

Learning Motion Primitives and Task Plan in Teleoperated Robot Motion through Multi-modal Interfaces

Sihui Li, Haowei Zhao, Ozan Akyildiz, Zhi Li

I. INTRODUCTION

Tele-nursing robots extend medical workers' physical capabilities to perform patient-caring tasks in remote and/or quarantine environments. Recently developed tele-nursing robots [15] are equipped with multiple manipulator arms, hands and mobile bases. However, performing complex and dexterous motion coordination is still difficult even if the robots are under direct teleoperation of expert users. It is also hard for the teleoperator to perceive the remote environment and tasks through camera views, which affects the users' situation awareness, motion fluency, and motion control accuracy. Previous research has compared teleoperation interface by their hardware capability and system configuration (e.g., field of view, camera viewpoints, depth perception, video frame rate, bandwidth limitations), time delay and control stability, sensory feedback channels (e.g., visual, audio, tactile display), motion control interfaces (e.g., voice, gesture, motion mapping control) [6, 22, 2], and augmented reality [8, 9, 19, 3]. However, research so far hasn't systematically compared teleoperation interfaces by regularity and variability of the motion primitives and complexity of task plan in the teleoperated motion coordination.

To study the motion coordination problem, we propose to learn from the teleoperated motion coordination demonstrated by expert users, to extract the robot's low-level motion primitives and high-level task plan for nursing tasks that involves arm-hand coordination, bimanual coordination, and loco-manipulation. We further compare the teleoperated motion coordination controlled through different user interfaces, and evaluate the human performance and interface usability in teleoperation. Our proposed research contributes to the understanding of how human motor control adapts to the motion and perception capabilities of remote robotic surrogates. It also informs the design of human-robot teleoperation interface that can facilitate the learning of motion coordination for novice users.

II. RELATED WORK

Tele-robotic systems synergize human and robot capabilities through teleoperation interfaces. To an expert user, the task performance primarily depends on the usability of teleoperation interface, given the teleoperation task and robot capability. Our preliminary evaluation of a mobile humanoid nursing robot (Tele-Robotic Intelligent Nursing Assistant (TRINA) [15]) has demonstrated difficulties in the direct teleoperation of arm-hand, bi-manual and loco-manipulation motion coordination tasks. The teleoperation interfaces constrain the set and quality of motion primitives a robot can employ, which further affect a teleoperator's task

planning. **Measurement metrics for teleoperation interfaces** proposed in previous research focused on evaluating fixed scalar values in the task completion process to compare teleoperation interfaces, including the task completion time, the number of collisions, travel distances, the number of motion primitives [9], smoothness of the trajectory [3], time delay [8], user questionnaire scores [2]. These aspects are useful in comparing performances of teleoperation interfaces. However, a simple combination of these metrics cannot form a framework that can systematically evaluate the usability of a teleoperation interface, by how well it can map the *low-level motion primitives* and *high-level task plan* from human to the teleoperated robot.

To fill this gap, we propose to extract motion primitives from the teleoperated robot coordination and use a symbolic representation of motion primitives to construct the task plan. The motion primitive models we consider include Dynamic Movement Primitive (DMP) [11], Gaussian Mixture Model(GMM) and Gaussian Mixture Regression(GMR) method [4, 5] and Probabilistic movement primitives (ProMP) [16, 18]. DMP learns from single motion trajectory with temporal and spatial scaling ability for reproduction. GMM/GMR can learn from multiple motion trajectories to extract both mean and variances information. ProMP is suitable for trajectory learning and reproduction in changing environments. To model the high-level task plan, we consider the problem in the state-action framework. Methods include constructing skill tree [14], Bayesian model based methods [21], Finite State Automaton (FSA) [17], semi-Markov decision process [13], and inverse reinforcement learning [1, 20].

Another aspect of data analysis is to evaluate human performance. Gawron [10] proposes 12 considerations before conducting an experiments and points out the importance of selecting proper human performance measurement. Kaber and Onal [12] evaluate the performance of the human operator in teleoperation tasks from following aspects: mean time for task completion, the number of motion errors and collisions, situation awareness and NASA-TLX workload score. All of these considerations will be reflected in our data collection and post-study questionnaire.

III. METHODOLOGY

Our user study aims to investigate how expert teleoperators control the motion coordination of a mobile humanoid robot through different teleoperation interfaces. We focus on the set of motion primitives developed through the usage of various robot control interfaces, and the task plan frequently used in arm-hand, bimanual and loco-manipulation coordination.

A. Platform

Our experiments and user study is based on the nursing robot platform TRINA [15]. On the robot side, the hardware includes a humanoid robot torso (Rethink Robotics Baxter), an omnidirectional mobile base (HStar AMP-I) and two three-fingered grippers (Righthand Robotics Reflex grippers). Shown in Fig. 1, our preliminary work has set up six input interfaces to control the motion of a humanoid robot and its mobile base, including (1) keyboard and mouse, (2) gamepad, (3) Geomagic touch haptic devices, (4) RoboPuppet [7], and motion mapping from (5) Vicon motion capture system and (6) Kinect v2 camera. The platform also enables a teleoperator to use voice control to switch camera views, from the RGB cameras attached to the left and right hands, mounted on robot head, and from a standalone camera that looks at the robot and workspace.

B. Experiment

Our user study will train 15 subjects to be fluent in teleoperating a mobile humanoid robot and perform the following motion coordination tasks designed based on our previous nursing task collection [15]:

- pick and place task: several objects will be placed randomly on a table. The objects have different shapes and weights. The goal is to pick up each of them and place them on a tray (will include locomotion).
- bi-manual task: 1) pick up and hold a tray using two hands, then place the tray to a different table nearby (requires locomotion). 2) pick and insert a straw into a cup, perform brushing & rotation motion.
- physical human-robot interaction: handing over different objects to a human in the same robot work-space.
- loco manipulation task: hold the handle of a cart and move the cart to a different location.
- precise perception: this is the category of tasks that require relatively precise perception, including pressing a button, scan a barcode, open cabinet door.
- handling irregular items: Lift large, heavy and/or irregular items with both arms. The object might be transformable, such as a pillow or a bag of rice.

The subjects will start with a training session. If the subject can successfully perform a task consecutively for three times with stable task completion time (variance within 10% of mean), the subject is considered an expert user for the specific task and the input device. After the training session, each subject will have ten trials for each control mode and each task. The order of the control mode is randomly selected. There will also be a post-study questionnaire for the subjective views during the study.

We will record synchronized teleoperated robot motion, control interface input and interface visual display. We segment the collected data to extract motion primitives from the motion robot hand, arm, and mobile bases, and coupled them if the motions are performed simultaneously. In addition, we also extract information for motion-perception coordination. We associate the camera view being used with the motion being performed. We also record task performance including

task completion time, collisions with unwanted objects, and success rate over repetitive trials.

C. Modeling motion primitives and task plan

After robot states data is collected for the entire motion sequence, the first step is to segment the data into compact and repeatable pieces. The segments of data would be modeled with motion primitive learning methods. We will use different modeling methods to learn motion primitives to extract features that can address different aspects of the low-level motor skills. For instance, DMP can be used to learn from an individual demonstration of reaching motion and normalize them across their traversed time and distances. We further use GMM/GMR to extract the regularity and variability of the normalized motion primitives and use ProMP to model the reaching motion performed with respect to moving landmarks in the environment. For task plan modeling, we will also use multiple modeling methods. Constructing skill tree can be applied to compare task plan complexity using skill tree levels. Constructing skill tree can be combined with the motion primitive modeling results directly because they are all trajectory based. Bayesian model based methods can provide a measure of confidence in their model regression results because of their probability nature. FSA constructs original plan structures which can be modified with according to different task situations. For example, based on a grasping plan of an object very close to the robot end-effector, more nodes for locomotion can be added to the task plan to model grasping of an object further away. The methods developed based on semi-Markov decision process require the set of motion primitives to be pre-specified. Motion primitives in this context can be series of learned trajectories, which have semantic meaning by itself, for example, "turn on light", "open door". Skill tree modeling can serve as primitives here.

IV. FUTURE WORK

Our motion analysis based on a novel framework of modeling teleoperated motion coordination will contribute to the understanding of how human motor control adapts to the capabilities of a physical robotic embodiment. By evaluating various teleoperation interfaces, we aim to inform the human-robot interface designs to facilitate the natural mapping of low-level motor skills and task plan. We also aim to identify the low-level tasks difficult for direct teleoperation but easy for autonomous control. For instance, we can autonomously adjust the mobile base to facilitate the manipulation tasks being performed. From observing teleoperators' choice of camera perception, we can also learn when and how to adjust the perception camera to facilitate the action arm motion. Under the framework of shared autonomous control, these tasks can be automated using geometric motion planning methods and/or through learning from demonstration. A robot control interface with intelligent motion coordination assistance will reduce medical workers' physical and cognitive efforts to control a nursing robot so that they can focus on the decision-making and patient interaction.

