies). These VDEs are also used for evaluating simulated interactions with autonomous or manually teleoperated robots, e.g., in situations where collocated interaction is infeasible or hazardous for humans (i.e., working near large industrial robots, testing space exploration robots in zero gravity environments, etc.). Simulated Robots are also useful for user training without putting robot hardware at risk of being damaged by inexperienced users and for training real robots through virtual-to-real-world transfer learning techniques [44] that require direct user interaction, such as learning from demonstration [4, 21, 58, 85] (see Figure 3(D)).

Robot Digital Twins operate in tandem with real robots and provide users with an immersive virtual robot that can be interacted with in lieu of a real, physical robot. By interacting first with a Robot Digital Twin, users can 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 virtual robot. For instance, mappings between the Robot Digital Twin and real robot include instantaneous duplication (the physical robot moves to match the Robot Digital Twin's position/attitude immediately), delayed duplication (the physical robot moves to match the Robot Digital Twin's position/attitude after a set period of time), confirmed duplication (the physical robot matches the Robot Digital Twin when triggered by the user), and more planning-oriented systems, such as using the Robot Digital Twin to denote waypoints or actions for future execution by the physical robot [36, 52, 87, 96] (see Figure 3(C)).

Virtual Entities – Control Objects

These VDEs represent virtual objects that users can interact with to send direct commands that control robotic systems. Control Objects can be 2D or 3D and can be user-anchored (such as 2D buttons on an AR tablet that remain in static positions regardless of where the display is pointed—see Figure 4(D)) or environment-anchored (such as a virtual 3D handles that remain fixed to a robot chassis—see Figure 5(F)). We identify two sub-classes of Control Objects that allow for developers to rapidly prototype and evaluate various interfaces and designs for robot input without procuring or engineering actual hardware:

Panels & Buttons are virtual control objects that look and act like panels and buttons found in real life and 2D GUIs (e.g., buttons, sliders, switches) [54] (see Figure 4(D)).

Controllers emulate physical 3D input devices that leverage the 3D capabilities of VAM displays. Controllers often act as metaphors for existing physical control input devices (i.e., levers, handles, joysticks, etc.). However, Controller VDEs have also opened a nascent design space that allows robot designers to create interface input devices that are unconstrained by physics in the form of objects unable to be implemented on Earth or in reality, such as manipulable control toruses [36] or floating control spheres [51], allowing for robot interactions that would otherwise be impossible to implement and/or evaluate with traditional non-VAM interfaces (see Figure 5(F)).

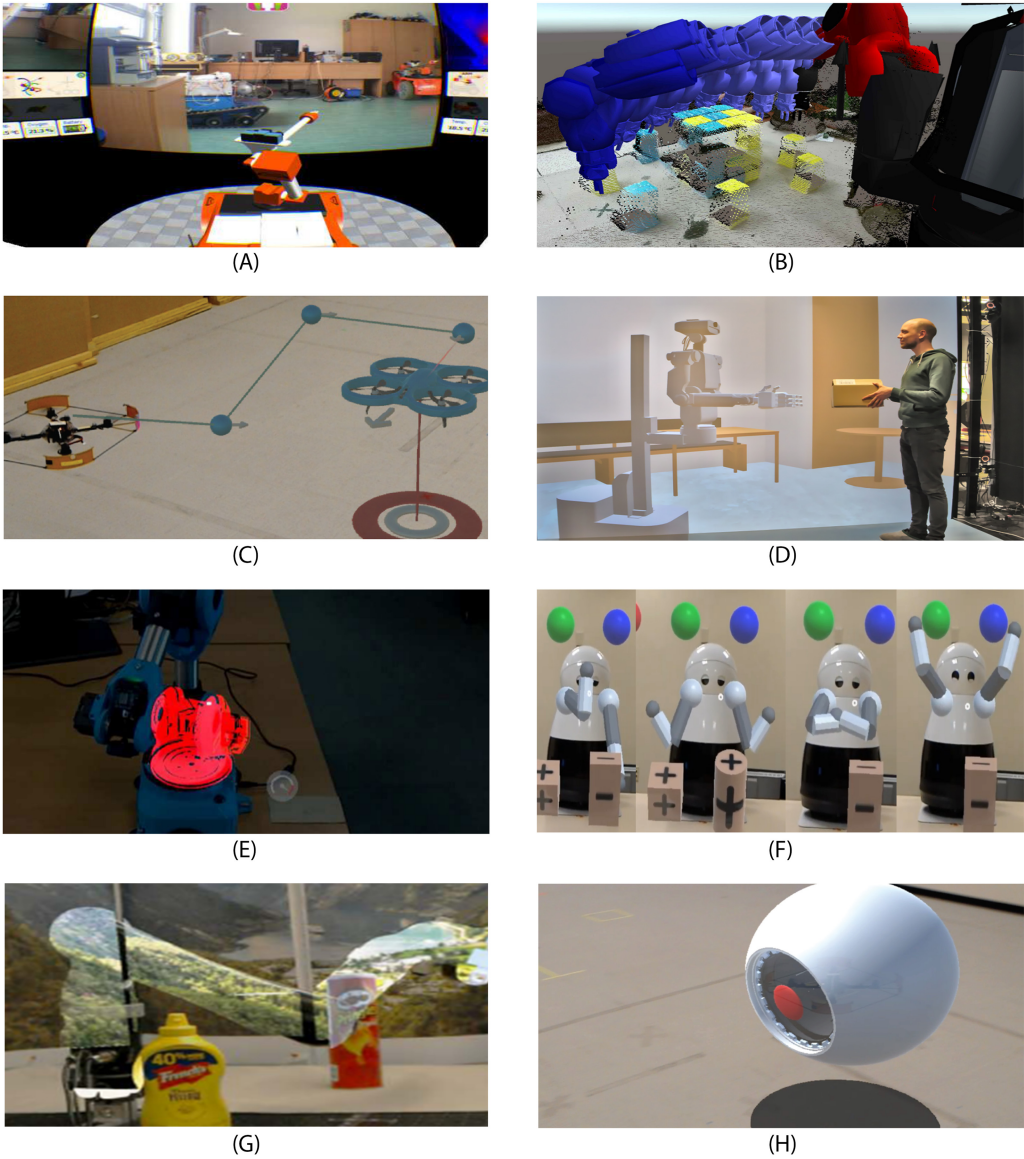


Fig. 3. A: Virtual control room, visualization robot, and internal readings [50], B: Cyber-physical interface, visualization robot, spatial previews, external sensor 3D Data, and Trajectories [76], C: Robot digital twin, headings, waypoints, and trajectories [96], D: Simulated robots, simulated agents, and simulated environment [58], E: Cosmetic alterations and internal readiness [6], F: Body extensions and callouts [35], G: Body diminishment [92], H: Form transformation and heading [95].

Virtual Entities – Environmental

The class of Environmental *Virtual Entities* encapsulates VAM-based visualizations of entities found in a user or robot's environment. For instance, virtual representations may be used to simulate agents, objects, and entire environments that do not physically exist in a robot's current development environment but will physically exist when a robot is deployed (e.g., to enable testing

in a laboratory environment that mimics conditions a robot would encounter in the field as robots are unable to distinguish between the simulated virtual objects and real objects). In addition, these VDEs can be used for simulating how a robot might interact with such objects. These simulated entities often enhance interactions between robot and developers when debugging/assessing robotic systems by freeing developers from real agents/terrain to test a system. Alternatively, Environmental *Virtual Entities* can be associated with physical agents, terrain, and objects already present in a robot's environment in the form of digital twins. We detail five VDEs in the Environmental sub-class of *Virtual Entities*:

Simulated Agents simulate, with virtual imagery, physical entities that normally have independent agency (e.g., autonomous robots, humans, animals). In the case of these simulations, the robot is not able to differentiate between real agents and Simulated Agents. The primary use of these VDEs is to enable robot testing and debugging without requiring the presence of key agents with whom robots would need to interact. For example, a large industrial robot might practice object handovers with a simulated human, without putting any human lives in harm's way, while the robot's developers observe and evaluate the mock interactions. Alternatively, a Simulated Agent VDE may be used in tandem with an autonomous Simulated Robot VDE for real humans to interact with, allowing for the testing of autonomous interactions with intelligent robots in situations where a physical robot is unavailable or currently infeasible [58] (see Figure 3(D)).

Simulated Objects use virtual imagery to simulate the presence of physical objects in a robot's environment. It is important to note that these nonexistent virtual objects hold no association with any real objects in a robot's setting. These nonexistent objects can be used to simulate obstacles in debugging sessions with real robots (e.g., a virtual wall, table, chair), without robot developers needing to procure physical objects, enabling rapid modification, deletion, and duplication of objects [11] (see Figure 5(G)).

Simulated Environments use virtual imagery to synthetically create the presence of environment areas or terrains. These VDEs can be used to evaluate autonomous robot responses to hazardous terrain (e.g., loose gravel, sand, water features) without endangering robots. Simulated Environments can also be used to evaluate robot interactions in environments that are difficult to find or recreate on Earth such as a lunar Moon with decreased gravity. Finally, Simulated Environments can provide a realistic setting to evaluate interactions (e.g., object hand-offs between user and robot) between Simulated Robots and users [58] (see Figure 3(D)).

Object Digital Twins act as virtual replicas of associated real objects to sequence actions to be taken on their real-world equivalents. These VDEs can allow users to preview actions to be taken on the real object prior to robot execution. For example, a real cup to be moved by a real robot might have a virtual cup overlaid on its current position. A user could then interact with the virtual cup and move it to a new location, fine-tune its final placement, and then command a robot to move the real cup to the position of the virtual cup. This interaction pattern can also enable robot action previewing similar to (and potentially in conjunction with) Robot Digital Twins [28, 52] (see Figure 5(H)).

Environment Digital Twins are virtual replicas of real environments rendered as a VAM-based visualization. These environments can be man-made structures/areas or outdoor terrain that are made to be exact replicas of their associated real world environment. As in Object Digital Twins, users could alter the state of the Environment Digital Twin to have a robot take action on the real environment so its state matches its digital twin. Examples of such systems include interfaces that visualize real satellite terrain data as an Environment Digital Twin VDE to test and/or

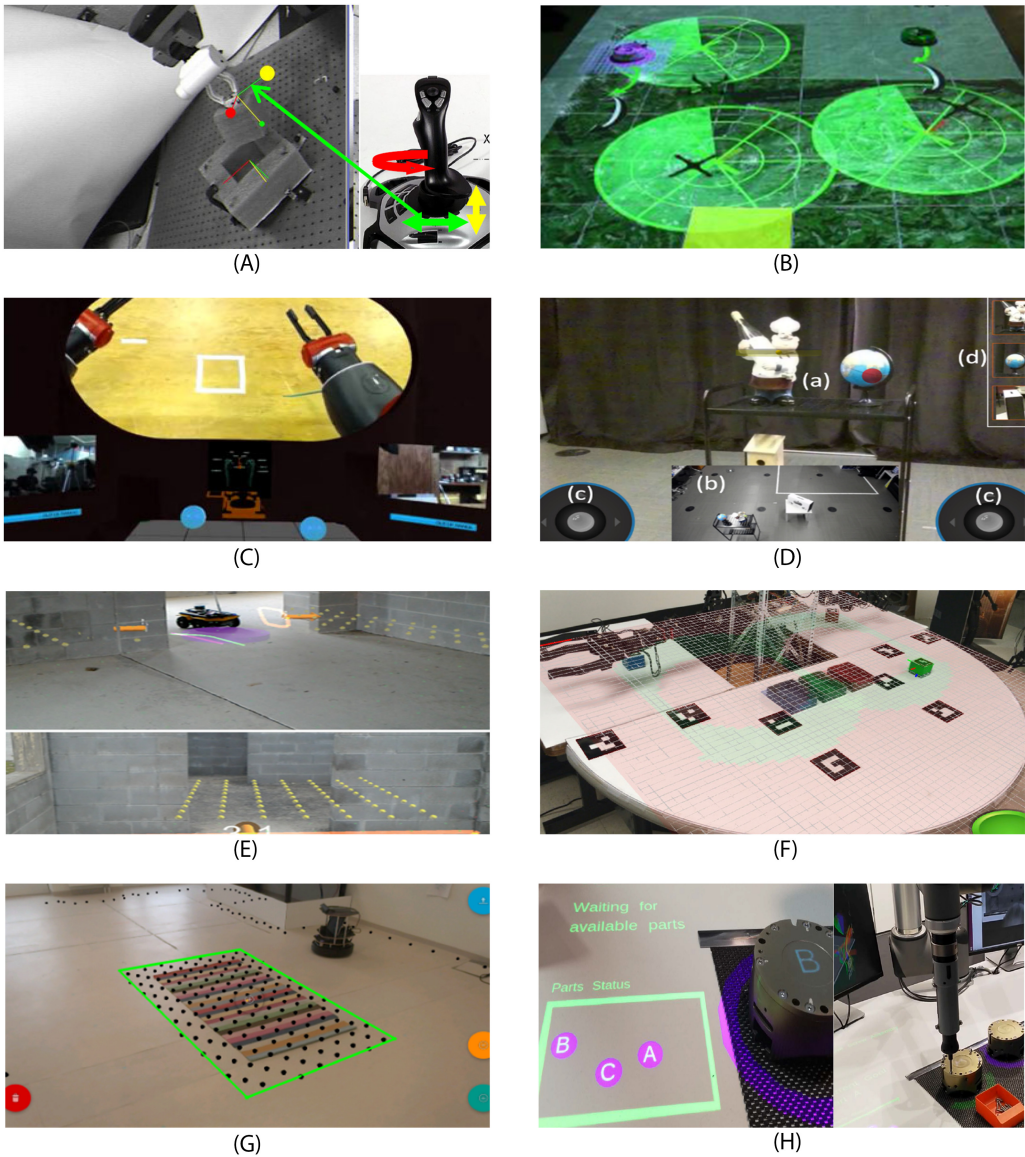


Fig. 4. A: Robot poses [65], B: External sensor purviews and environmental digital twin [68], C: Virtual control room and external sensor numerical readings [55], D: External sensor images & videos and virtual panels & buttons [54], E: Sensed spatial regions [74], F: Robot inherent spatial regions [28], G: User-defined spatial regions [83], H: Entity labels, entity locations, and task status [10].

supervise aerial robot systems scouting a wildland forest fires across the associated real expanse of wilderness [68] (see Figure 4(B)).

5.2 Virtual Alterations

The second category of VDEs we will examine are *Robot Virtual Alterations*: graphical elements that allow a robot's appearance to become a design variable that is fast, easy, and cheap to prototype

and manipulate. This category of VDE enables exciting new opportunities for HRI researchers and designers, especially since modifications to robot morphology are traditionally prohibitive due to cost, time, and/or constraints stemming from task or environment. We divide this category into classes involving (1) superficial alterations to robot appearance and (2) morphological alterations that substantially adjust robot form and/or perceived capabilities.

Virtual Alterations – Superficial

Superficial *Virtual Alterations* use virtual imagery to change the appearance of physical parts of the robot. This change in appearance does not occur by altering the robot's form or morphology (i.e., adding a virtual arm, making the head invisible, etc.) but instead by changing the appearance of robots' physical surfaces or the space adjacent to those surfaces. We identify two sub-classes of such elements:

Cosmetic Alterations alter the color, pattern, or texture of the robot's physical surfaces. The manipulation of robot surfaces enable new interaction patterns. These VDEs are considered cosmetic with respect to the robot's morphology and can be combined with additional VDEs to provide a function. For instance, changing the color of a robot arm in a manufacturing context might call attention to a malfunctioning actuator, indicate a hot surface temperature, or discourage touching. In an educational setting, superficial alterations might change the texture of a robot arm to look soft or furry to encourage interaction with children. Additionally, as robots increasingly expand to new consumer domains in the near future, designers could use Cosmetic Alterations to make robots more eye-catching in public spaces or enable end-user customization of personal robots in private living spaces to match home decor or personal taste [6] (see Figure 3(E)).

Special Effect Alterations add virtual imagery around robots' physical surfaces to change their appearance indirectly. These effects can take various forms such as a glow effect added around a robot's body, virtual streamers that render behind a robot's arm as it moves, virtual flames that spray out of a robot's end effectors, or virtual light sources that indirectly alter the reflective appearance of a robot's physical surface. Although we did not come across this VDE during our literature survey, VAM-HRI is still a growing field, and we envision this VDE holding value for manipulating human-robot interactions (i.e., adding virtual sparkles to a robot to make it more engaging to children in educational settings).

Virtual Alterations – Morphological

Morphological *Virtual Alterations* connect or overlay virtual imagery on a robot platform to fundamentally alter a robot's perceived form and/or function by creating new "virtually/physically embodied" cues, where cues that are traditionally generated using physical aspects of the robot are instead generated using indistinguishable virtual imagery. For example, rather than directly modifying a robot platform to include signaling lights as in Reference [91], an AR interface might overlay virtual signaling lights on the robot in an identical manner. Alternatively, virtual imagery might be used to give anthropomorphic or zoomorphic features to robots that do not have this physical capacity (e.g., adding a virtual body to a single manipulator or a virtual head to an aerial robot). Virtual imagery might also be used to obscure or make more salient various aspects of robot morphology based on user role (e.g., an override switch might be hidden for normal users but visible for a technician). These alterations may also enable new forms of interaction not previously possible for a given morphology, such as enabling functional robots to provide gestural cues [91]. We identify three sub-classes of such morphological *Virtual Alterations*:

Body Extensions add virtual parts to a robot without changing its underlying form such as an aerial robot is still recognizable as a UAV even if a virtual arm is added to its chassis, which would

not be the case if virtual imagery were overlaid on the aerial robot to make it look like a floating robotic eye [95] (see Figure 3(H)). Extensions do not necessarily need to be traditional robot parts and might instead appear as human heads, animal limbs, or even imagined parts like magical wings [14, 35] (see Figure 3(F)).

Body Diminishments visually remove, rather than add, portions of a robot (e.g., grippers, heads, arms, wheels) through **diminished reality (DR)** techniques [62]. An important use of this VDE is to resolve teleoperation occlusions that occur when a robot arm blocks the line-of-sight between its camera and the object being manipulated [92] (see Figure 3(G)).

Form Transformations overlay virtual imagery onto real robots to change the robot's underlying form and/or make it appear as something other than a robot entirely. Similar to Robot Diminishments, this VDE can change the form of the robot to make it more or less appealing to a targeted user group or utilize new communication methods, depending on the designer's intentions. These form alterations need not be limited to that of new mechanical forms, but can include any form such as that of a human, animal, or fictional character, all of varying degrees of realism [95, 105] (see Figures 3(H)).

5.3 Robot Status Visualizations

The third category of VDEs we present are *Robot Status Visualizations*: a set of elements focused on enabling designers to rapidly and easily assess the current state of a robot. We divide this category into classes focused on (1) internal and (2) external robot status.

Robot Status Visualizations – Internal

Internal *Robot Status Visualization* VDEs convey internal sensor readings and/or the operational status of robot sensors and actuators. We identify two sub-classes of such elements:

Internal Reading VDEs display data returned *from* internal sensors (e.g., battery levels, robot temperatures, wheel speeds). Making this information more readily available to users may enhance situational awareness and prevent mishaps such as robots running out of battery in the middle of mission-critical tasks or traveling too fast through environmental hazards [50] (see Figure 3(A) in which robot battery levels are displayed next to the stereo video stream).

Internal Readiness VDEs display data *about* sensors and actuators, such as whether a sensor is ready to collect data or whether an actuator is ready to function, and if not, why (e.g., whether a sensor is disconnected, an actuator is experiencing a fault). Such information may improve debugging and help prevent robots from operating as black-boxes, with users left wondering why a robot is not functioning as expected [6] (see Figure 3(E)).

Robot Status Visualizations – External

External *Robot Status Visualization* VDEs communicate the robot's external state (as the robot perceives it) by providing information regarding its current pose and location. We identify two sub-classes of such elements:

Robot Pose VDEs convey a robot's knowledge of its own pose (i.e., configuration and orientation). These VDEs can take different forms such as a textual display of numerical joint angles, a 3D model of a State Visualization Virtual Robot, or a rendering of a virtual axis anchored to a robot's joints [50, 65] (see Figure 4(A)).

Robot Location VDEs convey where a robot is in the environment, e.g., as an occluded robot's outline with a Spatial Visualization Virtual Robot or a virtual indicator on the other side of a

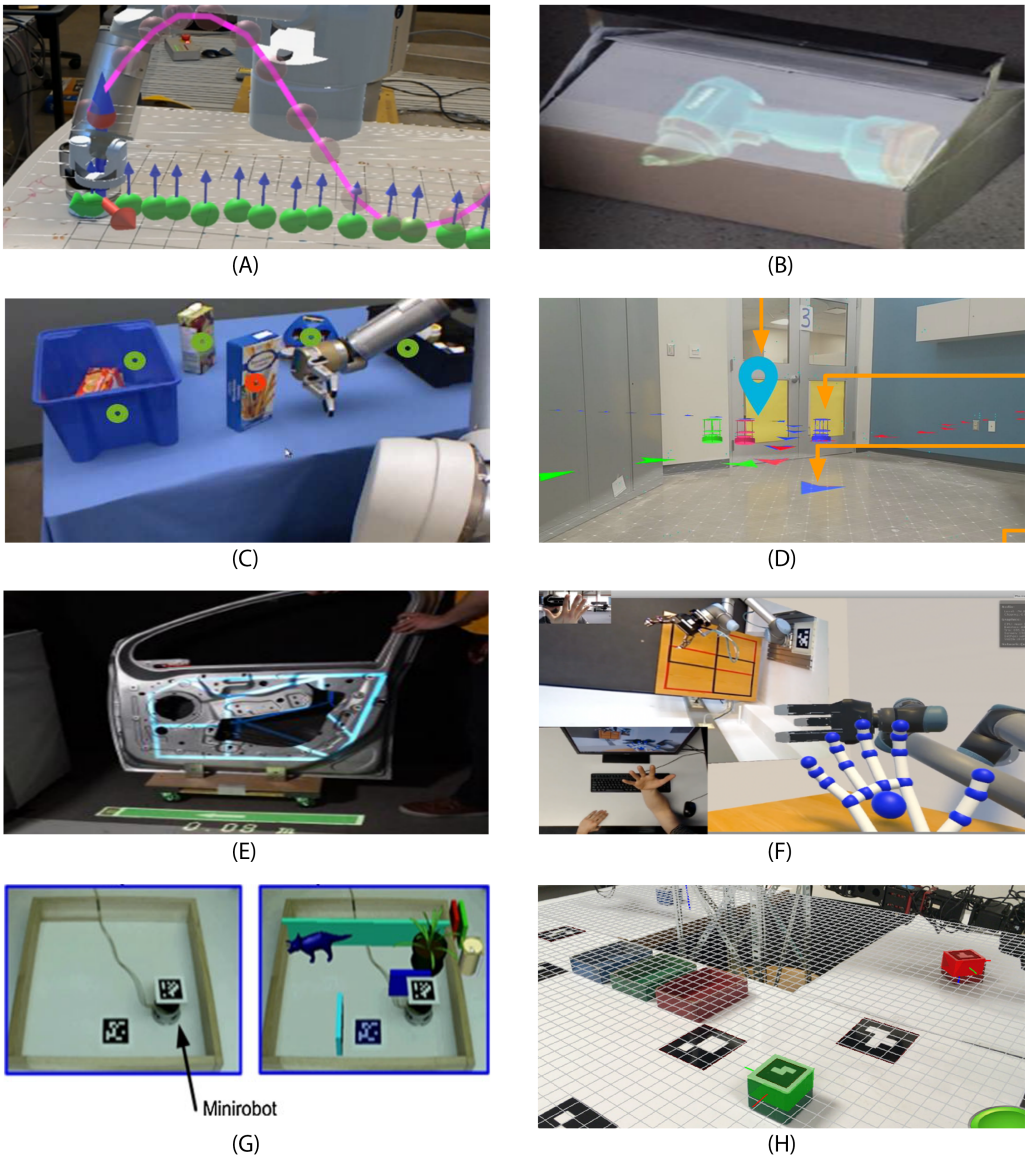


Fig. 5. A: Entity attributes and alteration previews [15], B: Entity appearances [31], C: Command options [72], D: Visualization robots, robot locations, and entity labels [16], E: Task instructions [31], F: Virtual controllers [51], G: Simulated objects [11], H: spatial previews and object digital twins [28].

wall or door, as a top-down radar-like display showing user and robot locations, or as an off-screen indicator to direct a user's attention to a robot outside the current field-of-view [16, 95] (see Figure 5(D)).

5.4 Robot Comprehension Visualizations

The fourth and final category of VDEs we examine are *Robot Comprehension Visualizations*: visualizations that convey what a robot believes about its environment, and its current or planned

task. VAM-HRI VDEs present a powerful medium for conveying this information, as VDEs can be directly overlaid on a robot's environment. For example, a visual trajectory spline rendered on a floor can wrap around a wall, indicating not only that the robot sees the wall, but also that the robot's planned actions will avoid the wall.

Robot Comprehension Visualizations – Environment

Environment-based *Robot Comprehension Visualization* VDEs communicate to the user what the robot believes about its environment, including where environment information is being collected from, what environment information has been collected, and/or what a robot has inferred from such information. We identify seven sub-classes of such elements:

External Sensor Purviews are environment-anchored visualizations that show users where a robot's external sensors (LiDAR, cameras, etc.) are collecting data and/or how and where those sensors are positioned [37, 49] (see Figure 4(B)).

External Sensor Numerical Readings convey numerical data returned from a robot's external sensors. These readings can be user- or environment-anchored and can be shown explicitly with digits or as more abstract visualizations such as progress bars, virtual thermometers, or virtual weight scales [14, 55] (see Figure 4(C) in which end effector range data is displayed as a blue progress bar below the end effector video streams). Note that the external sensors need not physically attached to a robot, such as data from motion capture cameras regarding the distance between a robot and a nearby object.

External Sensor Images & Videos allow the user to see remote environments or to see from a robot's perspective. Images and videos can be presented as either user- or environment-anchored visualizations from cameras on the robot or in the robot's environment. When stereo cameras are paired with a stereo interface (e.g., an HMD, 3D monitor, or CAVE), users can see the images and videos with stereoscopic depth, granting enhanced immersion and teleoperation capability [50, 54, 55] (see Figure 4(D)).

External Sensor 3D Data convey depth information through the recreation of remote robots' environments. These reconstructions can take various forms (e.g., point clouds, voxel maps, 3D meshes) and aim to present sensed depth data in a manner that allows users to perceive remote environments as if they were there in-person. Unlike Images and Video VDEs, the 3D reconstructions utilized by External Sensor 3D Data VDEs enable immersive and free exploration of a remote robot's environment, without being restricted to the robot's perspective [76, 87] (see Figure 3(B)). As in the case of the previous two VDEs, the external sensors do not need to be physically attached to the robot, and this VDE includes 3D data collected from user-worn HMDs, such as the spatial map generated from a HoloLens.

Sensed Spatial Regions, the first of three region-based VDEs, visualize regions that a robot has identified within its environment. These visualized regions are produced by a robot through the analysis of sensed environment data, such as exploration frontiers or traversable vs. non-traversable areas depicted through an occupancy grid (unlike 3D reconstructions that do not perform any logical analysis and categorization of environmental regions) [74] (see Figure 4(E) in which regions are annotated with information regarding the estimated information gained if the unexplored area were to be explored).

Robot Inherent Spatial Regions are not informed by data sensed from the environment, but are instead inherent to the robot based on its form or operational mode. For example, these may depict the regions a robot can physically reach or show areas a robot may operate best in, such as

the optimal area in which to perform an object handover [28] (see Figure 4(F) in which the green and red areas denote where the robot’s arm can or cannot reach).

User-defined Spatial Regions are regions defined by user input, rather than by the robot’s sensors or physical constraints (e.g., a user drawing a bounding box on the floor of a robot’s environment with a tablet display). These regions have many potential uses but are most commonly used to define areas a robot should not enter, in the form of virtual boundaries [83] (see Figure 4(G)).

Robot Comprehension Visualizations – Entity

Entity-based *Robot Comprehension Visualization* VDEs convey what a robot knows or believes about an entity (i.e., an object, human, another robot, etc.), such as where an entity is, what an entity is, and attributes an entity holds. We identify four sub-classes of such elements:

Entity Labels act as identifiers for entities known by a robotic system. These visualizations enable users to easily reference entities in spoken commands without the need for referring expressions such as enabling users to instruct robots through commands such as “pick up cube B” [10, 81] (see Figure 4(H)). Additionally, these VDEs allow robots to label points of importance in the environment such as a room’s primary access point [16] (see Figure 5(D)).

Entity Attributes convey information a robot knows about an entity’s characteristics, such as whether an entity is heavy, delicate, or dangerous; information known about the entity’s affordances; the current state of an entity (e.g., an entity that is too hot, in a dangerous location, still drying, sleeping, charging its battery); or an entity’s geometry and shape (e.g., optimal grasp points or surface normals) [15] (see Figure 5(A)).

Entity Locations highlight the locations of entities within the robot’s environment through rings, arrows, bounding boxes, and so on. This VDE is especially useful when the location of an entity is occluded by walls or containers or outside of the user’s field-of-view [23]. These can also be used to highlight task- and dialogue-relevant entities, either by allowing robots to passively highlight entities that are of interest to the current task or the subject of the robot’s current attention, or to actively call interlocutors’ attention to entities in the same way that humans typically would through deictic gaze and deictic gesture [10, 72, 81, 99] (see Figures 4(H)).

Entity Appearances show what an entity looks like, unlike Entity Location VDEs that communicate the location of an entity. These VDEs are primarily used when an entity is occluded from view and draws analogies to “X-ray vision” by showing users the appearance of real-life entities that are partially or fully visually hidden within the users’ environment (e.g., inside a box, behind a robot arm, or on the other side of a wall or door). This interface feature may be particularly useful in environments, such as warehouses, where objects are stored in sealed containers with contents only knowable through data representations that are exclusively computer-readable (e.g., barcodes) [31] (see Figure 5(B)).

Robot Comprehension Visualizations – Task

Task-based *Robot Comprehension Visualization* VDEs display what a robot understands about its current or planned task, including where to move, how to move, what objects to act upon, and how to act on those objects. These VDEs can also convey information regarding general task understanding, such as task status and outcomes. We identify nine sub-classes of such elements:

Headings, the first type of VDE in this class, do not show the actual path the robot or its manipulators will take, but simply the direction they are currently traveling in or will be traveling next. These visualizations commonly take the form of arrows pointing in the direction of planned

movement; however, they can also take more unique forms (i.e., the utilization of a Form Transformation VDE to provide an eyeless robot with virtual eyes that designers can use to provide gaze cues that communicate future movement intentions [95] (see Figure 3(H)). Headings may be useful for autonomous robots in crowded, shared spaces with human pedestrians or dynamic obstacles, in which navigational plans need to be recalculated by the robot quickly and frequently. Researchers have also shown how headings can be particularly effective when displayed using projectors, enabling all bystanders to see the intended movements of robots without observers needing to each wear or use specialized hardware [80] (see Figure 3(C)).

Waypoints are environmentally anchored visualizations of intermediate navigation points. These are typically used to visualize robot intentions but can also be used by robots to suggest spatial destinations indicating where users should move. Waypoints provide another method of previewing robot motion and can either be automatically placed in an environment by a robot trajectory planner or manually placed in the environment by a user. Waypoints are also often combined with other VDEs to show additional information known about each waypoint or what will be performed at each waypoint [15, 95, 96] (see Figure 3(C)).

Callouts are visualizations that communicate where a user should focus their attention. These VDEs use visualizations to attract attention to an object or location, such using a virtual arrow to show where a robot heard a sound or pointing at an object a user should look at [35] (see Figure 3(F)).

Spatial Previews use environment-anchored visualizations to show future poses of robots, objects, and other environmental entities. These VDEs can explicitly communicate the expected future position and/or orientation of an entity during a task. A common use for these previews is to depict where a robot will move or where a robot will move an object during a manipulation task. However, robots can also use these VDEs to make requests to users by indicating where a user should place an object. These VDEs can be depicted in various ways, including 2D circles on the ground, complex 3D wireframe or shaded models, combining waypoints with flags indicating orientation at those waypoints, or with one or more Spatial Visualization Virtual Robots [28, 76] (see Figures 3(B) & 5(H)).

Trajectories display spatial paths that a robot intends to follow or that it believes an object or agent will follow. These environment-anchored visualizations can show both robot navigation paths and manipulator paths. Trajectories can be visualized in various ways such as lines, splines, or dense stroboscopic Future Robot Pose VDEs in the form of State Visualization Virtual Robots. For example, one common implementation of Trajectory VDEs consists of rendering a trajectory for each wheel on a ground robot, which helps users to anticipate whether or not a wheel will collide with or fall into a terrain hazard. Trajectories can be used to not only indicate the path a robot will take in the future, but also the path a robot (or human) has taken in the past. Trajectory VDEs are typically only used to enhance a user's view into a robot's internal model but can also be used to provide new opportunities for control over the robot (e.g., a user directly manipulating the trajectory visualization by grabbing the trajectory line or spline and moving it with their hands). Finally, trajectories can encode data along their paths, such as robot velocities [53, 76, 95, 96] (see Figure 3(C)).

Alteration Previews show the intended permanent modifications a robot plans to make on an object. These rendered previews give the user a chance to verify if the robot's plan matches that of the user's task goal(s) prior to execution and cancel or modify them if needed. These modifications typically show how an object will appear during or after the modifications take place (e.g., displaying the proposed path of an etching tool on an object's surface, where holes will be drilled, how a

wall will look after being painted, how a steel bar will look after being bent). Additionally, these visualizations may communicate what actions will be applied to an object, such as varying pressures applied along the surface on object. This type of VDE is especially useful in circumstances where robot errors arising from command misinterpretation mean the object being acted upon will no longer be usable, potentially wasting hours or days of time and resources to replace the incorrectly modified object [15, 53] (see Figure 5(A)).

Command Options present to users what actions a robot can (or cannot) take in a given state. This may take the form of a robot displaying virtual imagery that indicates what object(s) it can currently pick and/or potential grasp points the robot can utilize [72]. In addition to showing user what actions a robot can take, these VDEs allow a robot to inform users that it is incapable of performing an action it believes the user wishes or might wish it to perform, sometimes with an explanation as to why they are not possible. These VDEs may reduce user frustrations with robotic systems by avoiding situations where a robot silently fails to execute commands and improve user efficiency by aiding users in preemptively realizing that a task will not be performed correctly (or at all) [3] (see Figure 5(C)).

Task Status convey beliefs regarding the status of a task currently or previously executed. These VDEs may be represented as traditional textual or numerical visualizations, or as more abstract visual representations, such as progress bars. These visualizations can facilitate human-robot collaborative task planning by helping users quickly and easily understand the current task state a robot is executing, improve debugging by enabling users to compare the state a robot thinks it is in with its actual state, or how much longer a task will take to complete [10, 31, 95] (see Figure 4(H)). Additionally, these VDEs can convey the status of a concluded task, whether it has resulted in success, failure, or error. These visualizations help users understand what a robot believes the outcome of a task to be (even if incorrect, which may aid in robot debugging). These visualizations can inform a user as to why a task resulted in failure or error, which can often be a mystery to users who would otherwise need to consult complex error logs [21, 31].

Task Instructions enable humans and robots to effectively instruct and guide each other by communicating next steps in collaborative tasks. These instructions can take various forms, such as explicit instructions written in text or more abstract instructions that inform a user what to do to accomplish a task such as using a Robot Extension Morphological Alteration VDE to add virtual arms to a robot that point at an object for a user to interact with [31, 35] (see Figures 5(E)).

6 ROBOT FUNCTIONALITIES SUPPORTED BY VAM-HRI

VAM-HRI interface VDEs have been leveraged to enhance a number of fundamental robotic needs (e.g., navigation, manipulation, training) that are applied across robot domains (healthcare, education, entertainment, etc.). To examine how VDEs have been implemented to augment or enhance these generic robot capabilities, we divide these robotic functionalities into the following categories: navigation, object manipulation, prototyping, human training, robot training, debugging, swarm supervision, and social interaction. In this section, we will describe the way in which VDEs have been used for each of these purposes.

6.1 Robot Navigation

For robotic applications that do not involve stationary robots, indoor and outdoor environments of varying scale must be safely and efficiently navigated. Our survey of the literature identified 51 papers in which AR and VR were used for this purpose. Kästner et al. [46], for example, present an approach in which *trajectories* are used to visualize a robot's intended navigation path within human teammates' HMDs. Moreover, Kästner et al. allow users to re-specify the robot's destination

through manipulation of a virtual arrow used to denote its destination. Similarly, Stotko et al. [86] present a VR-based system for remote robotic teleoperation, in which robotic sensor data is visualized directly within the user's HMD to allow them to explore the robot's environment through that robot's perspective with high levels of immersion.

6.2 Robot Manipulation

While not necessary in some social domains, most task-oriented robotics applications require robots to physically manipulate objects in their environment, e.g., in assembly or sorting tasks. Our survey of the literature identified 105 papers in which AR or VR were used for this purpose. Many of these papers used AR to help human teleoperate robotic manipulation more accurately and safely. Krupke et al. [52], for example, present a system that superimposes a virtual Robot Digital Twin over a real robot in augmented reality. Users of this system can then control the real robot to perform a pick-and-place task through manipulation of the Robot Digital Twin and Object Digital Twins shown in their HMD. After each user command, the virtual robot simulates performance of the commanded task. If the user is satisfied, then they can then trigger the real robot to perform that command. While in this use case, the robot operator is present alongside the robot, VAM techniques can also be used to help users to remotely control robot manipulation. Naceri et al. [64], for example, present a VR interface for real-time robot teleoperation. In this interface, the remote environment is visualized using streaming External Sensor Images and Videos and External Sensor 3D Data, to make robot teleoperation more effective.

6.3 Robot Training

In many circumstances, robot end-users are tasked with teaching and/or programming a more general-purpose robot to perform a specific job, such as how to sort dishes, open and pour a bottle, or build furniture. More traditional robot training methods, such as learning from demonstration [4], can be especially challenging tasks when minimal feedback from the robot is provided to the user regarding how well the robot is learning, why the robot is not learning, or how to provide teaching inputs to the robot. VR and AR can be used in these contexts to visualize task-relevant objects and obstacles as virtual objects. Our survey of the literature identified 31 papers that used VR or AR for such purposes. Sprute et al. [84], for example, present an AR system for teaching robots the extent of its operating environment through User-Defined Spatial Regions that serve as virtual walls. The virtual borders are defined with an AR tablet interface, which the robot uses to subsequently generate its navigation plans.

6.4 Human Training

Alternatively, in some domains, humans need to be trained to operate robots within a safe and constrained environment. Additionally, robots may need to share task-relevant information with their human teammates, such as instructions on how to complete a human-robot shared task. Our survey of the literature identified 35 papers in which AR and VR were used for this purpose. In these cases, VR and AR are often used to develop personalized, low-cost training environments. Many of these use cases are in the context of training users to use vehicles (autonomous or otherwise) [5, 75] or in the context of training humans to train robots [30]. Driving simulations provide learners a safer place to improve their skill without worrying about causing disturbances to others.

6.5 Robot Debugging

Similarly, even outside of training procedures, robot programmers as well as end-users often need to determine on the fly why a robot is acting in a particular way, especially when unexpected behavior is displayed. Current robots typically require these users to parse detailed error logs

to answer fairly simple questions, such as why a robot's end effector stopped moving. This sort of question can have multiple answers, ranging from motor faults to payloads that exceed maximum weights. Our survey of the literature identified five papers in which AR and VR were used to enhance robot debugging. VAM techniques can be added to robotic systems to aid in quickly answering common debugging questions. For example, robot debugging can be enhanced by rendering virtual imagery to localize and efficiently explain robot faults. VAM-HRI systems have been designed that visualize robotic faults within user HMDs through the use of Virtual Cosmetic Alterations on the robot that highlight the robot parts that are currently experiencing faults with paired visualizations that provide at-a-glance information about fault type (e.g., sensor fault, servo fault, end effector overloading) [6, 21].

It is important to note that HRI is not restricted to interactions between robots and end-users, but between robots and developers as well. When going through the process of creating a new robot or robot algorithm, robot designers go through extensive cycles of debugging. To iterate and improve a robotic system, these users must understand why a robot/algorithm is not performing as expected. During this testing phase VAM-HRI techniques can allow robot designers to more easily see what a robot is thinking through virtual imagery, such as if a robot detects an obstacle while testing an autonomous navigation algorithms [49].

6.6 Robot Prototyping

Similarly, while many of the previous use cases have focused on on-line robot tasking, VAM can also be used in the initial design of robots before they are ready for deployment. When VAM is used for robot prototyping, virtual imagery is used to preview robot designs and/or functionality. This virtual imagery can be used either to represent a completely virtual robot or add virtual parts to a robot—all without using physical robotic hardware. VAM robot prototyping saves on both the monetary costs of robotic hardware as well as hours of labor that would otherwise be needed to install or program robot parts during the design process. Our survey of the literature identified 10 papers in which AR or VR were used for this purpose. Cao et al. [14], for example, introduce a mixed reality robot prototyping system for people building DIY robots that allows users to virtually assemble and construct robots in AR with Visualization Robots and Body Extensions. Using this system, hobbyists can test their designs in mixed reality before executing those designs in real life with Simulated Robots and Simulated Objects.

6.7 Social Interactions

In the domain of social robotics, robot developers use a variety of design strategies to manipulate users' perceptions of robots as being more trustworthy, engaging, and/or approachable. Our survey of the literature identified 17 papers in which AR or VR were used for this purpose. Many of these approaches have operated by altering robot appearance or by enhancing robots' communicative capabilities to allow communication that would otherwise have been impossible given their inherent morphologies. For example, Zhang et al. [105] present a system to enhance human perceptions of interaction proxemics with a mixed reality robotic avatar. In this case, the physical robot is non-humanoid, but a Form Transformation VDE is utilized by overlaying a 3D AR avatar of a human above the real robot that mimics a human's gaze and body motions while moving. Through arm swinging frequency, this visualization allows the robot to effectively communicate its moving speed to nearby humans and improve subjective perceptions about the robot.

6.8 Swarm Supervision

Finally, while many of the approaches above have focused on single robots, management of multi-robot systems is also a major challenge for robot designers and users. As more robots are added to a

robotic system, the system becomes increasingly difficult to supervise and/or control. The number of robots that can be operated simultaneously is called the fan-out of a human-robot team, with robots that have high neglect tolerance and lower interaction time achieving higher fan-out [67]. Our survey of the literature identified three papers in which AR or VR were used for this purpose. In these papers, VAM-HRI researchers have investigated how VAM technologies can increase the fan-out of robotic systems and decrease the mental load of robot operators, for example, by rendering virtual imagery to display the location and status of many robots. For example, Ghiringhelli et al. [33] present an AR interface for swarm supervision, in which the supervisor is presented with Robot Status Visualizations VDEs (Robot Location, Robot Pose, Waypoints, Headings, and Trajectories) overlaid over each robot.

7 VAM-HRI FOR ROBOT DOMAINS

While in the previous section we explored the different core robotic functionalities that VAM-HRI technologies are being used to augment, in this section, we turn our attention to the high-level application domains in which such solutions are being designed and deployed. Our coverage of these domains is guided and organized according to the set of application domains delineated by Bartneck et al. in their recent textbook [8]: customer service robots, robots for learning, robots for entertainment, robots for healthcare and therapy, service robots, collaborative robots, self-driving cars, and remotely operated robots (See Figure 6).

7.1 Collaborative Robots

Collaborative robots, unlike traditional industrial robots, have safety features and human-friendly designs that allow human to work closely with them in manufacturing, shop floor, or maintenance contexts. Mixed and virtual reality can provide otherwise unobservable information about the robots working in these domains, such as visualization of the areas in which such robots work, the areas reachable by such robots, and the regions such robots are likely to move in the future. Making all of this information available to human teammates can make collaborative robots safer and more effective for those teammates to work with. Our survey identified 34 papers in which VR and AR are being used in these domains.

Matsas et al. [56], for example, prototype techniques in virtual reality to improve safety in human-robot collaborative manufacturing. Specifically, in this system, a red-bordered circle is overlaid on human teammates' view of their environment through a virtual reality interface to show the limits of the robot's workspace, and a yellow wedge is drawn to show the movement range of the robot's arms. If humans enter these spaces, then this wedge turns red, and a warning is displayed.

7.2 Customer Service Robots

Customer service robots have been used as tour guides (e.g., providing information to visitors about points of interest and taking visitors to requested location), receptionists (e.g., providing check-in processing for hotel guests), and sales promoters (e.g., providing store promotion information to customers). VAM technologies have not yet been applied to this domain widely: Our survey identified a single paper within this domain. Specifically, Pereira et al. [69] demonstrated the use of an AR dialog interface that provides a compelling visual interface for customers to interact with robots in restaurant contexts.

7.3 Robots for Learning

Robots for learning are those operating in classroom environments, as teachers, tutors, peers, or teaching assistants—or as the objects of study themselves—to help make lessons more effective and

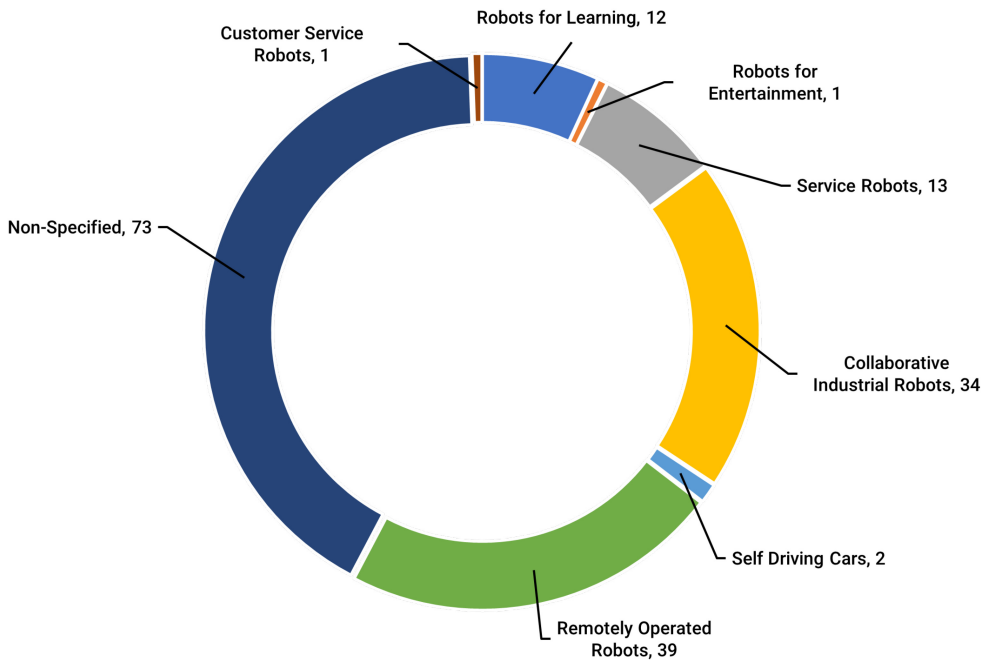


Fig. 6. The above graph shows the distribution of VAM-HRI papers surveyed that either aimed at enhancing robotics in a general sense (i.e., non-specified) or targeted specific robotic domains, such as education.

engaging. With augmented reality, visual educational material can be deployed in tandem with a physical robots to further enhance the student learning experience [18]. New knowledge can be overlaid on physical robots so students can see this information in a spatially situated manner. Alternatively, learning content can be projected onto a surface on which students and peer robots can engage with that content together.

Our survey of prior literature identified 12 papers in this area. Johal et al. [45], for example, present an educational robot that is used to teach optical concepts regarding the visible and **infrared (IR)** light spectrum to K–12 students during Physics classes. Johal et al. use AR in this context to visualize information about the robot’s IR sensors, such as the direction of IR emitters, cone range, and intensity of the IR signal, all of which are visualized to students through Android tablets.

7.4 Robots for Entertainment

Robots have also been used for entertainment, as pets, toys, exhibitions, or in the performing arts. VAM has not been applied widely in entertainment robotics: We identified only one paper in this domain. The robot in this paper [94] is not a traditional entertainment robot of the sort mentioned above but is instead a multipurpose wearable (wrist-worn) robot that has many different functionalities invoked by AR system. It consists of six interlinked servomotors fastened together using plastic brackets. This robot follows users’ commands to change speeds, turn to specific angles, or set torque limits, and can wrap around or stand straight up from a user’s wrist. Through an AR headset, a window appears on top of the robot, from which the users can see the robot status display, shape-changing menus, a media player on which they can watch videos, and a robot pose controller.

7.5 Robots for Healthcare and Therapy

Robots are widely applied to healthcare and therapy. They offer support for senior citizens by helping to detect adverse medical events or by providing enhanced mobility. They are also used in the context of therapy, e.g., for people with autism spectrum disorder, people undergoing rehabilitation, and people undergoing surgery (e.g., by providing assistance in laparoscopic surgery). To support robotic surgery, AR helps to highlight anatomical structures, overlay surgical plans, and display robot and instrument status on the doctor's main visual source (e.g., patient body, video stream, etc.). AR can also be used to visualize the planned trajectory of the surgical robot's needle so the surgeon can check if it is valid.

Our survey identified 71 papers in which AR and VR are used with robots in such domains. AR has been used in this domain primarily to support **robotic-assisted surgery (RAS)**. Qian et al. [71], for example, present an AR system called ARssist to aid the first assistant in robotic-assisted laparoscopic surgeries. This system visualizes the robotic instruments and endoscope inside the patient body using the assistant's HMD. This approach has the potential to improve efficiency, navigation consistency, and safety for instrument insertion. VAM environments have also been used in the context of therapeutic and assistive robots such as robotic wheelchairs. Zolotas et al. [107], for example, present an augmented reality system that can help users control their wheelchairs safely and independently. Users of this system observe a mini-map utility that shows the wheelchair's future trajectory, potential obstacles, and potentials for collision.

7.6 Service Robots

Service robots (as opposed to the previously discussed customer service robots) perform simple and repetitive tasks in service of humans, such as house-cleaning, delivery, or security operations, as well as other dull or dangerous tasks such as space exploration and emergency response. Many service robots work remotely by themselves. In these cases, the remote environment is displayed to the operator using 3D rendering from point cloud data or video streams. If humans and service robots are co-located, such as in certain types of search and rescue tasks, then robot state information such as location, battery, condition, trajectory, and so forth, can be shown within their teammates' HMDs. Our survey of the literature found 13 papers in this domain.

Our survey identified 15 papers in which AR and VR are used with robots in such domains. Martín et al., for example, [77] present a multimodal AR system that allows robots to warn human teammates more effectively about hazards during navigation through unfamiliar spaces, by way of hazard area visualizations in teammates' HoloLens interfaces. Service robots can also be helpful in facilitating repetitive tasks in field domains such as agriculture. Huuskonen et al. [41], for example, provide an AR interface that allows farmers to simultaneously monitor multiple autonomous tractors.

7.7 Self-driving Cars

Self-driving cars (and other autonomous vehicles) are robots that can automatically navigate between locations in large-scale human environments, especially those that can operate on semi-structured human transportation infrastructure such as roads and highways. One of the main proposed benefits of self-driving cars is to allow autonomous driving when humans are too fatigued to safely operate their vehicles. However, current self-driving cars have a number of limitations that require humans to be able to quickly take over and intervene. Since on-road autonomous car training is dangerous and expensive, VAM technology has been applied to design autonomous car training programs. During VR training, the simulators are shown in HMDs to show users information about the virtual car such as speed limit, distance traveled, and current speed. In AR training programs, the user is trained in a real car on a designated road while

wearing a see-through AR HMD that can display introductory videos, car's state information, and instruction to the user. Our survey revealed two papers in this domain. For example, Sportillo et al. [82] present a virtual reality training program for autonomous vehicle operators that can help train such operators to improve their ability to quickly retake vehicle control when necessary.

7.8 Remotely Operated Robots

Remotely operated robots are robots that are controlled by humans from different places (a use case that overlaps with some of the other application domains described above). In this scenario, human operators usually receive the visualization of the remote environment in point cloud data rendering or video stream. Via VAM interface, operators can see robot status, future trajectory, heading, grasping point, which are overlaid on the environment visualization.

Our survey identified 39 papers in this domain. For example, Zollmann et al. [106] present an AR interface for piloting **unmanned aerial vehicles (UAVs)**. In this paper, information about the UAV is displayed in the user's AR HMD: A virtual sphere acts as waypoint for the UAV, a virtual shadow is displayed on the ground, and a line is drawn between the waypoint and the shadow to show how high the UAV is from the ground. Many virtual spheres are connected to be UAV's trajectory, which help the user to supervise the UAV's path and intervene if they detect any potential collision. However, Kent et al. [47] design an AR interface to help users teleoperate robots in an object manipulation task. In the AR interface, when the remote operator clicks on an object, a semi-transparent sphere appears overlaying the selected object with several blue grasping points. After the operator chooses one of the grasp points, a 3D model of the robot's hand/end-effector is presented to the user to preview the grasping angle. After confirming the grasping pose, the real robot arm executes the grasp.

Finally, Gharaybeh et al. [19] present an MR system for teleoperating robot arms to defuse hazardous undetonated underwater munitions; an otherwise very dangerous task for human divers. After submerging a robot arm, its teleoperator can use visualizations of the arm's LiDAR point cloud sensor data to see the ocean floor, the undetonated munition, and other helpful visual aids, including a 3D model of the robot arm. This 3D information brings depth perception to the operators, enhancing control of the remote robot arm to defuse munitions more safely and easily.

8 TAXONOMY VISUALIZATION AND COMMUNITY ENGAGEMENT

As a final contribution of this work, we produce an online version of this taxonomy accessible at: vamhri.com. This online resource presents the categorization of all 175 papers, through various forms of visualization to enable the researcher to explore the space of research in this field through the taxonomic classification proposed in this article. The interface has been built using a survey visualizer tool [108] that is open-sourced at: <https://github.com/TathagataChakraborti/survey-visualizer>.

The primary view is presented in Figure 7. This allows the user to interact with the categorization by searching with keywords, and filtering with selections over timelines, categories, and so on. By selecting any specific categories, the user can also view the timeline of research in that category (Figure 7, left inset). Additionally, the tool presents two alternative views into the survey data, as follows:

- The affinity view allows the user to explore the papers contained in the survey in the space of document similarity [20]—here, a document is a collection of the paper metadata, including authors, venue, keywords, title, abstract, and so on, and the taxonomic classes describing the paper. An illustration is presented in Figure 8(a), where we have selected a cluster of papers on the right. The collection of tags underneath the selection describes the selection

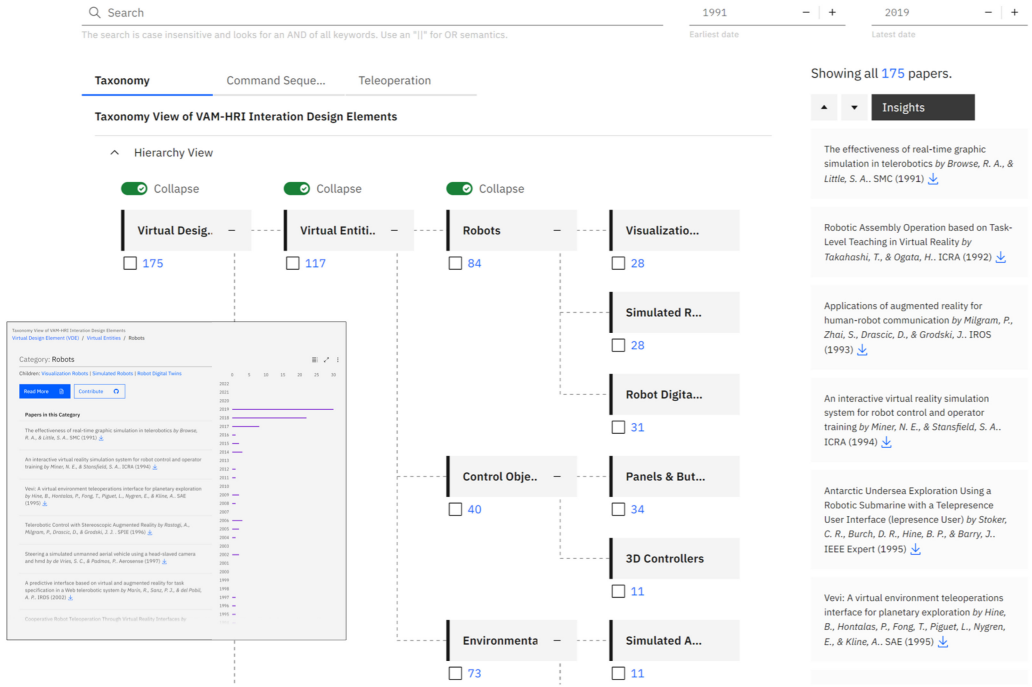
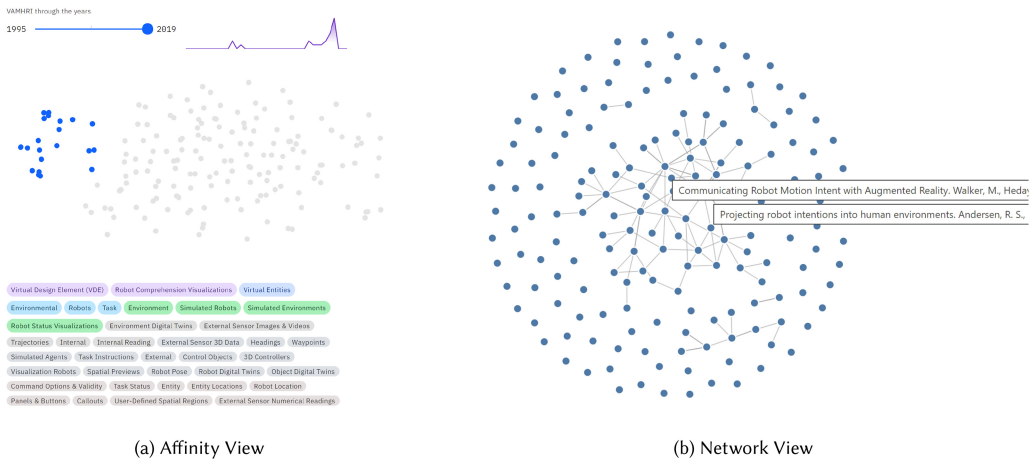


Fig. 7. Interactive visualization of the VAM-HRI taxonomy at vamhri.com.



(a) Affinity View

(b) Network View

Fig. 8. Affinity view and network view of the VAM-HRI taxonomy at vamhri.com.

in terms of its categories (the selection in the example seems to be dealing with virtual and simulated design elements). The sparkline on top aims to give a sense of how active the research described by the selection of sub-categories is.

- The network view, however, presents the influence network in the field by mining for how papers covered in the survey point to each other. For example, one of the hubs in the network (as it appears on hovering over a node), presented in Figure 8(b), is Reference [2], a highly cited work in the field. While one of the hubs identified in the network does point back to

a heavily cited work [95] from the authors on this survey, as is expected from any selection and categorization scheme, the large number of orphaned nodes in the network also indicate the diversity of topics covered here as well as the scope for further contributions from the VAM-HRI community.

The VAM-HRI community began its journey half a decade ago in 2018 as a workshop [102] at the ACM/IEEE International Conference on Human Robot Interaction. It has grown in size and scope significantly since then, and the diversity of research covered in this article has hopefully captured that exciting journey. The online interactive platform enabled by this work accompanies the workshop today, as a tool for new students to explore and become quickly initiated with the field, as well as for researchers, new and experienced, to identify gaps in existing literature and push the community forward. At the 2022 edition of the workshop, the tool was used to categorize the program and ground the discussion during the breakout sessions.¹ Through curated issue templates on GitHub,² any VAM-HRI researcher is invited to create a pull request to add their research to this categorization scheme and grow the scope of work over time, both in terms of updates and enhancements to the taxonomy as well as new entries into the list of papers covered by it. It is our hope that this platform serves as a living resource that the VAM-HRI community may use to track the continued progress and growth of our nascent field.

9 TAXONOMY SURVEY DISCUSSION AND CONCLUSION

VAM-HRI Trends in Robot Domain

As seen in Figure 6, the majority of VAM-HRI papers surveyed did not fall into a specified robotic domain, where instead the VAM-HRI research aimed to improve robot interactions in a more general sense by enhancing common robot functionalities (i.e., navigation, manipulation, etc.) that are required across a wide range of domains.

The domain most frequently targeted by VAM-HRI interfaces is Remotely Operated Robots. VAM-HRI interfaces have found to hold major promise in enhancing robot operator capabilities, especially in terms of telepresence and situational awareness. VAM technology allows for a natural presentation of three-dimensional data (unlike that seen on traditional two-dimensional monitors) that robots collect from their remote environment using modern sensors, such as stereo cameras or LiDAR laser scans. Stereoscopic depth shown in HMD interfaces or 3D reconstructions presented in augmented virtuality interfaces (e.g., cyber-physical interfaces) are posed to be a revolutionary new way for teleoperating modern robots across of variety of industries, such as emergency first response (e.g., wildland firefighting, search and rescue) or space exploration where in both domains safety can be preserved by sending forth robots to explore hazardous environments instead of humans.

The second-most targeted domain in recent VAM-HRI research is that of collaborative industrial robots. This trend is to be expected as traditional industrial robots are frequently being replaced with a new generation of robots that are safer and more easily deployable and programmable in-house. In the past, traditional robots were housed behind fences due to their large size and ability to easily harm nearby humans, but collaborative robots are breaking down this physical barrier and working alongside their human counterparts. With humans now directly working with robots in industrial settings, more effort has been made to improve interactions with such robots, including the use of VAM-HRI interfaces to assist in communicating between human and robot during collaborative tasks.

¹<https://vam-hri.github.io/program/>.

²<https://github.com/TathagataChakraborti/survey-visualizer/issues/new/choose>.

Virtual Design Element (VDE) Usage Trends

As a second step in this analysis, we consider VDE use trends (see Figure 9). When looking at the usage data across the 175 VAM-HRI papers surveyed, we can see that Trajectories are the most commonly used VDE. Trajectory VDEs have shown great effectiveness in explicitly communicating robot motion plans and/or previously navigated paths taken by a robot. Spatial Previews, Waypoints, and Entity Location VDEs also have a high usage rate among the implemented VAM-HRI interfaces. These visualizations are used to convey robot intentions, preview object manipulations, or communicate to users the location of entities relevant to the current task. Like that of Trajectories, these three VDEs take advantage of the advanced capabilities of VAM displays to contextually place 3D visualizations within a user's environment, which have shown to be one of the main strengths held by VAM-HRI interfaces over existing traditional methods of robot interaction.

An additional takeaway from Figure 9 is that the Panels and Buttons VDE is heavily relied upon when designing VAM-HRI user interfaces that accept direct user input. The popularity of this VDE likely stems from its ease of implementation as well as its ability to leverage a design familiar to users due to the Panels and Buttons VDE drawing directly from traditional 2D graphical user interfaces that use panels/windows and buttons to control computer systems and applications (e.g., websites, touch screen phones, word processors). Meanwhile, the more novel virtual 3D Controllers VDE is used much less frequently, even though this VDE makes use of all three dimensions that VAM displays afford. The underutilization of this VDE presents an opportunity for future VAM-HRI research, where the development of novel virtual 3D input devices (that are unconstrained by the laws of physics, which is a significant restriction of physical 3D input devices) could enhance the ways in which robot operators interact with their robots.

When looking at the VDE usage data, it becomes apparent that there is an entire VDE Category that has only begun to be explored: Virtual Alterations, which were utilized by only 13 of the 175 VAM-HRI papers surveyed. This VDE Category holds both Superficial and Morphological VDE Classes that alter the robot's appearance at a surface level or their physical form to varying degrees. These VDEs hold promise in manipulating social interactions between humans and robots, such as by changing the robot's texture via a Cosmetic Alteration VDE (e.g., adding virtual fur to a robot to encourage children to touch and better engage with a robot in a classroom, altering a robot's surface to appear covered in sharp spikes to deter nearby humans from touching a dangerous or delicate robot). Additionally, rapid prototyping could be achieved by making use of Body Extension VDEs, where a prototype robot arm could be evaluated virtually prior to spending the time and money to physically manufacture a real-world prototype. This gap in VAM-HRI research represents a real opportunity for researchers to expand upon and improve how Virtual Alteration VDEs impact the field of VAM-HRI.

Finally, the External Sensor Images and Video VDE proved to be the most popular Environment-based Robot Comprehension Visualization. Although 3D environment reconstructions are still frequently used, they do not match the usage of the more traditional robot teleoperation and supervision methods that use images and video streams for operator telepresence and situational awareness. This high rate of External Sensor Images and Video VDE usage could stem from the enhanced stereoscopic 3D image and video viewing capabilities VAM-based interfaces hold. Additionally, this relatively higher usage is likely due to the ease of working with camera streams versus the difficulties that modern technology still has in creating the 3D reconstructions that are associated with External Sensor 3D Data and Sensed Spatial Region VDEs. As technology improves to allow for more accurate collection of 3D environment data (e.g., improved methods of SLAM with loop closure), the rendering of high-volume visual data (e.g., rendering large-scale dense point clouds of large buildings or outdoor environments that may have millions or billions of 3D RGB points), and the improvement of network capabilities to transport the large amounts of

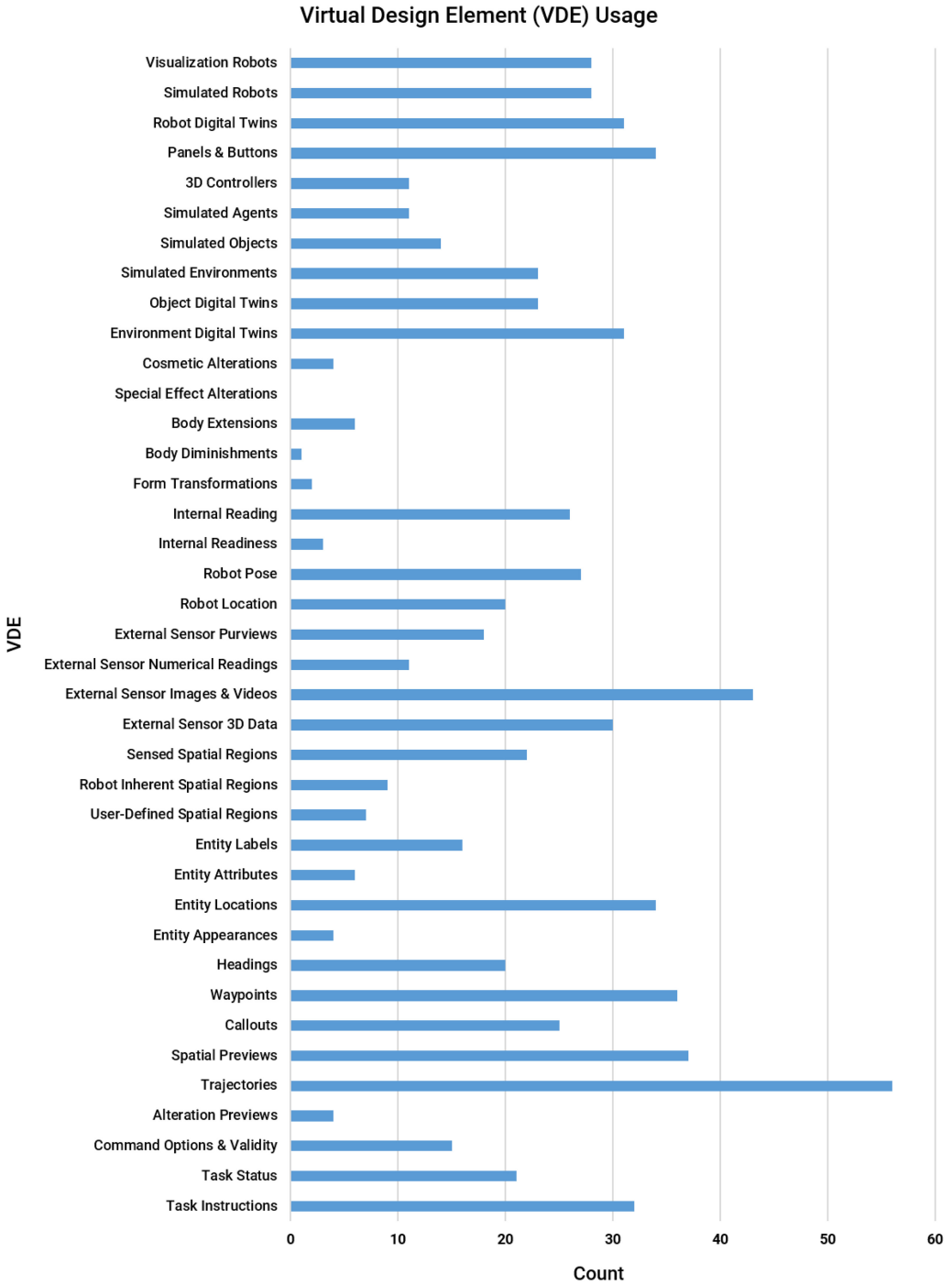


Fig. 9. The count of VDEs implemented in the 175 VAM-HRI research papers surveyed.

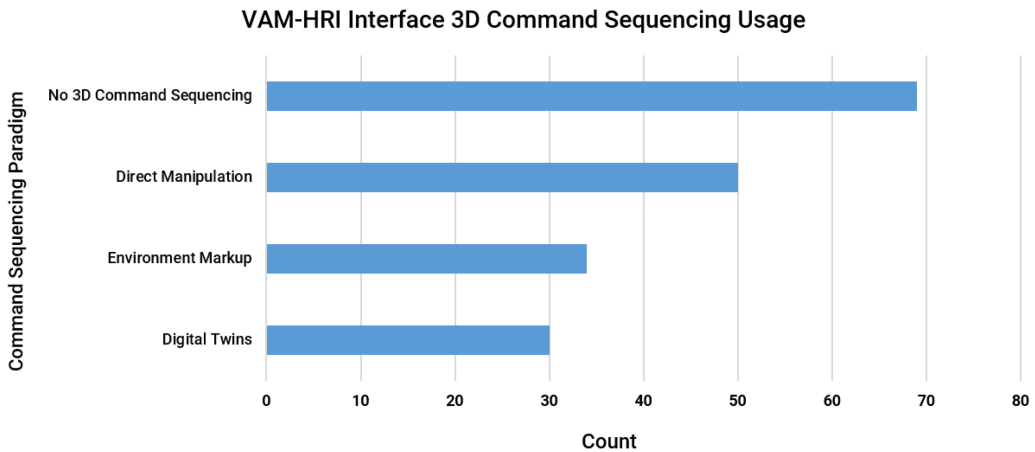


Fig. 10. The count of command sequencing paradigms used in the 175 VAM-HRI research papers surveyed.

data associated with 3D reconstructions (e.g., upgraded network infrastructure aided by edge computing), one would expect the usage of 3D reconstruction-based VDEs to increase across VAM-HRI interfaces.

VAM-based 3D Command Sequencing Paradigm Usage Trends

Our next analysis explores use trends of VAM-HRI Interface 3D Command Sequencing (see Figure 10). This usage data reveals that the Environment Markup and Digital Twin Paradigms were used at nearly the same frequency as one another, which is interesting due to the arguably orthogonal methods the two paradigms use to accomplish the same goal of 3D command sequencing. Although on its own, the more simplistic and more traditional Direct Manipulation Paradigm was the most commonly used paradigm, researchers more frequently opted to implement one of the more nascent and advanced methods of command sequencing: Environment Markup or Digital Twin.

Additionally, 40% of the surveyed papers did not allow for 3D command sequencing. This indicates the large portion of VAM-HRI research that either evaluates collaborative and/or passive user interactions that do not require user input for robot control or do not leverage the 3D capabilities of VAM displays to control their robotic systems.

HMD-based Remote Robot Teleoperation Paradigm Usage Trends

As a final step in this analysis, we examine the use trends of HMD-based Remote Robot Teleoperation Paradigms. As seen in Figure 11, augmented virtuality-based Cyber-Physical Interfaces have been the most explored paradigm for teleoperating remote robots with HMDs in VAM-HRI research. This paradigm leverages the 3D immersion provided by HMDs to enhance operator telepresence and situational awareness through the accurate rendering a digital twin of the remote robot, entities collocated with the remote robot, and/or the robot's remote environment in the form of a 3D reconstruction. The popularity of this paradigm can additionally be attributed to the robot-exocentric perspective provided by Cyber-Physical Interfaces allows users to "step out" of the robot's perspective (unlike that of interfaces that utilize camera video streams) and freely explore the recreated virtual scene from any angle with HMD-based body and head tracking, which helps mitigate issues such as object occlusion when line-of-sight is blocked by the robot's appendages or chassis.

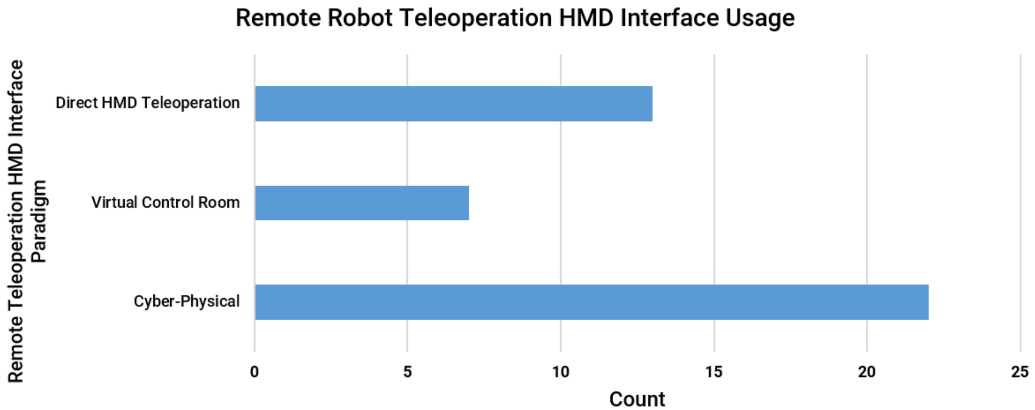


Fig. 11. The count of the remote robot teleoperation HMD interface paradigms implemented in the 42 VAM-HRI research papers (out of 175) that were used for teleoperating remotely located robots.

In contrast, the Virtual Control Room Paradigm is the least-explored HMD-based teleoperation paradigm and is implemented the most infrequently, even in regard to the more naive Direct HMD Teleoperation Paradigm, which is implemented at nearly double the rate. This is surprising due to the many benefits that Virtual Control Rooms hold over that of the directly streaming stereo video streams to the HMDs lens. Virtual Control Rooms still provide robot-egocentric stereoscopic 3D video feeds (as seen in the Direct HMD Teleoperation method), while untethering user perspectives from the stereo video stream to mitigate nausea arising from proprioception miscues. Additionally, this augmented virtuality paradigm operates within a virtual environment that allows for capacity to implement additional interface capabilities, such as virtual Controller VDEs, Visualization Robot VDEs, or even virtual avatars for remote users to upgrade the Virtual Control Room into a collaborative virtual workspace. Although this paradigm is the least popular method of teleoperation based on our survey, there is great potential for further VAM-HRI research and advancement that fully leverages the capabilities this design paradigm has to offer robot operators.

Conclusion

Overall, in this article, we have explored a wide range of research that demonstrates that VAM-HRI is an active and rapidly growing research area. We believe, however, that this field is currently hindered by a lack of precise terminology and theoretical models that explain how work in the field may connect and build off each other. The taxonomy we have presented for identifying, grouping, and classifying key design elements across VAM-HRI systems helps to address this issue and highlights potential design elements that have yet to appear in the research literature, which may serve as fertile ground for future research. For example, in the creation of this taxonomy it was realized that Special Effect Alteration VDEs had yet to be explored in scientific literature; however, when viewing the VAM-HRI work done to date from the high-level view of the VDE Table, this hole in the research landscape is made clear. It is our hope that other researchers are able to use the table in a similar fashion to guide their own efforts, and that our work will help VAM-HRI grow into a mainstream field by providing researchers in the community with the necessary lexicon for easily understanding, describing, and referencing the designs in their own systems with other relevant work being performed, while also helping researchers reason about what areas require further exploration or represent entirely novel areas of inquiry.

Additionally, the VAM VDE Taxonomy can be used as a catalogue and/or cookbook for robot designers interested in enhancing their robots and their interactions through VAM technology. We

envision developers being able to see all the VDEs available, including their hierarchical categories and classes, when deploying VAM-HRI systems and being able to pick and choose the VDE(s) that best addresses their VAM-HRI design's challenges and purpose.

Finally, it is important to note that as research and technology in this emerging field becomes more mature, the taxonomy presented in this article is destined to change and grow as time goes on. This taxonomy thus serves as a jumping off point for organizing and inspiring the future work performed within this field.

REFERENCES

- [1] Jordan Allspaw, Jonathan Roche, Nicholas Lemiesz, Michael Yannuzzi, and Holly A. Yanco. 2018. Remotely teleoperating a humanoid robot to perform fine motor tasks with virtual reality. In *1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI'18)*.
- [2] Rasmus S. Andersen, Ole Madsen, Thomas B. Moeslund, and Heni Ben Amor. 2016. Projecting robot intentions into human environments. In *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'16)*. IEEE, 294–301.
- [3] Stephanie Arévalo Arboleda, Tim Dierks, Franziska Rücker, and Jens Gerken. 2020. There's more than meets the eye: Enhancing robot control through augmented visual cues. In *ACM/IEEE International Conference on Human-Robot Interaction*. 104–106.
- [4] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robot. Auton. Syst.* 57, 5 (2009), 469–483.
- [5] David F. Arppe. 2020. UniNet: A mixed reality driving simulator. University of Ontario Institute of Technology (Canada), 2020. <https://www.proquest.com/docview/2576937121?pq-origsite=gscholar&fromopenview=true>.
- [6] Giancarlo Avale, Francesco De Pace, Claudio Fornaro, Federico Manuri, and Andrea Sanna. 2019. An augmented reality system to support fault visualization in industrial robotic tasks. *IEEE Access* 7 (2019), 132343–132359.
- [7] Kim Baraka. 2016. *Effective Non-verbal Communication for Mobile Robots Using Expressive Lights*. Dissertation, Carnegie Mellon University, Pittsburgh, PA.
- [8] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. 2020. *Human-robot Interaction: An Introduction*. Cambridge University Press.
- [9] Antal K. Beczcy, Won S. Kim, and Steven C. Venema. 1990. The phantom robot: Predictive displays for teleoperation with time delay. In *IEEE International Conference on Robotics and Automation*. IEEE, 546–551.
- [10] Gabriele Bolano, Christian Juelg, Arne Roennau, and Ruediger Dillmann. 2019. Transparent robot behavior using augmented reality in close human-robot interaction. In *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN'19)*. IEEE, 1–7.
- [11] A. Mejías Borrero and J. M. Andújar Márquez. 2012. A pilot study of the effectiveness of augmented reality to enhance the use of remote labs in electrical engineering education. *J. Sci. Educ. Technol.* 21, 5 (2012), 540–557.
- [12] J. Bosch, P. Ridao Rodríguez, R. García, and N. Gracías. 2016. Towards omnidirectional immersion for ROV teleoperation. In *XXXVII Jornadas de Automática*. Comité Español de Automática, 293–299.
- [13] Grigore C. Burdea. 1996. Virtual reality and robotics in medicine. In *5th IEEE International Workshop on Robot and Human Communication*. IEEE, 16–25.
- [14] Yuanzhi Cao, Zhuangying Xu, Terrell Glenn, Ke Huo, and Karthik Ramani. 2018. Ani-Bot: A modular robotics system supporting creation, tweaking, and usage with mixed-reality interactions. In *12th International Conference on Tangible, Embedded, and Embodied Interaction*. 419–428.
- [15] Wesley P. Chan, Camilo Perez Quintero, Matthew K. X. J. Pan, Maram Sakr, H. F. Machiel Van der Loos, and Elizabeth Croft. 2018. A multimodal system using augmented reality, gestures, and tactile feedback for robot trajectory programming and execution. In *ICRA Workshop on Robotics in Virtual Reality*. 21–25.
- [16] Kishan Chandan, Vidisha Kudalkar, Xiang Li, and Shiqi Zhang. 2019. Negotiation-based human-robot collaboration via augmented reality. *arXiv preprint arXiv:1909.11227* (2019).
- [17] Chih-Wei Chang, Jih-Hsien Lee, Chin-Yeh Wang, and Gwo-Dong Chen. 2010. Improving the authentic learning experience by integrating robots into the mixed-reality environment. *Comput. Educ.* 55, 4 (2010), 1572–1578.
- [18] Mark Cheli, Jivko Sinapov, Ethan E. Danahy, and Chris Rogers. 2018. Towards an augmented reality framework for k-12 robotics education. In *1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI'18)*.
- [19] Howard Chizeck et al. 2019. *Telerobotic Control in Virtual Reality*. Ph.D. Dissertation, University of Washington.
- [20] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. 2020. SPECTER: Document-level representation learning using citation-informed transformers. In *Meeting of the Association for Computational Linguistics*.

- [21] Francesco De Pace, Federico Manuri, Andrea Sanna, and Davide Zappia. 2018. An augmented interface to display industrial robot faults. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 403–421.
- [22] Sjoerd C. de Vries and Pieter Padmos. 1997. Steering a simulated unmanned aerial vehicle using a head-slaved camera and HMD. In *Head-Mounted Displays II*, Vol. 3058. International Society for Optics and Photonics, 24–33.
- [23] Elijs Dima, Kjell Brunnström, Märten Sjöström, Mattias Andersson, Joakim Edlund, Mathias Johanson, and Tahir Qureshi. 2020. Joint effects of depth-aiding augmentations and viewing positions on the quality of experience in augmented telepresence. *Qual. User Exper.* 5, 1 (2020), 2.
- [24] Anca D. Dragan, Kenton C. T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and predictability of robot motion. In *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI'13)*. IEEE, 301–308.
- [25] Jonatan S. Dyrstad, Elling Ruud Øye, Annette Stahl, and John Reidar Mathiassen. 2018. Teaching a robot to grasp real fish by imitation learning from a human supervisor in virtual reality. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 7185–7192.
- [26] Rolando Fernandez, Nathan John, Sean Kirmani, Justin Hart, Jivko Sinapov, and Peter Stone. 2018. Passive demonstrations of light-based robot signals for improved human interpretability. In *27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'18)*. IEEE, 234–239.
- [27] Jared A. Frank, Matthew Moorhead, and Vikram Kapila. 2016. Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks. In *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'16)*. IEEE, 302–307.
- [28] Jared A. Frank, Matthew Moorhead, and Vikram Kapila. 2017. Mobile mixed-reality interfaces that enhance human-robot interaction in shared spaces. *Front. Robot. AI* 4 (2017), 20.
- [29] Eckhard Freund and Juergen Rossmann. 1999. Projective virtual reality: Bridging the gap between virtual reality and robotics. *IEEE Trans. Robot. Automat.* 15, 3 (1999), 411–422.
- [30] Samir Yitzhak Gadre, Eric Rosen, Gary Chien, Elizabeth Phillips, Stefanie Tellex, and George Konidaris. 2019. End-user robot programming using mixed reality. In *International Conference on Robotics and Automation (ICRA'19)*. IEEE, 2707–2713.
- [31] Ramsundar Kalpagam Ganesan, Yash K. Rathore, Heather M. Ross, and Heni Ben Amor. 2018. Better teaming through visual cues: How projecting imagery in a workspace can improve human-robot collaboration. *IEEE Robot. Automat. Mag.* 25, 2 (2018), 59–71.
- [32] Abraham Prieto García, Gervasio Varela Fernández, Blanca María Priego Torres, and Fernando López-Peña. 2011. Mixed reality educational environment for robotics. In *IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems*. IEEE, 1–6.
- [33] Fabrizio Ghiringhelli, Jérôme Guzzi, Gianni A. Di Caro, Vincenzo Caglioti, Luca M. Gambardella, and Alessandro Giusti. 2014. Interactive augmented reality for understanding and analyzing multi-robot systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1195–1201.
- [34] Jason M. Gregory, Christopher Reardon, Kevin Lee, Geoffrey White, Ki Ng, and Caitlyn Sims. 2019. Enabling intuitive human-robot teaming using augmented reality and gesture control. *arXiv preprint arXiv:1909.06415* (2019).
- [35] Thomas Groechel, Zhonghao Shi, Roxanna Pakkar, and Maja J. Mataric. 2019. Using socially expressive mixed reality arms for enhancing low-expressivity robots. In *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN'19)*. IEEE, 1–8.
- [36] Sunao Hashimoto, Akihiko Ishida, Masahiko Inami, and Takeo Igarashi. 2011. TouchMe: An augmented reality based remote robot manipulation. In *21st International Conference on Artificial Reality and Telexistence*.
- [37] Hooman Hedayati, Michael Walker, and Daniel Szafrir. 2018. Improving collocated robot teleoperation with augmented reality. In *ACM/IEEE International Conference on Human-Robot Interaction*. 78–86.
- [38] K. Higuchi and J. Rekimoto. 2013. Flying head: a head motion synchronization mechanism for unmanned aerial vehicle control. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2029–2038. https://dl.acm.org/doi/abs/10.1145/2468356.2468721?casa_token=5N3-qIBVXokAAAAA:KJNTL_sdkJWt0wJqFabxB-VyD_ONMtaWUCXZoSMMaAjVWzFTWWmyAgAoSg46gOT9T9xWY9ZVrMzP.
- [39] B. Hine, P. Hontalas, T. Fong, L. Piguet, E. Nygren, and A. Kline. 1995. VEVI: A virtual environment teleoperations interface for planetary exploration. *SAE Transactions* 1 (1995), 615–28.
- [40] Burkhard Hoppenstedt, Thomas Witte, Jona Ruof, Klaus Kammerer, Matthias Tichy, Manfred Reichert, and Rüdiger Pryss. 2019. Debugging quadcopter trajectories in mixed reality. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 43–50.
- [41] Janna Huuskonen and Timo Oksanen. 2019. Augmented reality for supervising multirobot system in agricultural field operation. *IFAC-PapersOnLine* 52, 30 (2019), 367–372.
- [42] Roman Ibrahimov, Evgeny Tsykunov, Vladimir Shirokun, Andrey Somov, and Dzmitry Tsetserukou. 2019. DronePick: Object picking and delivery teleoperation with the drone controlled by a wearable tactile display. In *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN'19)*. IEEE, 1–6.

- [43] Kentaro Ishii, Shengdong Zhao, Masahiko Inami, Takeo Igarashi, and Michita Imai. 2009. Designing laser gesture interface for robot control. In *IFIP Conference on Human-Computer Interaction*. Springer, 479–492.
- [44] Michael L. Iuzzolino, Michael E. Walker, and Daniel Szafrir. 2018. Virtual-to-real-world transfer learning for robots on wilderness trails. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 576–582.
- [45] Wafa Johal, Olguta Robu, Amaury Dame, Stephane Magnenat, and Francesco Mondada. 2019. Augmented robotics for learners: A case study on optics. In *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN'19)*. IEEE, 1–6.
- [46] Linh Kästner and Jens Lambrecht. 2019. Augmented-reality-based visualization of navigation data of mobile robots on the Microsoft HoloLens—Possibilities and limitations. In *IEEE International Conference on Cybernetics and Intelligent Systems (CIS'19) and IEEE Conference on Robotics, Automation and Mechatronics (RAM'19)*. IEEE, 344–349.
- [47] David Kent, Carl Saldanha, and Sonia Chernova. 2017. A comparison of remote robot teleoperation interfaces for general object manipulation. In *ACM/IEEE International Conference on Human-Robot Interaction*. 371–379.
- [48] Won Kim, Frank Tendick, and Lawrence W. Stark. 1987. Visual enhancements in pick-and-place tasks: Human operators controlling a simulated cylindrical manipulator. *IEEE J. Robot. Automat.* 3, 5 (1987), 418–425.
- [49] Kazuhiko Kobayashi, Koichi Nishiwaki, Shinji Uchiyama, Hiroyuki Yamamoto, Satoshi Kagami, and Takeo Kanade. 2007. Overlay what humanoid robot perceives and thinks to the real-world by mixed reality system. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 275–276.
- [50] Tomáš Kot and Petr Novák. 2014. Utilization of the Oculus Rift HMD in mobile robot teleoperation. In *Applied Mechanics and Materials*, Vol. 555. Trans Tech Publications, 199–208.
- [51] Dennis Krupke, Lasse Einig, Eike Langbehn, Jianwei Zhang, and Frank Steinicke. 2016. Immersive remote grasping: Realtime gripper control by a heterogenous robot control system. In *22nd ACM Conference on Virtual Reality Software and Technology*. 337–338.
- [52] Dennis Krupke, Frank Steinicke, Paul Lubos, Yannick Jonetzko, Michael Görner, and Jianwei Zhang. 2018. Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 1–9.
- [53] Florian Leutert, Christian Herrmann, and Klaus Schilling. 2013. A spatial augmented reality system for intuitive display of robotic data. In *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI'13)*. IEEE, 179–180.
- [54] Jiannan Li, Ravin Balakrishnan, and Tovi Grossman. 2020. StarHopper: A touch interface for remote object-centric drone navigation. *Graphics Interface 2020*. <https://openreview.net/forum?id=BviYjfnlk>.
- [55] Jeffrey I. Lipton, Aidan J. Fay, and Daniela Rus. 2017. Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robot. Automat. Lett.* 3, 1 (2017), 179–186.
- [56] Elias Matsas, George-Christopher Vosniakos, and Dimitris Batras. 2018. Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality. *Robot. Comput.-Integ. Manuf.* 50 (2018), 168–180.
- [57] Johannes Meyer, Alexander Sendobry, Stefan Kohlbrecher, Uwe Klingauf, and Oskar Von Stryk. 2012. Comprehensive simulation of quadrotor UAVs using ROS and Gazebo. In *International Conference on Simulation, Modeling, and Programming for Autonomous Robots*. Springer, 400–411.
- [58] Sebastian Meyer zu Borgsen, Patrick Renner, Florian Lier, Thies Pfeiffer, and Sven Wachsmuth. 2018. Improving human-robot handover research by mixed reality techniques. In *Inaugural International Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction*.
- [59] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* 77, 12 (1994), 1321–1329.
- [60] Paul Milgram, Anu Rastogi, and Julius J. Grodski. 1995. Telerobotic control using augmented reality. In *4th IEEE International Workshop on Robot and Human Communication*. IEEE, 21–29.
- [61] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. 1993. Applications of augmented reality for human-robot communication. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*, Vol. 3. IEEE, 1467–1472.
- [62] Shohei Mori, Sei Ikeda, and Hideo Saito. 2017. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Trans. Comput. Vis. Applic.* 9, 1 (2017), 1–14.
- [63] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D. Tardos. 2015. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* 31, 5 (2015), 1147–1163.
- [64] Abdeldjalil Naceri, Dario Mazzanti, Joao Bimbo, Domenico Prattichizzo, Darwin G. Caldwell, Leonardo S. Mattos, and Nikhil Deshpande. 2019. Towards a virtual reality interface for remote robotic teleoperation. In *19th International Conference on Advanced Robotics (ICAR'19)*. IEEE, 284–289.
- [65] Aditya Nawab, Keshav Chintamani, Darin Ellis, Gregory Auner, and Abhilash Pandya. 2007. Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In *IEEE Virtual Reality Conference*. IEEE, 263–266.

- [66] Donald A. Norman and Stephen W. Draper. 1986. *User Centered System Design; New Perspectives on Human-computer Interaction*. L. Erlbaum Associates Inc.
- [67] Dan R. Olsen Jr and Stephen Bart Wood. 2004. Fan-out: Measuring human control of multiple robots. In *SIGCHI Conference on Human Factors in Computing Systems*. 231–238.
- [68] Shayegan Omidshafiei, Ali-Akbar Agha-Mohammadi, Yu Fan Chen, Nazim Kemal Üre, Jonathan P. How, John L. Vian, and Rajeev Surati. 2015. MAR-CPS: Measurable augmented reality for prototyping cyber-physical systems. In *AIAA Infotech@ Aerospace*.
- [69] André Pereira, Elizabeth J. Carter, Iolanda Leite, John Mars, and Jill Fain Lehman. 2017. Augmented reality dialog interface for multimodal teleoperation. In *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'17)*. IEEE, 764–771.
- [70] Luis Pérez, Eduardo Diez, Rubén Usamentiaga, and Daniel F. García. 2019. Industrial robot control and operator training using virtual reality interfaces. *Comput. Industr.* 109 (2019), 114–120.
- [71] Long Qian, Anton Deguet, Zerui Wang, Yun-Hui Liu, and Peter Kazanzides. 2019. Augmented reality assisted instrument insertion and tool manipulation for the first assistant in robotic surgery. In *International Conference on Robotics and Automation (ICRA'19)*. IEEE, 5173–5179.
- [72] Camilo Perez Quintero, Oscar Ramirez, and Martin Jägersand. 2015. VIBI: Assistive vision-based interface for robot manipulation. In *IEEE International Conference on Robotics and Automation (ICRA'15)*. IEEE, 4458–4463.
- [73] Christopher Reardon, Kevin Lee, and Jonathan Fink. 2018. Come see this! Augmented reality to enable human-robot cooperative search. In *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR'18)*. IEEE, 1–7.
- [74] Christopher Reardon, Kevin Lee, John G. Rogers, and Jonathan Fink. 2019. Augmented reality for human-robot teaming in field environments. In *International Conference on Human-Computer Interaction*. Springer, 79–92.
- [75] Sandro Ropelato, Fabio Zünd, Stéphane Magnenat, Marino Menozzi, and Robert Sumner. 2018. Adaptive tutoring on a virtual reality driving simulator. *Int. Ser. Inf. Syst. Manag. Creat. EMedia* 2017, 2 (2018), 12–17.
- [76] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2020. Communicating robot arm motion intent through mixed reality head-mounted displays. In *Robotics Research*. Springer, 301–316.
- [77] Ane San Martín and Johan Kildal. 2019. Audio-visual AR to improve awareness of hazard zones around robots. In *CHI Conference on Human Factors in Computing Systems*. 1–6.
- [78] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W. McOwan, and Ana Paiva. 2011. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *6th International Conference on Human-Robot Interaction*. 305–312.
- [79] A. Segal, D. Haehnel, and S. Thrun. 2009. Generalized-icp. *Robotics: Science and Systems* 2, 4 (2009), 435.
- [80] Moondeep C. Shrestha, Tomoya Onishi, Ayano Kobayashi, Mitsuhiro Kamezaki, and Shigeki Sugano. 2018. Communicating directional intent in robot navigation using projection indicators. In *27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'18)*. IEEE, 746–751.
- [81] Elena Sibirtseva, Dimosthenis Kontogiorgos, Olov Nykvist, Hakan Karaoguz, Iolanda Leite, Joakim Gustafson, and Danica Kragic. 2018. A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction. In *27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'18)*. IEEE, 43–50.
- [82] Daniele Sportillo, Alexis Paljic, and Luciano Ojeda. 2018. Get ready for automated driving using virtual reality. *Accid. Anal. Prevent.* 118 (2018), 102–113.
- [83] Dennis Sprute, Klaus Tönnies, and Matthias König. 2019. A study on different user interfaces for teaching virtual borders to mobile robots. *Int. J. Soc. Robot.* 11, 3 (2019), 373–388.
- [84] Dennis Sprute, Philipp Viertel, Klaus Tönnies, and Matthias König. 2019. Learning virtual borders through semantic scene understanding and augmented reality. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 4607–4614.
- [85] Michael Stilman, Philipp Michel, Joel Chestnutt, Koichi Nishiwaki, Satoshi Kagami, and James Kuffner. 2005. *Augmented Reality for Robot Development and Experimentation*. Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-05-55 2, 3 (2005).
- [86] Patrick Stotko, Stefan Krumpfen, Max Schwarz, Christian Lenz, Sven Behnke, Reinhard Klein, and Michael Weinmann. 2019. A VR system for immersive teleoperation and live exploration with a mobile robot. *arXiv preprint arXiv:1908.02949* (2019).
- [87] Da Sun, Andrey Kiselev, Qianfang Liao, Todor Stoyanov, and Amy Loutfi. 2020. A new mixed-reality-based teleoperation system for telepresence and maneuverability enhancement. *IEEE Trans. Hum.-Mach. Syst.* 50, 1 (2020), 55–67.
- [88] Ivan E. Sutherland. 1965. The ultimate display. *Multim.: Wagner Virt. Real.* 1 (1965).
- [89] Ivan E. Sutherland. 1968. A head-mounted three dimensional display. In *Proceedings of the December 9–11, 1968, fall Joint Computer Conference, part I*. ACM, 757–764.

- [90] Daniel Szafr, Bilge Mutlu, and Terrence Fong. 2014. Communication of intent in assistive free flyers. In *ACM/IEEE International Conference on Human-Robot Interaction*. 358–365.
- [91] Daniel Szafr, Bilge Mutlu, and Terrence Fong. 2015. Communicating directionality in flying robots. In *10th ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. IEEE, 19–26.
- [92] A. V. Taylor, A. Matsumoto, E. J. Carter, A. Plopski, and H. Admoni. 2020. Diminished reality for close quarters robotic telemanipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 11531–11538.
- [93] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. 2014. *Asking for Help Using Inverse Semantics*. Robotics: Science and Systems Foundation.
- [94] Jaryd Urbani, Mohammed Al-Sada, Tatsuo Nakajima, and Thomas Höglund. 2018. Exploring augmented reality interaction for everyday multipurpose wearable robots. In *IEEE 24th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'18)*. IEEE, 209–216.
- [95] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafr. 2018. Communicating robot motion intent with augmented reality. In *ACM/IEEE International Conference on Human-Robot Interaction*. 316–324.
- [96] Michael E. Walker, Hooman Hedayati, and Daniel Szafr. 2019. Robot teleoperation with augmented reality virtual surrogates. In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI'19)*. IEEE, 202–210.
- [97] David Whitney, Eric Rosen, Daniel Ullman, Elizabeth Phillips, and Stefanie Tellex. 2018. ROS reality: A virtual reality framework using consumer-grade hardware for ROS-enabled robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 1–9.
- [98] Luc Wijnen, Séverin Lemaignan, and Paul Bremner. 2020. Towards using virtual reality for replicating HRI studies. In *ACM/IEEE International Conference on Human-Robot Interaction*. 514–516.
- [99] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 2019. Mixed reality deictic gesture for multi-modal robot communication. In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI'19)*. IEEE, 191–201.
- [100] Tom Williams, Leanne Hirshfield, Nhan Tran, Trevor Grant, and Nicholas Woodward. 2020. Using augmented reality to better study human-robot interaction. In *International Conference on Human-Computer Interaction*. Springer, 643–654.
- [101] Tom Williams, Daniel Szafr, and Tathagata Chakraborti. 2019. The reality-virtuality interaction cube. In *2nd International Workshop on Virtual, Augmented, and Mixed Reality for HRI*.
- [102] T. Williams, D. Szafr, T. Chakraborti, and H. Ben Amor. 2018. Virtual, augmented, and mixed reality for human-robot interaction. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*. 403–404.
- [103] Tom Williams, Daniel Szafr, Tathagata Chakraborti, and Heni Ben Amor. 2018. Virtual, augmented, and mixed reality for human-robot interaction. In *ACM/IEEE International Conference on Human-Robot Interaction*. 403–404.
- [104] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. 2017. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. *arXiv preprint arXiv:1711.00199* (2017).
- [105] Jingxin Zhang, Omar Janeh, Nikolaos Katzakis, Dennis Krupke, and Frank Steinicke. 2019. Evaluation of proxemics in dynamic interaction with a mixed reality avatar robot. In *International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments (ICAT-EGVE'19)*. 37–44.
- [106] Stefanie Zollmann, Christof Hoppe, Tobias Langlotz, and Gerhard Reitmayr. 2014. FlyAR: Augmented reality supported micro aerial vehicle navigation. *IEEE Trans. Visualiz. Comput. Graph.* 20, 4 (2014), 560–568.
- [107] M. Zolotas and Y. Demiris. 2019. Towards explainable shared control using augmented reality. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 3020–3026.
- [108] Tathagata Chakraborti, Jungkoo Kang, Christian Muise, Sarath Sreedhathan, Michael Walker, Daniel Szafr, and Tom Williams. 2022. A tool for exploring data in academic survey papers.

Received 17 June 2021; revised 17 January 2023; accepted 13 April 2023