Modeling human perception-action coordination in VR telepresence

Alexandra Valiton, Heramb Nemlekar, and Zhi Li[1]

I. INTRODUCTION

Robot teleoperation enables the control of complex robotic systems to reliably perform tasks in unstructured environments. 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). 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. For a complex telepresence tele-action robot, perception assistance that automates the active perception actions (e.g., camera selection and control) and manipulation actions can significantly reduce the operator's cognitive workload. However, the perception assistance may not be effective and intuitive to its human users unless it complies with human behavior and preference in perception action coordination. In this paper, we propose to model the human behavior and preference we have observed in our prior study on the perception-action coordination VR telepresence. This model will enable us to plan the appropriate perception (manipulation) actions given the manipulation (perception) controlled by a human teleoperator.

II. RELATED WORK

Perception assistance for robot teleoperation - 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. The coordination of vision and movement is an essential perception-action coupling skill in human motor control. When performing tasks via telepresence and teleoperation interfaces, human teleoperators use active perception to determine what action to take as well as how, where, and when that action should interact with the environment [1]. Previous research efforts have demonstrated a limited understanding of the complex relationship between *camera selection and control* and *robot action choices and parameters*. Without considering human preference in camera selection and control, it is unclear whether the shared or full autonomous perception-action coordination for a telepresence tele-action robot will adversely affects the performance, situational awareness and comfort of human operators.

Natural perception assistance - Developing intuitive models for shared autonomy relies upon (1) user studies that collect data on human behavior and preference of perception-action coordination in telepresence tele-action tasks, and (2) learning and planning the coordinated perception and manipulation actions and motions natural to teleoperators. Our prior work has studied the camera selection and control for environment exploration, gross and fine manipulation, given the visual display from multiple wearable cameras attached to human head, body and hands [2]. Despite the consistent behavior and preference across participants, we also noticed differences in camera preference between genders, and difficulty in learning active perception control for the participants with worse spatial skills. On the other hand, available modeling approaches for human-robot collaborations (such as sMDP, POMDP, MOMDP, Dynamic Bayesien Network, see a review [3]) may be adapted to predict and plan the action sequences that consist of teleoperated perception action and autonomous manipulation actions (or vice versa).

III. PRELIMINARY RESULTS FROM USER STUDY

Fig. 1: The cameras in VR telepresence task simulate the cameras equipped on a tele-presence tele-action nursing robot [4]. Participants typically executed exploring, gross manipulation (reaching, carrying), and fine manipulation (grasp, place) actions to complete the cup-stacking task.

Our user study examined human perception-action coupling in manipulation tasks performed with visual feedback from wearable telepresence cameras. 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 (see Fig. 1). Data was collected from participant surveys evaluating each camera and describing camera preference, participant performance evaluated based on task duration and the number of errors committed, and motion capture data for future analysis of natural human motion.

Observation of participant behavior revealed the set of perception actions and manipulation actions performed under each camera view. The participants tend to be more effective when control the perception actions with a more intuitive camera views (e.g., Head Camera), while with less intuitive

---

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

camera views (e.g.,Action Hand Camera) the participants struggled more to adapt to the uncomfortable perception-action coupling. Among trials using the same camera, we observed some consistent behaviors across subjects to overcome the unfamiliar perception-action coupling, including unnatural locking of manipulation arm joints to limit motion complexity, maintaining a static relationship between the head and active camera to prevent confusion during camera motion, and reducing the dimensionality of control by using haptic feedback from the environment to restrict and inform motion in the absence of sufficient visual information from the environment.

Performance revealed a significant disadvantage when using the Action Hand Camera, which was mirrored in participants' evaluation and stated camera preferences. While the Head Camera was highly evaluated and most frequently preferred, performance was not significantly better than with the Clavicle Camera, Workspace Camera, or Perception Hand Camera. Survey results show that men found the Clavicle Camera significantly easier to use than women, while women evaluated the Workspace Camera significantly easier than men. We conclude that an effective shared autonomy system will adapt to the demonstrated perception-action coupling preference of the human teleoperator.

IV. PROPOSED MODEL

Autonomy in Perception and Manipulation - Preliminary analysis of experiment video shows a clear delineation between perception actions (A_p), which primarily gather information from the environment, and manipulation actions (A_m), which directly contribute to task completion. This distinction lends itself to shared autonomous control of a robotic system; such a system could allow the human to focus on manipulation (perception) actions and allow the robotic system to select and execute perception (manipulation) actions. We propose a model adapted from a Markov Decision Process (MDP) in which the perception actions are selected by the robot and manipulation actions are selected by a human, so that the human can effectively make decisions about completing tasks in the remote workspace without experiencing motion sickness or heightened mental fatigue over time.

Perception and Manipulation Actions - In addition to reasonable manipulation actions (A_m) for a cup stacking task, we observed a number of distinct perception actions(A_p), including Exploring the workspace (A_{p-E}) and observing Gross Manipulation (A_{p-G}) and Fine Manipulation (A_{p-F}). These behaviors reveal important insights into perception-action coupling adopted by human operators in a novel context. These actions are summarized in Table I, and form the action space of the MDP.

of the low-level world states to them; (2) use a high-level symbol learning method [5] to learn the symbols that

State Space in Cup-stacking Task - We consider two methods to generate the state space for MDP planning: (1) use human expertise to define the high-level states that satisfy the pre-conditions and post-conditions of the perception and manipulation actions, and associate a probabilistic distribution sufficiently represent the initiation and effect sets given the sets of perception and manipulation actions.

Actions	Description
Perception Actions A_p	
A_{p-E}	Explore environment to gather information
A_{p-G}	Observe gross manipulation actions
A_{p-F}	Observe fine manipulation actions
Manipulation Actions A_m	
A_{m-R}	Reach through workspace toward target (gross)
A_{m-G}	Grasp target (fine)
A_{m-C}	Carry object toward target (gross)
A_{m-P}	Place object at/on target (fine)
Additional Actions	
A_{c-X}	Switch to camera X

TABLE I: Actions observed in the human study. Some perception actions differ by active camera, like A_{p-E}. Manipulation actions exhibiting variance across cameras should converge with natural perception control.

Pairing Perception and Manipulation Actions in State Transitions - In order to represent the interaction between perception and manipulation actions, this model will execute them alternately such that in each time step, a perception action is selected to update the belief state and elicit the highest reward based on expected human selection of manipulation action. This model is shown in Fig. 2.

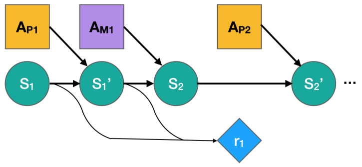

Fig. 2: Adaptation of the MDP model to include two actions (A_p and A_m) at each step.

Paired Action Rewards - In order to support collaborative efforts between human and robot controllers, the reward values in the paired-action MDP should be tuned for each possible combination of A_p and A_m at each step. The rewards should prompt the robot to select actions that support selection of manipulation actions to complete the task, and should update its reward matrix model based on the human operator's demonstrated response to increase the likelihood of intuitive, natural perception actions in the future.

REFERENCES

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

[2] A. Valiton, O. Akyildiz, and Z. Li, "Perception-Action Coordination in VR Telepresence," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Submitted.

[3] L. M. Hiatt, C. Narber, E. Bekele, S. S. Khemlani, and J. G. Trafton, "Human modeling for human–robot collaboration," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 580–596, 2017.

[4] 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.

[5] G. Konidaris, L. P. Kaelbling, and T. Lozano-Perez, "From skills to symbols: Learning symbolic representations for abstract high-level planning," *Journal of Artificial Intelligence Research*, vol. 61, pp. 215–289, 2018.