2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France
[Award version is not citable, please cite the contributed session version]

Perception-Action Coupling in Usage of Telepresence Cameras

Alexandra Valiton and Zhi Li[1]

Abstract— **Telepresence tele-action robots enable human workers to reliably perform difficult tasks in remote, cluttered, and human environments. However, the effort to control coordinated manipulation and active perception motions may exhaust and intimidate novice workers. We hypothesize that such cognitive efforts would be effectively reduced if the teleoperators are provided with autonomous camera selection and control aligned with the natural perception-action coupling of the human motor system. Thus, we conducted a user study to investigate the coordination of active perception control and manipulation motions performed with visual feedback from various wearable and standalone cameras in a telepresence scenario. Our study discovered rich information about telepresence camera selection to inform telepresence system configuration and possible teleoperation assistance design for reduced cognitive effort in robot teleoperation.**

I. Introduction

Robot teleoperation enables the control of complex robotic systems to perform difficult tasks in cluttered human environments. Such tasks usually require human expertise for decision-making and/or human intervention for handling uncertainty and recovering from failures and therefore cannot be reliably performed under fully autonomous control [1]. For instance, through direct teleoperation, a healthcare worker can control multiple action components (e.g. arms, grippers, mobile base) of a mobile humanoid nursing robot in coordination with its perception components (e.g., moving cameras attached at robot head, torso, and wrists).

Although complex nursing tasks become feasible via teleoperation, the overwhelming physical and mental effort required to coordinate motion and perception components makes tele-nursing robot use difficult [2]. Among many aspects of motion control, the coordination between perception and action is critical to tele-nursing task performance. For example, a teleoperator may need to switch between cameras to explore and inspect the remote environment from distinct perspectives or adjust a camera attached to one arm to observe the precise manipulation performed by the other arm (see Fig. 1). To reduce the effort required in these complex perception-action coordination tasks, prior research has explored approaches for autonomous camera control to facilitate teleoperated manipulation actions [3], [4]. However, without understanding natural human methods of perception-action coordination, these approaches may select robot actions that are disruptive to the teleoperator's situational awareness and negatively affect their decision-making and high-level task planning [5].

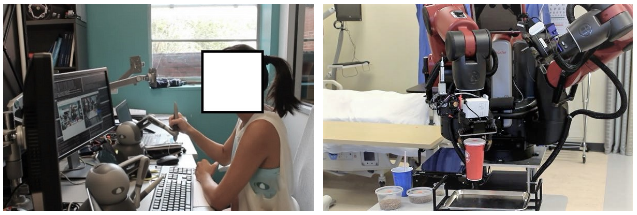

Fig. 1: A human teleoperator controlling a mobile humanoid nursing robot to perform a food serving task. The camera attached to the left robot arm is posed to observe the manipulation of the beverage container.

This work aims to understand human preference in camera selection and control in the teleoperation of a complex robot system with multiple perception and action components. We investigate natural human perception-action coordination in a simulated telepresence setting, where manipulator motion control is not an issue. Specifically, we conduct a user study ($n = 16$) for a cup-stacking task, which involves gross and precise manipulation as well as active perception. This task was performed with visual feedback limited to cameras attached to the head, chest, and hands of a human subject, made available via a two-dimensional head-mounted display. Participants' camera selection and usage was recorded and analyzed along with a NASA-TLX workload assessment and camera preference survey. Our study found that participants discovered effective and interesting ways to adapt the various camera views to the task needs. Participants' camera selection and control strategies fell within two categories with different primary perception control objectives, but standard performance metrics in the single-camera trials were unable to predict participant behavior or their performance in the final trial. Additionally, reported camera preference did not match objectively measured camera selection trends. We present a discussion of preferred camera selection among a diverse student population, the limitations of standard teleoperation performance metrics in perception-action coordination in telepresence, and their implications for customized shared autonomy design in telepresence tele-action human-robot systems.

II. Related Work

The coordination of vision and movement is an essential perception-action coupling skill in human motor control. Experimental studies on eye-hand and eye-foot coordination tasks have revealed temporal and spatial coupling of vision and movements [6], [7]. Particularly in precise manipulation, humans rely heavily on visual feedback for movement corrections [8]. When performing tasks via telepresence and teleoperation interfaces, human teleoperators use active perception to gather environment information to make motion control decisions for manipulation and navigation tasks being

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

performed [9]. Camera selection and control in telepresence and teleoperation scenarios are influenced by human preferences in perception-action coordination. The use of active perception also reflects how a teleoperator visualizes and understands a remote space [10]. Perception assistance for robot teleoperation, in terms of autonomous camera selection and control, should be based upon an understanding of human preference, to better compensate for or augment the spatial skills of human teleoperators.

A. *Active and Interactive Perception*

Perception-action coordination, in terms of active and interactive perception, enables humans to effectively and efficiently explore an environment. Through active perception, humans use visual information to determine what action to take as well as how, where, and when that action should interact with the environment [9]. For example, visual search allows an agent to collect environmental cues that inform action selection [11] and visual feedback refines action execution [8]. On the other hand, interactive perception exploits interaction with the environment to create a rich sensory signal that would otherwise not be available, and utilizes knowledge of the regularity in the combined space of sensory data and action parameters to predict and/or interpret the sensory signal [12]. Thus far, approaches for active and interactive perception has been developed for an autonomous manipulator [13], humanoid robot [14], and telepresence mobile robot [15]. While developing autonomous camera control suitable for a specific robot platform and task scenario, these research efforts have limited understanding of the complex relationship between *camera selection and control* and *robot action and action phases*. Other recent work evaluates immersive teleoperation systems that use virtual reality to engage the operator's natural perception-action coordination [16], [17], although these systems received mixed user evaluations.

B. *Spatial Skills and Teleoperator Performance*

Spatial skills are essential to human performance in robot teleoperation tasks. While experts disagree on a precise taxonomy of spatial skills, common groupings include spatial visualization (perceiving objects among cluttered environments), mental rotation (visualizing an object in a new configuration or orientation), and perspective taking (visualizing an object from distinct observer perspectives) [18]. There is extensive literature on the correlation between spatial skills and teleoperation performance in diverse tasks [19]–[23]. Teleoperation systems that augment operators' spatial skills saw increased performance and reduced fatigue [17], [24]–[27]. Particularly in teleoperation of nursing robots, developing perception assistance to augment spatial skills is essential given the specific user population. The current US nursing population consists of more women than men, with an average age around 49 years old [28]. Research consistently shows that men tend to outperform women in spatial skill tasks [29] and there is some evidence that as individuals age their spatial skills decline [30] The differences in spatial skills may limit female and older nurses' ability and desire to use tele-nursing robots.

C. *Perception Assistance for Robot Teleoperation*

Perception-action coordination in robot teleoperation differs from that in human motion control due to the robot embodiment, robot sensing and motion capabilities, and the design of teleoperation interface. Research cites narrow field of view [31], lack of depth perception [26] and reduced situational awareness [32] as factors that reduce functional presence in a remote workspace [25]. Teleoperators with stronger spatial skills may be able to adapt to the limited visual feedback and successfully perform perception-action coordination tasks [24], [33]. However, extensive usage of spatial skills in unfamiliar environments can cause mental fatigue and reduced task performance [27], [34]. Robot teleoperation assistance thus far primarily focuses on automating tele-action tasks based on intent inference [35]–[38]. Research on perception issues in teleoperation have led to improved design and augmentation of user interfaces [17], [26], [39]–[41]. Recent work in alleviating the mental strain of robot teleoperation has proposed adoption of interface design from telepresence technologies [42], [43]. The assistance available for active perception control in coordination with teleoperated manipulation tasks is limited to the autonomous control of a moving camera [3], [44]. Without considering human preference in camera selection and control, it is unclear whether the autonomously controlled camera view will lead to an inconvenient and uncomfortable teleoperator experience. Particularly, this approach is limited to controlling a single moving camera which may not be suitable for the robot platforms that are equipped with multiple fixed and moving cameras [45]. To fully utilize the physical capabilities of such a platform, it is necessary to develop perception assistance that automates the camera selection and control in coordination with the robot's versatile action functions.

III. Experiment

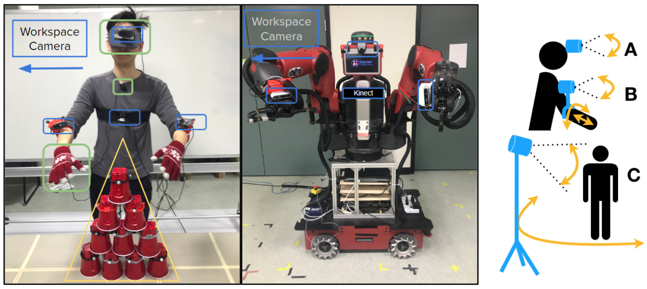

Fig. 2: The cameras in VR telepresence task simulate the cameras equipped on a tele-presence tele-action nursing robot [45]. The cameras can be adjusted before the experiment for a convenient view.

We conducted a user study to examine human perception-action coupling in manipulation tasks performed with visual feedback from telepresence cameras. In direct robot teleoperation, natural perception-action coupling in human motor control cannot be preserved due to the dissimilarity of human and robot embodiments. Additionally, the difficulty of controlling robot motions, even using a motion-capture system to map human motion directly to the robot [45], may discourage the teleoperator's active camera selection and control. A teleoperator may eventually adapt to the teleoperation interface and be able to successfully perform the required tasks. However, their developed teleoperation strategies may involve perception-action coordination that requires strong spatial skills and high mental effort, and could be uncomfortable examples for novice users to follow. As a result, we studied human perception-action coordination in a simulated telepresence setup, where participants wearing a head-mounted display received video feeds from cameras attached to their own body, thereby trivializing the manipulation component of the task to encourage active camera selection and control. Participants naturally reveal effective perception-action coordination strategies as they adapt to the camera configuration and discovers their preferred camera selection and control.

A. Experiment Setup

The participants were instructed to perform a cup-stacking task with the camera views from various wearable and standalone cameras streamed to a VR headset. Shown in Fig. 2, these telepresence cameras were chosen to simulate the perception cameras equipped on a mobile humanoid nursing robot, which can perform manipulation and navigation tasks under direct teleoperation [45]. The *Clavicle Camera* ($C_{clavicle}$) was attached to the chest above the sternum and between the underarms and mimicked the limited degrees of freedom and range of motion of a robot head camera. The *Action Camera* (C_{action}) and the *Perception Camera* ($C_{perception}$) were attached to the dominant hand primarily responsible for manipulation and the non-dominant hand that assists manipulation, respectively. The *Workspace Camera* ($C_{workspace}$) was located across the workspace from the participant on a stationary tripod. The *Head Camera* (C_{head}) was attached to the front of the VR headset, matching natural human eyesight. During the experiment, the participants wore thick gloves to minimize their dependence on precise haptic feedback and a wireless microphone to switch the cameras using voice commands. Before the experiment, participants were allowed to make small adjustments to the camera field of view to their preference. The available camera adjustments are shown in the far right of Fig. 2 and include:

- $\mathbf{C_{head}}$: The angle between the front of the VR headset and the camera lens.
- $\mathbf{C_{action}}$ and $\mathbf{C_{perception}}$: The location on the forearm (between the elbow and the wrist), the rotation of the mounting bracket around the forearm, and the angle between the mounting plate and the camera lens.
- $\mathbf{C_{clavicle}}$: The angle between the sternum mounting strap and the camera lens.
- $\mathbf{C_{workspace}}$: The location of the camera tripod relative to the participant and workspace, the angle between the tripod mount and the camera lens, and the focal length of the camera image. The $C_{workspace}$ image was flipped horizontally based on user feedback during a pilot study.

B. Participants, Tasks and Procedure

Our study recruited healthy participants ($N = 16$, 8 males, 8 females, average age = 23.4 ± 3.6) including student and general populations. Before the experiment, the experimenter equipped the participant with the wearable cameras, VR headset, microphone and gloves, and introduced the task of stacking lightweight plastic cups into a pyramid. The task was considered to be successful if it was completed within three minutes and if the cup stack remain standing for two seconds. We designed the task to be simple to understand and perform. Consequently, skill or experience played little role in successful completion of the task. The stacking task involved three distinct actions: (1) world exploration to observe the environment without interaction, (2) gross manipulation to reach for and carry objects, and (3) fine manipulation of objects with hands. These actions, and combinations thereof, span a wide variety of tasks a tele-manipulation system may need to perform. The cups were easy to grasp and manipulate, yet their low-friction surface and light weight made manipulation errors easy to observe.

Participants were first asked to stack six cups with the telepresence cameras (2 trials $\times$ 5 cameras = 10 trials). For each camera, a participant had a three-minute practice trial to get familiar with the selected camera view. In a following trial, the participant would perform the task as quickly as they felt comfortable. This second trial is used to evaluate the operator's skill and workload using the selected camera (*single-camera trials*). The order of camera selection was randomized for each participant to minimize task learning effects. Camera adjustment was permitted before, during and after the practice trial, but the wearable camera locations and angles with the mounting point remained static during the performance trial.

For the final trial, participants were instructed to stack ten cups and were able to use and switch the camera view at will (*multi-camera trial*). The available cameras did not include the head camera (C_{head}) because in practice, VR telepresence systems may be uncomfortable to use for long periods of time like traditional healthcare worker schedules [5]; we used the C_{head} condition to represent an ideal camera control baseline against which the other cameras can be compared.. The participants were instructed to perform the final trial at a comfortable pace within the three minute time limit. Before the final trial, participants practiced using voice commands to switch cameras such that their task performance flow would be minimally disrupted by camera switching.

C. Survey

After the experiment, the participants completed a workload assessment with relevant NASA-TLX questions (Q1-Q4) for mental task demand, effort, performance, and frustration, and additional questions about functional presence (Q5-Q8) on the Likert scale. Our workload assessment questions include:

- Q1 - Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?
- Q2 - Frustration Level: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during

the task?

- Q3 - Mental demand: How much mental and perceptual activity was required?
- Q4 - Overall Performance: How successful were you in performing the task?
- Q5 - Awareness of Hands: How aware were you of the position of your hands during the task?
- Q6 - Gross Manipulation: How intuitive was moving your hands (up, down, left, right) in this task?
- Q7 - Fine Manipulation: How intuitive was grasping in this task?
- Q8 - Awareness of Cups: How confident were you in the position of the cups in this task?

Since the task was not physically challenging for healthy adults and had an imposed time limit, the NASA-TLX questions regarding physical and temporal demand were omitted. The participants also completed a camera preference survey describing their choice of cameras in exploring, reaching, grasping, and overall. They were asked to provide any specific feedback about usage and preference of camera views and suggest improvements for camera configuration.

D. Data Analysis

The experiment was recorded along with the participant's view from the selected camera. The videos were annotated for the following:

- **Action Phase**: Classifying motion intent as exploration, gross manipulation, or precise manipulation.
- **Hand in View**: Determining whether the participants can see their hands within the camera view.
- **Camera Selection**: Labelling which camera is currently being viewed.
- **Errors**: Marking when an error is committed, including errors in grasping, placement, ineffective manipulation, and direct and indirect collisions.
- **Task Progress**: Marking when the participant successfully grasps a cup or places it in the pyramid formation.
- **Duration**: Marking the beginning and end of the task, as well as noting any interceding time when task progress was suspended by experimenters (e.g., cups fall off the table and must be reset).

Participant performance was evaluated based on the task duration and the number of errors committed [46]. In order to preserve the focus of the experiment on participant preference, the participants were not informed of the evaluation metrics, and were only told to ”complete the task as quickly as [they] feel comfortable”. Participant performance and survey responses were compared across cameras, gender, and engineering experience using ANOVA and Wilcoxon rank sum tests, respectively. The entire experiment was performed in a motion capture volume with motion capture markers located on the VR headset, wrist camera mounts and the cups. Later work will focus on analysis of this motion data to uncover trends in natural human motion for camera control.

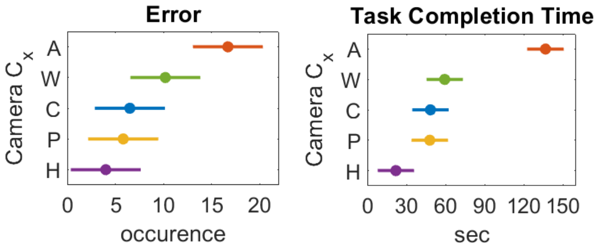

Fig. 3: Comparing task performance among cameras.

IV. Results

A. Single-Camera Trials

Performance metrics were compared across cameras, genders, and engineering education. Demographics (gender and education) had no significant effect on task performance. Fig. 3 shows that trials using the Action Camera (C_{action}) had significantly longer task completion time than the other camera views, and significantly more errors than Head (C_{head}), Clavicle ($C_{clavicle}$) and Perception Cameras ($C_{perception}$). Trials with C_{head}, $C_{clavicle}$, $C_{perception}$, and $C_{workspace}$ showed no significant difference in their effects on task performance among participants. We also considered relative task performance (e.g., errors committed using $C_{perception}$ compared with errors committed using C_{head}, or one participant's task completion time using $C_{clavicle}$ compared to the average task completion time using $C_{clavicle}$) and found no consistent pattern of performance across subjects.

We analyzed the workload assessment results for each of the single-camera trials to understand user preference among the camera views. The ranksum analysis on the responses is shown in Table I, which indicates clear preference rankings among camera views with significant differences ($p < 0.05$). C_{action} was consistently evaluated to be the worst camera view, and C_{head} the best. $C_{clavicle}$, $C_{perception}$, and $C_{workspace}$ received mixed evaluations. Analysis of the workload assessment responses shows that *gender plays a role in camera preference*: male participants reported using less effort ($p < 0.05$) and achieving higher performance ($p < 0.01$) with $C_{clavicle}$ than female participants, and female participants reported using less effort with $C_{workspace}$ than males ($p < 0.05$). No corresponding differences in performance metrics were observed.

B. Multi-Camera Trials

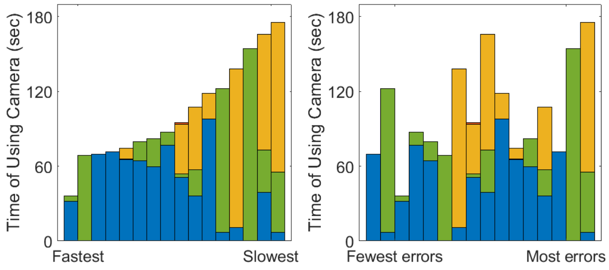

Fig. 4: Camera preference indicated by final trial camera selection.

Q1 (Effort):	$C_a < C_c , C_p, C_w < C_h$
Q2 (Frustration):	$C_a < C_c , C_p < C_h$ $C_a < C_w$
Q3 (Mental Demand):	$C_a < C_c , C_p , C_w < C_h$
Q4 (Performance):	$C_a < C_c, C_w < C_h$ $C_a < C_p$
Q5 (Awareness of Hand):	$C_a < C_c, C_w < C_h$ $C_a < C_p$
Q6 (Motion Intuitiveness):	$C_a, C_c, C_w < C_h$ $C_a < C_p$
Q7 (Grasp Intuitiveness):	$C_a < C_p, C_w < C_h$ $C_c < C_h$
Q8 (Awareness of Cups):	$C_a < C_w < C_p, C_h$ $C_a < C_c$

TABLE I: Comparing camera rankings through workload assessment responses. C_a represents the Action Camera; C_w represents the Workspace Camera; C_c represents the Clavicle Camera; C_p represents the Perception Camera; C_h represents the Head Camera. $<$ represents a statistically significant ranking ($p < 0.05$)

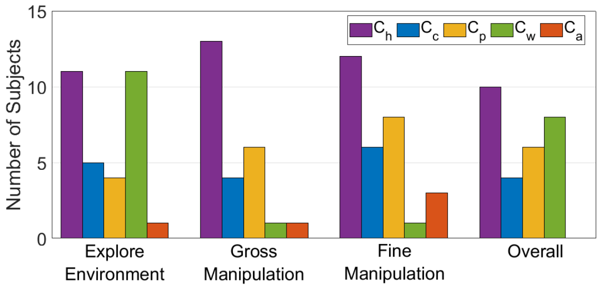

Fig. 5: Camera preference indicated by participant survey

We analyzed the camera preference as indicated by selection during the final trial as well as the camera preference survey. Fig. 4 orders the participants by their task completion time and error occurrences in the final trial, and indicates the corresponding task completion time and proportion of camera usage. Fig. 5 shows the number of participants that prefer each camera for each type of action.

We found *a marked inconsistency between the survey responses and final trial camera choice*. According to survey responses, the preference for C_{head} and dislike of C_{action} is evident. $C_{workspace}$, which has the widest field of view, was selected as often as C_{head} for exploring the environment, and was ranked second in overall preference. $C_{clavicle}$ and $C_{perception}$ were most preferred after C_{head} for fine manipulation. However, according to the final trial camera choice, $C_{clavicle}$ was selected by a majority of participants (14 out of 16) for a majority of the task time (mean: $52 \pm 40\%$ of task completion time). None used C_{action} significantly (threshold: 2% of task completion time). A participant's camera selection did not correlate significantly with their performance in the single camera trials, nor with their workload assessment responses in single camera trials. Roughly half of participants (9 of 16) used one or both of $C_{clavicle}$ and $C_{workspace}$ in the final trials, but not $C_{perception}$, a strategy requiring almost no camera control. The rest of the participant group used $C_{perception}$ in combination with $C_{clavicle}$ and/or $C_{workspace}$.

V. Discussion

A. Telepresence Tele-action Camera Evaluation

In terms of task performance and subjective workload evaluation, our work does not support a significant difference between C_{head} and $C_{clavicle}$, $C_{workspace}$, or $C_{perception}$. This finding implies that *appropriate usage (by human operator or autonomous controller) of $C_{clavicle}$, $C_{workspace}$, and $C_{perception}$ would yield comparable performance* in terms of task completion time and number of errors. While these metrics are standard for robot teleoperation evaluation, it is possible that other performance metrics may reveal the difference in impact on operator's remote perception capability between these cameras, which we will address in a future investigation.

B. Prediction of Camera Selection

Our results do not show a significant correlation between gender, education level, reported workload, or performance in single-camera trials with performance in the final trial in which users can select cameras freely. Particularly, neither the preference of nor relative success with individual cameras predict camera selection in final trial. These inconclusive results suggest that *standard metrics of teleoperation performance may not sufficiently capture operator preference* in remote teleoperation tasks. The results also don't reveal consistent user preference or performance across cameras. These outcomes imply that *customization of autonomous camera selection with respect to user groups, or even personalization, is necessary*.

C. Camera Usage

Objective analysis of final trial camera selection, subjective analysis of experiment video, and in-depth interviews with participants have revealed *multiple distinct objectives in view selection* during the final, multi-camera trial. A participant's weighting of these objectives impacted their camera selection in the final trial, as well as the task completion time.

User Group 1 (9/16 participants) prefers less camera switching and camera motion as long as necessary information (e.g., cup to be grasped, manipulation hand, goal position to place cup, etc.) is present the in camera view. Our data shows that these participants tend to use either a single camera (workspace or clavicle) or use one camera heavily (minimum 72% of final trial duration) with one other camera occasionally. Participants who fall into this group tend to have shorter-than-median task completion time (7/9 participants in this group).

User Group 2 (7/16 participants) prefers switching among multiple cameras and greater use of the perception hand camera (mean: $42\% \pm 28\%$ of final trial duration). We believe that participants in this group either want to ensure the task will be performed successfully, or cannot estimate the spatial relationship appropriately without additional camera perspectives. Participants who fall into this group tend to have longer-than-median task completion time (6 of 7 participants in this group).

The incompatibility of these two group's active perception behavior demands investigation into the contradictory objectives as well as the relative priority of these objectives among a diverse group of users.

D. Camera Selection Skill

Statistical analysis of each participants performance in the final trial showed no significant correlation with their performance in the single-camera trials. This indicates that when multiple cameras are available, the active perception skill not only depends on the capability of using each individual camera, *but also depends on the skill of selecting the camera suitable to the operator capability and task steps*. Development of shared autonomy for active telepresence needs to address camera selection in addition to individual camera control.

VI. Conclusions and Hypotheses for Future Investigation

Through this study we observed several interesting behaviors that could not be explained by our experimental metrics; we will investigate these behaviors in future user studies. We present here the hypotheses we have formed in response to this work that will be addressed in future research.

Hypothesis 1: Objective and subjective methods from cognitive science research can be used to more accurately evaluate telepresence skill and task performance.

Although we were able to compare task performance and workload across participants and across cameras, we do not believe that this work has precisely captured the variation in remote perception performance or spatial skill. For example, two participants that were clearly comfortable and capable had dramatically different camera usage and were ranked first and ninth by task completion time in the final trial. Another subject ranked second by task completion time offered very negative feedback about all camera views, indicating a complete lack of comfort. We believe that spatial skill evaluation methods from cognitive science research will offer better insight to participants' telepresence skill. Similarly, we believe that more specific assessments of cognitive demand will allow us to more clearly rank telepresence task performance with and without shared autonomous control.

Hypothesis 2: Decision support from learned models of camera preference will help operators converge on the most effective strategy sooner.

Our study demonstrates that participants did not always select the cameras most suited to their capabilities. For example, pairs of participants that had similar distributions of camera use in the final trial did not perform similarly (ranked ninth vs sixteenth, first vs fourteenth, sixth vs fourteenth in errors committed). We believe that participants with lower performance did not select the cameras that were most effective for them. A camera suggestion model that can evaluate an operators' spatial skill and suggest a compatible camera selection strategy to the operator would improve the operator's performance faster than letting the operator learn on their own.

Hypothesis 3: Operators with low spatial skill do not benefit from active perception like the previously identified camera user group.

Some participants ranked among the worst performance in all trials (single- and multi-camera trials). These participants appear in both of the user groups identified in section V-C, but their performance is similarly bad. We believe that these participants belong to a third user group which prefers minimal active perception (e.g., one stationary camera). This user group may be more sensitive to other developments for telepresence and teleoperation systems such as user interfaces which integrate multiple sensors or semi-autonomous tele-action.

References

[1] M. A. Goodrich, J. W. Crandall, and E. Barakova, "Teleoperation and Beyond for Assistive Humanoid Robots," in *Reviews of Human Factors and Ergonomics*, 2013, vol. 9, ch. 5, pp. 175–226.

[2] B. DeJong, J. Colgate, and M. Peshkin, "Mental transformations in human-robot interaction," in *Mixed Reality and Human-Robot Interaction*. Springer, 2011, pp. 35–51.

[3] D. Rakita, B. Mutlu, and M. Gleicher, "An Autonomous Dynamic Camera Method for Effective Remote Teleoperation," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, 2018, pp. 325–333.

[4] ——, "Remote Telemanipulation with Adapting Viewpoints in Visually Complex Environments," in *Robotics: Science and Systems*, Freiburg im Breisgau, 2019. [Online]. Available: https://github.com/uwgraphics/relaxed

[5] J. Chen, E. Haas, and M. Barnes, "Human performance issues and user interface design for teleoperated robots," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1231–1245, 2007.

[6] W. F. Helsen, D. Elliott, J. L. Starkes, and K. L. Ricker, "Temporal and spatial coupling of point of gaze and hand movements in aiming," *Journal of motor behavior*, vol. 30, no. 3, pp. 249–259, 1998.

[7] ——, "Coupling of eye, finger, elbow, and shoulder movements during manual aiming," *Journal of motor behavior*, vol. 32, no. 3, pp. 241–248, 2000.

[8] D. Elliott, W. F. Helsen, and R. Chua, "A century later: Woodworth's (1899) two-component model of goal-directed aiming." *Psychological bulletin*, vol. 127, no. 3, p. 342, 2001.

[9] R. Bajcsy, Y. Aloimonos, and J. K. Tsotsos, "Revisiting active perception," *Autonomous Robots*, vol. 42, no. 2, pp. 177–196, feb 2018.

[10] A. B. Yu and J. M. Zacks, "Transformations and representations supporting spatial perspective taking," *Spatial Cognition & Computation*, vol. 17, no. 4, pp. 304–337, 2017.

[11] M. Zehetleitner, M. Hegenloh, and H. J. Müller, "Visually guided pointing movements are driven by the salience map," *Journal of Vision*, vol. 11, no. 1, pp. 24–24, 2011.

[12] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive Perception: Leveraging Action in Perception and Perception in Action," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1273–1291, dec 2017.

[13] C. Eppner, R. Martin-Martin, and O. Brock, "Physics-Based Selection of Informative Actions for Interactive Perception," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, Australia: IEEE, 2018, pp. 7427–7432.

[14] Z. Peng, T. Genewein, F. Leibfried, and D. A. Braun, "An information-theoretic on-line update principle for perception-action coupling," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2017-Septe, pp. 789–796, 2017.

[15] D. J. Rea and J. E. Young, "It's All in Your Head: Using Priming to Shape an Operator's Perceptions and Behavior during Teleoperation," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, 2018.

[16] I. Aaltonen, S. Aromaa, K. Helin, and A. Muhammad, “Multimodality evaluation metrics for human-robot interaction needed: A case study in immersive telerobotics,” in *Advances in Human Factors in Robots and Unmanned Systems*, J. Chen, Ed. Cham: Springer International Publishing, 2018, pp. 335–347.

[17] H. Hedayati, M. Walker, and D. Szafir, “Improving Collocated Robot Teleoperation with Augmented Reality,” in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI ’18*, 2018, pp. 78–86.

[18] D. H. Uttal, N. G. Meadow, E. Tipton, L. L. Hand, A. R. Alden, C. Warren, and N. S. Newcombe, “The malleability of spatial skills: A meta-analysis of training studies,” *Psychological Bulletin*, vol. 139, no. 2, pp. 352–402, 2013.

[19] M. A. Menchaca Brandan, “Influence of spatial orientation and spatial visualization abilities on space teleoperation performance,” Ph.D. dissertation, Massachusetts Institute of Technology, 2007.

[20] C. Wang, Y. Tian, S. Chen, Z. Tian, T. Jiang, and F. Du, “Predicting performance in manually controlled rendezvous and docking through spatial abilities,” *Advances in Space Research*, vol. 53, pp. 362–369, 2014.

[21] M. Kozhevnikov, M. A. Motes, B. Rasch, and O. Blajenkova, “Perspective-Taking vs. Mental Rotation Transformations and How They Predict Spatial Navigation Performance,” *Applied Cognitive Psychology*, vol. 20, pp. 397–417, 2006.

[22] C. E. Lathan and M. Tracey, “The Effects of Operator Spatial Perception and Sensory Feedback on Human-Robot Teleoperation Performance,” *Presence*, no. 4, pp. 368–377, 2002.

[23] J. Y. C. Chen and M. J. Barnes, “HumanAgent Teaming for Multirobot Control: A Review of Human Factors Issues,” *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 1, pp. 13–29, 2014.

[24] J. L. Wright, J. Y. C. Chen, and M. J. Barnes, “Human-automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance,” *Ergonomics*, vol. 61, no. 8, pp. 1033–1045, 2018.

[25] M. Voshell, D. D. Woods, and F. Phillips, “Overcoming the keyhole in human-robot coordination: simulation and evaluation,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49. Sage Publications Sage CA: Los Angeles, CA, 2005, pp. 442–446.

[26] M. Mast, Z. Materna, M. Španěl, F. Weisshardt, G. Arbeiter, M. Burmester, P. Smrž, and B. Graf, “Semi-autonomous domestic service robots: Evaluation of a user interface for remote manipulation and navigation with focus on effects of stereoscopic display,” *International Journal of Social Robotics*, vol. 7, no. 2, pp. 183–202, 2015.

[27] B. DeJong, J. Colgate, and M. Peshkin, “Improving teleoperation: reducing mental rotations and translations,” in *IEEE International Conference on Robotics and Automation*, New Orleans, 2004, pp. 3708–3714.

[28] NCSBN, “The 2015 National Nursing Workforce Survey,” NCSBN, Tech. Rep., 2016. [Online]. Available: www.journalofnursingregulation.com

[29] M. C. Linn and A. C. Petersen, “Emergence and characterization of sex differences in spatial ability: A meta-analysis,” *Child development*, pp. 1479–1498, 1985.

[35] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. Andrew Bagnell, “Shared autonomy via hindsight optimization for teleoperation and teaming,” *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[30] S. D. Moffat, “Aging and spatial navigation: what do we know and where do we go?” *Neuropsychology review*, vol. 19, no. 4, p. 478, 2009.

[31] S. Johnson, I. Rae, B. Mutlu, and L. Takayama, “Can You See Me Now? How Field of View Affects Collaboration in Robotic Telepresence,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, 2015, pp. 2397–2406.

[32] J. Andersh, N. Papanikolopoulos, and B. Mettler, “Modeling the Human Visuo-motor System for Remote-control Operation,” *ProQuest Dissertations and Theses*, p. 173, 2018.

[33] H. Fabroyir and W.-C. Teng, “Navigation in virtual environments using head-mounted displays: Allocentric vs. egocentric behaviors,” *Computers in Human Behavior*, vol. 80, pp. 331–343, mar 2018.

[34] G. Doisy, A. Ronen, and Y. Edan, “Comparison of three different techniques for camera and motion control of a teleoperated robot,” *Applied Ergonomics*, vol. 58, pp. 527–534, 2017.

[36] J. Butepage, H. Kjellstrom, and D. Kragic, “Anticipating Many Futures: Online Human Motion Prediction and Generation for Human-Robot Interaction,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, Australia: IEEE, may 2018, pp. 4563–4570.

[37] A. Erdogan and B. D. Argall, “Prediction of user preference over shared-control paradigms for a robotic wheelchair,” *IEEE International Conference on Rehabilitation Robotics*, pp. 1106–1111, 2017.

[38] H. Admoni and S. Srinivasa, “Predicting user intent through eye gaze for shared autonomy,” Tech. Rep., 2016.

[39] R. Parasuraman, S. Caccamo, F. Båberg, P. Ögren, and M. Neerincx, “A New UGV Teleoperation Interface for Improved Awareness of Network Connectivity and Physical Surroundings,” *The Journal of Human-Robot Interaction (JHRI)*, oct 2017.

[40] S. H. Seo, D. J. Rea, J. Wiebe, and J. E. Young, “Monocle : Interactive Detail-in-Context Using Two Pan- and-Tilt Cameras to Improve Teleoperation Effectiveness,” in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Lisbon, Portugal, 2017, pp. 962–967.

[41] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, “Ecological interfaces for improving mobile robot teleoperation,” *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 927–941, 2007.

[42] I. Rae, B. Mutlu, and L. Takayama, “Bodies in Motion: Mobility, Presence, and Task Awareness in Telepresence,” in *CHI*, Toronto, 2014, pp. 2153–2162.

[43] R. Yanu Tara and W.-C. Teng, “Improving the visual momentum of tethered viewpoint displays using spatial cue augmentation,” *Intel Serv Robotics*, vol. 10, pp. 313–322, 2017.

[44] D. Nicolis, M. Palumbo, A. M. Zanchettin, and P. Rocco, “Occlusion-free Visual Servoing for the Shared Autonomy Teleoperation of Dual-Arm Robots,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1–1, 2018.

[45] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. Hauser, “Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas,” in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3581–3586.

[46] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, “Common Metrics for Human-Robot Interaction,” in *HRI*, Salt Lake City, 2006, pp. 33–40.