2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 1-5, 2023. Detroit, USA

Human Preferred Augmented Reality Visual Cues for Remote Robot Manipulation Assistance: from Direct to Supervisory Control

Achyuthan Unni Krishnan, Tsung-Chi Lin and Zhi Li[1]

***Abstract*— When humans control or supervise remote robot manipulation, augmented reality (AR) visual cues overlaid on the remote camera video stream can effectively enhance human's remote perception of task and robot states, and comprehension of the robot autonomy's capability and intent. In this work, we conducted a user study (N=18) to investigate: (RQ1) what AR cues humans prefer when controlling the robot with various levels of autonomy, and (RQ2) whether this preference can be influenced by the way humans learn to use the interface. We provided AR visual cues of various types (e.g., motion guidance, obstacle indicator, target hint, autonomy activation and intent) to assist humans to pick and place an object around an obstacle on a counter workspace. We found that: 1) Participants prefer different types of AR cues based on the level of robot autonomy; 2) The AR cues the participants prefer to use after hands-on robot operation converged to the recommendation of experienced users, and may largely differ from their initial selection based on video instruction.**

I. Introduction

This work aims to investigate what augmented reality (AR) visual cues humans prefer to use when controlling or supervising remote robot manipulation. To enhance remote perception, AR visual cues are overlaid on the video stream from remote cameras to indicate the robot and task states, and the spatial relationship in a 3D environment, which may be difficult for human operators to estimate precisely. They are also used to indicate the motion, action, path and task that robot autonomy plans to perform, in order to enhance human's understanding of robot's intent, behavior and capabilities. Thus far, effective AR visual cues are mostly developed case-by-case for remote robot manipulation under direct to supervisory control. It is still not clear what AR visual cues humans need or prefer to use to control a remote manipulator robot with various levels of autonomy.

To this end, we proposed systematic AR visual cues for remote robot manipulation assistance. These AR cues can guide human's control of robot motion toward the target and around the obstacles in a 3D workspace, and indicate whether the robot autonomy is activated,and its planned motion or action. We conducted a user study where the participants (N=18) controlled a remote manipulator robot with adjustable autonomy to move around an obstacle to pick and place an object on the counter space. The robot can operate under direct human control, or assist humans with autonomy to avoid obstacles and environmental constraints, to perform error-prone precise manipulation actions, or plan and execute the entire robot motion to pick and place an object under human supervision. We compared the AR cues the participants chose for each level of autonomy, to the choices of experienced interface users. We also compared the participants' initial choices based on video demos of AR features used to understand the interface to their choices after hands-on practice with the interface. Our results show that the preference for AR visual cues varied with the level of autonomy used to control the robot. From direct to supervisory control, the preference shifted from AR visual cues that guide their control of robot motion to cues that communicate the robot's intended action and path plan. Additionally, participants tended to change their initial preference for the AR visual cues after hands-on practice tending to agree with the recommendation of experienced users.

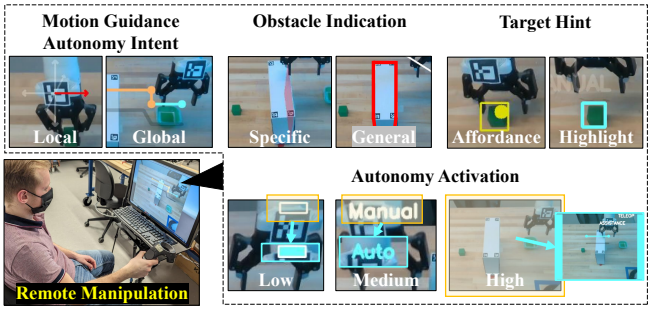

Fig. 1: Proposed AR visual cues to assist humans to control or supervise remote robot manipulation, and to communicate the robot autonomy's activation, capabilities, and intents.

II. Related Work

Our proposed AR visual interface aims to enable humans to effectively assist the robot with limited autonomy for structured pick-and-place object manipulation to handle a wide range of unstructured, complex manipulation tasks in a cluttered environment. Although the robot autonomy cannot handle the entire task, it can reliably enforce motion constraints, and plan and execute autonomous actions. Effective human-robot collaboration for such tasks enables humans and robots to contribute complementary skills to improve the overall task performance, reduce the human workload, and task complexity for robot autonomy [1]–[3]. Related work shows that humans can operate the robot to separate the cluttered and entangled objects based on their task knowledge and experience, rearrange the objects and workspace into more organized and predictable positions, in order to reduce the robot sensor occlusion and facilitate the robot's autonomous semantic segmentation of workspace [1]. To reduce the complexity of autonomous motion planning,

[1] Robotics Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA {aunnikrishnan,tlin2,zli11}@wpi.edu

2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-9190-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IROS55552.2023.10341969

humans can guide a robot towards target locations or objects across cluttered workspace [4], control nonprehensile object manipulation [5], select the grasping points to handle (deformable) objects [6]. In shared and supervisory control, humans can select or confirm the robot's actions [7], review and revise robot action sequence or task plan [8], and supervise the autonomy's execution [9].

Recent related work in literature has provided a taxonomy for the virtual, augmented, and mixed reality for human-robot interaction [10]. Overall, the AR visual cues can be augmented on the *robot*, *interface* and *environment*, in order to visualize the robot's internal and external states (e.g., internal reading and readiness, robot pose and location), as well as the robot's comprehension of *environment* (e.g., the purview, numerical readings, videos and images, 3D data from external sensors, robot-sensed/internal/user-defined spatial region,), *entities* (e.g., entity labels, attributes, locations and appearance), and *tasks* (e.g., heading, waypoint, call-out, trajectory, spatial preview, trajectory, alternation preview, command options, task statuses and instructions). Related work also recommends various effective integration of AR visual cues for manipulation tasks in the 3D environment (e.g., [11]–[16]), yet these case-by-case designs did not reveal how the AR visual cues should be designed for a human-robot collaborative manipulation with various levels of autonomy. Another recent work revealed similarity in the design principles for data visualization and AR for assisting remote manipulation control [17]. To enhance the operator's remote perception, it recommended to *find relevant information* by data salience and clutter, *synthesize data across sources* by comparison and multiple views, *identify anomalies* by data provenance and statistical estimation, *make predictions* by uncertainty and temporal data, *assess risks* by direct attention and value estimation. While the design principles are valid for manipulation tasks, it is unclear how to apply them to the AR visual designed for various human-robot collaborations for remote robot manipulation.

III. Proposed Method

Here we introduce the remote manipulation system and the human-robot task division in each control mode, in order to provide context for AR visual cue design, and present our systematic AR visual cue design for the human-robot collaborative control at four different levels of autonomy.

Remote Manipulation System: Our prior work [18] had integrated a robotic system with an adjustable level of autonomy for remote, unstructured manipulation. This system uses a HTC Vive handheld controller to control a 7 degrees of freedom Kinova Gen 3 robotic manipulator with a two-fingered Robotiq gripper. A desktop monitor displayed a graphical user interface (GUI) on a Unity 3D window (1920 $\times$ 1150 pixel) to stream the video from the workspace camera (RealSense D435f) at 30 Hz.

Control Modes: Table I illustrates the human-robot task division in each control mode. In the **Direct Control** mode, a human manually controls the robot through the entire task. In the **Assisted Control** mode, robot autonomy uses a virtual fixture to constrain human-controlled robot motion in directions that avoid collision with environmental constraints and obstacles as well as move to grasp/place an object. In the **Shared Control** mode, the robot assists human-controlled gross manipulation motions (e.g., approaching or moving an object) with automated motions for collision avoidance, and for precise manipulation (e.g., grasping or placing) when the robot end-effector is close enough to the target object or location. In the **Supervisory Control** mode, robot autonomy plans the motion for the entire pick-and-place task, while humans supervise its execution. The robot's autonomous perception using AruCo markers [19] are available for all the control modes, in order to locate the target object, container, obstacle, and track the robot end-effector.

TABLE I: Human (**H**) and Robot (**R**) task division in each control mode.

Control Mode	Gross Manipulation	Obstacle Avoidance	Precise Manipulation	Autonomy Activation
Direct	H	H	H	N/A
Assisted	H/R	H/R	H/R	Auto
Shared	H/R	R	R	Auto
Supervisory	R	R	R	Auto

AR Visual Cues: Shown in Fig. 1, we propose five types of AR visual cues and representation options:

- **Motion Guidance** indicates the robot's instantaneous motion direction using a *3-axis arrows* overlaid on the robot end-effector or display the robot's suggested path [20].
- **Obstacle Indication** has the options to highlight close-to-collision features on the obstacle (e.g., plane, edge or vertex, as in [7]), and to display the obstacle's 3D bounding box (as in [21]).
- **Target Hint** provides the grasp/place *affordance* [15] by changing the color of the square around the target and the dot overlaid on the robot end-effector. Alternatively, the target can be *highlighted* [4] to provide an intent inference to the user.
- **Autonomy Activation** indicates if the robot autonomy has been activated. We provide three options for the representation of different visual salience, including a blue light (low salience, as in [22]), text with "AUTO" (medium salience, as in [23]), and blue bars on both sides of the camera view (high salience, as in [7]).
- **Autonomy Intention** has the option to indicate the robot's motion, action, and path plan (as in [22], [24]). The autonomy intention is displayed similar to the visual cues used for motion guidance.

TABLE II: AR visual cue choices recommended by experienced users.

Control Mode	Motion Guidance	Obstacle Indication	Target Hint	Autonomy Activation	Autonomy Intention
Direct	Arrow	Planes	Affordance	–	–
Assisted	Arrow	Box	None	Low	Path
Shared	Path	None	Highlight	High	Path
Supervisory	–	None	None	Medium	Path

Our user study allows participants to choose the types and options of AR visual cues for each control mode. Table II

shows the choices recommended by experienced users (users (N=5) with at least 100 hours of experience in remote manipulation.)

IV. User Study

Research Questions: We conduct a user study to investigate: **(RQ1)** what AR visual cues humans prefer when controlling the robot with various levels of autonomy, and **(RQ2)** whether this preference can be influenced by the way humans learn to use the interface.

Participants and Task: We recruited 18 participants (13 male and 5 female, 25.4 ± 6.9 years) from the WPI campus to perform a single object pick-and-place task with an obstacle between the object and the box. Shown in Fig. 2, participants controlled the robot to pick up an object on the other side of an obstacle (box of size 200 mm $\times$ 80 mm $\times$ 200 mm), and bring it back and place it in a small container. Participants performed the task under three different initial task states: the object location remained the same, while the container and robot were placed in three different ways such that the robot planned path would move around different sides of the obstacle to reach the object and container.

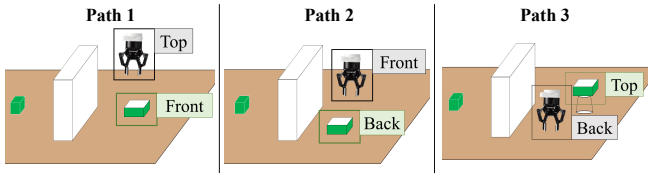

Fig. 2: Worskpace configurations for the pick and place experimental task. The robot end-effector starts towards the top, front and back of the container for Paths 1,2 and 3 respectively. The target container is towards the front, back and top of the obstacle for Paths 1,2 and 3 respectively.

Experiment Procedure: Participants were trained on the baseline control interface and three levels of assistive autonomy before the experiment. The experimenter then provided video instructions to demonstrate each AR feature and gather their preferences (Preference 1) for each level of autonomy. Participants performed a single trial of the task in Path 1 using their selected AR visual cues for all control modes. Following the completion of the task in Path 1, participants were given the opportunity to switch their preferences (Preference 2) based on their hands-on experience and then perform the same task again with Paths 2 and 3 for one trial each. Lastly, participants were required to perform the same task using the AR feature proposed by the experienced user for Paths 2 and 3 for one trial each and made their final selection of AR preferences (Preference 3). Note that a trial was skipped if the preferred AR combination was the same between Preference 1, 2, and 3.

Data Collection and Analysis: To assess control efficiency, we recorded the complete length of the trajectory covered by the handheld controller. Previous research has shown that the size of the pupil increases as the level of stress rises [25]. We utilized pupil diameter as a measure for estimating subject-specific cognitive workload. To establish a baseline for each participant, we instructed them to look at a blank screen for 30 seconds, assuming a stress-free state. The cognitive workload was calculated by finding the average difference between real-time pupil diameter and baseline value, which was then normalized by the maximum average distance between real-time pupil diameter and baseline value across all trials for each participant.

V. Results And Discussion

A post-hoc power analysis with (1-β) = 0.8 and α = 0.05 found an observed power of 0.85 (d = -1.05) for the participant size (N=18). For the comparisons in control efficiency and cognitive workload, we first used F-test to compare the variances of the two sample groups, and then used Student's t-test (equal variance) and Welch's t-test (unequal variance) with $p < .05$. Note that preferences 1, 2, and 3 are referred to as P1, P2, and P3 in this section.

A. AR Preferences for Each Level of Autonomy

Fig. 3.a presents participants' final selection (after using the recommendation of experienced users, refer to as P3) of each AR visual cue for different levels of robot autonomy.

Direct Control — Most of the participants (13 out of 18) selected local information (arrow) as a preferred way to continuously guide their motion to perform the task and avoid an obstacle. All the participants (18 out of 18) preferred having detailed (planes of the obstacle) information indicating possible collisions and specific (target affordance) methods to indicate if the position of the robot end-effector is good to perform the precise manipulation.

Assisted Control — Most of the participants (15 out of 18) still selected local information (arrow) as a preferred way to guide their motion, especially in the aspect of explicitly showing the required control direction while the assisted autonomy is activated. Half of the participants preferred notification of the activation of the autonomy with a low salience (light-up a small square) method while using global information (path) to be informed of the autonomy intention had been selected by most of the participants (14 out of 18). Most of the participants (11 out of 18) preferred having a general obstacle indication (highlight with a red boundary) that potentially provides a reason for the activation of assisted autonomy to avoid an obstacle. Most of the participants (11 out of 18), however, preferred not to have AR visual cues for precise manipulation with grasping and placing an object because they feel assisted autonomy will handle it and like to have minimal visual clutter on the screen.

Shared Control — In contrast to the direct and assisted control, most of the participants (12 out of 18) selected global information (path) as a preferred AR visual cue to guide their motion to help approach the autonomy zone around the targets and obstacle. Almost all the participants (17 out of 18) preferred the most salience (highlight the entire user interface with color and two thick bars) method to indicate the activation of autonomy and global information (path) to indicate the intent of autonomy. This way the participants could get a better sense of the timing of resuming control when the autonomy was completed. Most of the participants (13 out of 18) preferred to have no AR visual cue for

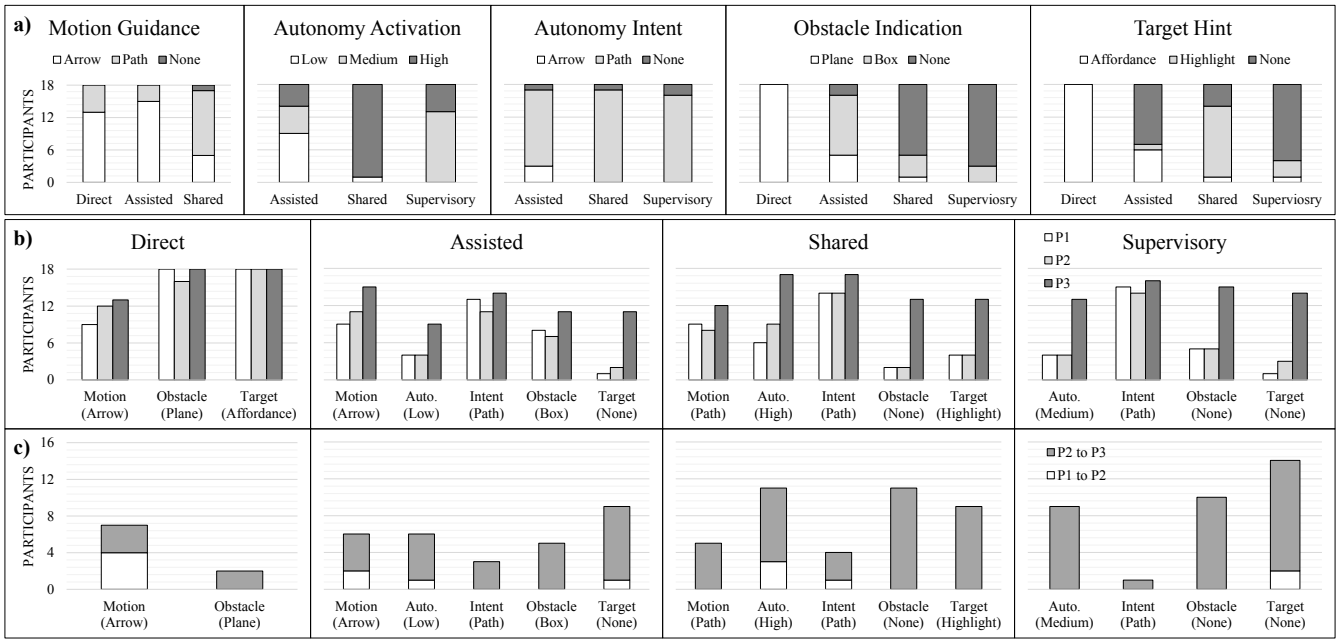

Fig. 3: a) The final selections of the participants for the different AR features for all the control modes; b) The number of participants who selected the recommended AR feature for all the control modes; c) The number of participants who changed their preferences from other AR features to the recommended AR feature when moving from P1-2 and P2-3.

obstacle indication as they feel the autonomy will handle it automatically, and use general information (highlight the current target) to ensure the autonomy is being applied to the correct target.

Supervisory Control — Most of the participants (13 and 16 out of 18 participants respectively) preferred having a medium salience (pop-up text) to indicate autonomy activation and global information (path) to demonstrate the autonomy intention. No AR visual cues for obstacle indication and target hint was selected as they were deemed not necessary by most of the participants (15 and 14 out of 18) as these will be handled by robot autonomy.

TABLE III: Usefulness of AR features for each control mode.

Interface	Priority
Direct	Motion Guidance>Obstacle>Target
Assisted	Motion Guidance>Autonomy>Intent>Target>Obstacle
Shared	Autonomy>Motion Guidance>Intent>Target>Obstacle
Supervisory	Intent>Autonomy>Target>Obstacle

Discussion — To address **RQ1**, in addition to the participants' final selection of AR features, Table III shows their subjective feedback on the usefulness rankings of AR features for each control mode. As the level of autonomy transitions from direct to supervisory control, we found that: (1) humans' priority for AR visual cues shifts ***from guiding robot control to communicating autonomy activation and intention*** based on their subjective feedback in Table III; (2) humans' preference for AR visual cues changes ***from providing local information to offering global guidance in robot control*** as indicated by a large portion of participants (13 and 12 out of 18) selecting arrow for direct control and the path for shared control. However, the use of ***global information to display the autonomy's intention*** remains consistent across all interfaces with autonomy, as more than 14 participants preferred having a planned path to display it; (3) the ***efficacy of the AR features that share the same purpose as the robot autonomy is decreased*** as observed by most of the participants preferring not to have any AR visual cues for both obstacle indication and target hint in supervisory control mode.

B. Influence of Interface Learning Method

Fig. 3.b shows the users' preference changes for each AR feature suggested by the experienced user from P1 to P3.

Video Instruction-Based (Initial Selection) — In **direct control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. All the participants already preferred having detailed information (planes of possible collision with the obstacle) for avoiding the obstacle and grasp/place affordance to assist precise manipulation. In **assisted control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. Few participants selected the low salience (light-up small square) method to be informed of the autonomy activation while most of the participants already chose the global information (planned path) to show the autonomy intention. Similar to direct control, most of the participants still preferred having detailed information (planes) over the general method (highlight obstacle boundary) to avoid the obstacle. Most of the participants chose grasp/place affordance to assist precise manipulation even if assisted autonomy was provided. In **shared control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. Few of the participants selected the high salience (highlight the entire screen with

bars) method to show the autonomy activation while most of the participants already chose the global information (planned path) to show the autonomy intention. Most of the participants selected the general information for obstacle avoidance, over having no AR visual cues, and grasp/place affordance to assist precise manipulation. In **supervisory control** mode, most of the participants preferred having the most salience method to indicate the activation of the autonomy while most of the participants already chose the global information (planned path) to show the autonomy intention. Similar to direct control, most of the participants selected the general information for obstacle avoidance over having no AR visual cues and grasp/place affordance to assist precise manipulation even the autonomy will handle both obstacle avoidance and precise manipulation.

Hands-On Engagement-Based (Intermediate Selection) — In **direct control** mode, a few participants changed their preferences for motion guidance from global information to local information in the form of arrows. The preferences for obstacle collision, and target grasping and placing largely remained the same. For **assisted control** and **shared control** mode, the preferences for the AR features for motion guidance, obstacle collision formation, target grasping/placing, and autonomy indication and intent generally remained the same. This trend continued for **supervisory control** mode as well with minimal changes in the selections for the AR features for obstacle collision information and autonomy indication and intent from the previous selections.

Expert Recommendation-Based (Final Selection) — For **direct control** mode, minimal changes happened across all the AR features between the intermediate selection and the final selection with most of the participants selecting the local information for motion guidance and all the participants wanting detailed information for obstacle collision (planes) and target grasp/place (affordance). For the **assisted control** mode, most of the participants selected local information in the form of arrows for motion guidance with some participants changing their preferences from global information in the form of path AR. Half the participants selected the low salience light notification for autonomy indication with a few participants changing their choices from high salience to low salience when moving from intermediate to final selection. Most participants continued to select the path AR to provide global information about the intent of autonomy. For information about a collision with the obstacle, most participants selected general information in the form of box AR. Finally, for the target hint, most participants chose to have no hint to help with target grasp/place with nearly half the participants changing their intermediate selections. Most of the participants selected global information for motion guidance while making their final selection for **shared control** mode, with some participants changing their selections from local information for motion guidance. Nearly all the participants selected the high salience autonomy indication with nearly half the participants changing their selections from the intermediate selection. Similar to assisted control, the selections for autonomy intent remained largely unchanged, with nearly all the participants preferring to learn about robot autonomy intent through the global information provided by path AR. More than half the participants changed their preference for obstacle information, with most of the participants now preferring no information about the obstacle collision. Most of the participants prefer to highlight the target while grasping or placing with half the participants changing their preference from the previous intermediate selection. For **supervisory control** mode, most of the participants selected the medium salience text indication for autonomy indication with half the participants changing their selection to medium salience. Similar to both assisted and shared control, most participants selected the global information via path AR for autonomy intent with minimal participants changing their preferences. Most of the participants preferred no information about collisions with the obstacle with nearly half the participants changing their preferences from other AR features. Finally, most of the participants now preferred no information regarding the target with most of the participants changing their preferences from the intermediate selection.

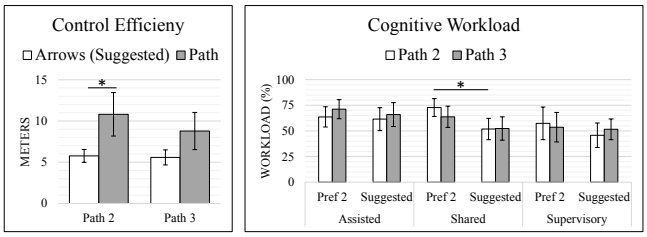

Fig. 4: Control efficiency evaluated by the handheld controller's trajectory length in direct control and cognitive workload.

Discussion — Regarding the influences of how participants learn to use the AR visual cues (**RQ2**), we found that ***participants' preference for AR visual cues converged to the recommendation of experienced users***, which is observed by a large change after hands-on robot control using suggested AR features (Fig. 3.c) and most of the participants selecting the suggested AR features as their final choice across all control modes. Additionally, the participants commented: *"...would like to have clear guidance on how I should move the robot when not much robot autonomy is available."* which was supported by the total trajectory lengths of the handheld controller being significantly shorter ($p < .05$) while using the suggested AR visual cue (arrows) in direct control mode (Fig. 4). The participants also commented: *"...the most obvious way to inform the activation of the autonomy will be preferred for shared control so that I do not need to put too much effort when the autonomy is on."* and this was supported by the significantly lower cognitive workload ($p < .05$) while using the suggested AR visual cues in shared control (Fig. 4). We also found that ***video instruction and hands-on practice tend to provide sufficient information for the selection of AR visual cues in direct control without autonomy*** while experience and proficiency play a role in selecting suitable AR features when various autonomy is available. This is supported by the observation that participants' final preference for AR visual

cues remained consistent with their initial selection when using direct control, but there were notable differences when using assisted, shared, and supervisory control.

VI. Conclusion

In this paper, we presented multiple control modes with varying levels of autonomy for a pick-and-place remote robot manipulation task. We also provided several AR features that participants could select to create their ideal visual interface for each control mode. Our user study investigated how participants select AR cues based on the level of robot autonomy. The participants' priority for AR visual cues shifted from guiding human motion for robot control to communicating autonomy activation and intention as the level of autonomy transitioned from direct to supervisory control. With the increasing levels of autonomy, their preference for AR cues shifted from providing local information to global guidance for robot control. Additionally, AR cues that served the same purpose as the robot autonomy had reduced efficacy and was generally not preferred by the participants. We also identified that the participants' preference for AR visual cues tended to converge to the recommendation of users with hands-on experience using the visual interfaces regardless of their initial selections based on video instructions.

Limitations — Although our user study only focused on a pick-and-place task, we anticipate these findings to have broader implications for AR interface design in remote manipulation, and we plan to investigate their applicability to other manipulation tasks (stacking, inserting, boundary-tracing) in future work. We also plan to expand the scope of the study to include more participants and investigate the impact of augmented reality features on nullifying the effect of varying backgrounds of the operators, including their robot operation or gaming experience to determine if the interfaces can have similar performance between participants regardless of their experience level.

References

[1] M. T. Mason, "Toward robotic manipulation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 1–28, 2018.

[2] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[3] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, J. Kober, X. Li *et al.*, "Challenges and outlook in robotic manipulation of deformable objects," *IEEE Robotics & Automation Magazine*, vol. 29, no. 3, pp. 67–77, 2022.

[4] T.-C. Lin, A. U. Krishnan, and Z. Li, "Shared autonomous interface for reducing physical effort in robot teleoperation via human motion mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9157–9163.

[5] M. Selvaggio, J. Cacace, C. Pacchierotti, F. Ruggiero, and P. R. Giordano, "A shared-control teleoperation architecture for nonprehensile object transportation," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 569–583, 2021.

[6] A. E. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow, "Strategies for human-in-the-loop robotic grasping," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 1–8.

[7] S. Arevalo Arboleda, F. Rücker, T. Dierks, and J. Gerken, "Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, 2021, pp. 1–14.

[8] S. S. White, K. W. Bisland, M. C. Collins, and Z. Li, "Design of a high-level teleoperation interface resilient to the effects of unreliable robot autonomy," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 519–11 524.

[9] D. Schitz, S. Bao, D. Rieth, and H. Aschemann, "Shared autonomy for teleoperated driving: A real-time interactive path planning approach," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 999–1004.

[10] M. Walker, T. Phung, T. Chakraborti, T. Williams, and D. Szafir, "Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy," *arXiv preprint arXiv:2202.11249*, 2022.

[11] C. P. Quintero, O. Ramirez, and M. Jägersand, "Vibi: Assistive vision-based interface for robot manipulation," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4458–4463.

[12] D. Krupke, L. Einig, E. Langbehn, J. Zhang, and F. Steinicke, "Immersive remote grasping: realtime gripper control by a heterogenous robot control system," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016, pp. 337–338.

[13] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 179–186, 2017.

[14] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating robot arm motion intent through mixed reality head-mounted displays," in *Robotics Research: The 18th International Symposium ISRR*. Springer, 2020, pp. 301–316.

[15] T.-C. Lin, A. U. Krishnan, and Z. Li, "Comparison of haptic and augmented reality visual cues for assisting tele-manipulation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9309–9316.

[16] W. P. Chan, C. P. Quintero, M. K. Pan, M. Sakr, H. M. Van der Loos, and E. Croft, "A multimodal system using augmented reality, gestures, and tactile feedback for robot trajectory programming and execution," in *Virtual Reality*. River Publishers, 2022, pp. 142–158.

[17] D. Szafir and D. A. Szafir, "Connecting human-robot interaction and data visualization," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 281–292.

[18] A. U. Krishnan, T.-C. Lin, and Z. Li, "Design interface mapping for efficient free-form tele-manipulation," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 6221–6226.

[19] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[20] M. D. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, pp. 241–248, 2014.

[21] C. Piyavichayanon, M. Koga, E. Hayashi, and S. Chumkamon, "Collision-aware ar telemanipulation using depth mesh," in *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2022, pp. 386–392.

[22] J. Larsson, M. Broxvall, and A. Saffiotti, "An evaluation of local autonomy applied to teleoperated vehicles in underground mines," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 1745–1752.

[23] J. F. Mullen, J. Mosier, S. Chakrabarti, A. Chen, T. White, and D. P. Losey, "Communicating inferred goals with passive augmented reality and active haptic feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8522–8529, 2021.

[24] K. Chintamani, A. Cao, R. D. Ellis, and A. K. Pandya, "Improved telemanipulator navigation during display-control misalignments using augmented reality cues," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 29–39, 2009.

[25] A. D. Souchet, S. Philippe, D. Lourdeaux, and L. Leroy, "Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review," *International Journal of Human–Computer Interaction*, vol. 38, no. 9, pp. 801–824, 2022.