

An Open Tele-Impedance Framework to Generate Data for Contact-Rich Tasks in Robotic Manipulation

Alberto Giammarino *Member, IEEE*, Juan M. Gandarias, *Member, IEEE* and Arash Ajoudani, *Member, IEEE*

Abstract—Using large datasets in machine learning has led to outstanding results, in some cases outperforming humans in tasks that were believed impossible for machines. However, achieving human-level performance when dealing with physically interactive tasks, e.g., in contact-rich robotic manipulation, is still a big challenge. It is well known that regulating the Cartesian impedance for such operations is of utmost importance for their successful execution. Approaches like Reinforcement Learning (RL) can be a promising paradigm for solving such problems. More precisely, approaches that use task-agnostic expert demonstrations to bootstrap learning when solving new tasks have a huge potential since they can exploit large datasets. However, existing data collection systems are expensive, complex, or do not allow for impedance regulation. This work represents a first step towards a data collection framework suitable for collecting large datasets of impedance-based expert demonstrations compatible with the RL problem formulation, where a novel action space, namely Variable Impedance Control in End-effector Space (VICES), is used. The framework is designed according to requirements acquired after an extensive analysis of available data collection frameworks for robotics manipulation. The result is a low-cost and open-access tele-impedance framework which makes human experts capable of demonstrating contact-rich tasks.

I. INTRODUCTION

Performing contact-rich manipulation tasks is still a big challenge in robotics. Solving them would unlock the use of robots for applications like logistics and assembly, where their use can have a strong impact. Different aspects make contact-rich manipulation particularly hard to solve. The dynamics of the system around contacts are difficult to model, and small modeling errors can induce task failure. Moreover, perception of the scene is hard due to the frequent co-existence of multiple different objects along with the manipulating hand in a small region.

Imitation Learning (IL) [1] is a viable approach for such problems, and one widely used IL approach is behavioral cloning [2]. Although it has been successfully applied, it suffers from distributional shift [3], meaning that it struggles outside the manifold of the demonstration data. An alternative IL approach is Inverse Reinforcement Learning (IRL) [4]. Despite the appealing idea, the framework makes the limiting assumption of having expert task-specific demonstrations and the joint optimization of policy and reward renders the training process difficult to stabilize.

The authors are with the HRI² Lab, Istituto Italiano di Tecnologia, Genoa, Italy. {alberto.giammarino, juan.gandarias, arash.ajoudani}@iit.it

This work was supported by the ERC-StG Ergo-Lean (Grant Agreement No.850932).

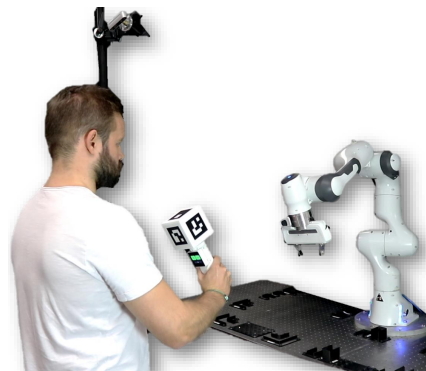


Fig. 1. A human expert teleoperating a robotic manipulator thanks to the proposed open tele-impedance framework¹. A video of the experimental session and a demonstration of the framework using a real Franka Emika Panda robotic manipulator is included².

Another attractive framework is Reinforcement Learning (RL), where the agent learns to accomplish a task by trial and error, maximizing an expected future cumulative reward. In the context of robotic manipulation, a natural reward choice is sparse and binary (i.e., +1 when the task succeeds, 0 otherwise). Before the agent is able to receive useful feedback, it has to explore the state-action space until task completion, which makes RL algorithms extremely sample inefficient for this application. One possible solution to this problem is reward shaping. However, designing a reward is often counterintuitive, requires handcrafting, might lead to unexpected undesired behaviors, and is prone to local optima [5].

An interesting alternative to reward shaping is the integration of task demonstrations into RL algorithms [6]–[11] (Fig. 1). The additional knowledge encoded into the demonstrations is used to bootstrap learning, and speed it up. A class of these approaches leverages task-agnostic demonstrations to learn a new task [8], [10]. In this case, demonstrations are used as generic prior experience to learn aspects of environment and task. We deem that these methods are among the most promising ones since they can make use of large datasets, which have resulted in remarkable results in other machine learning fields, e.g. computer vision and natural language processing. Moreover, these methods resemble how humans learn new tasks. Prior experience allows them to narrow down the search space and as a consequence, they are able to master a new task after a few trials. In this paper, we refer to these approaches as task-

¹<https://gitlab.iit.it/hrii-public/teleimpedance-m5core2>

²<https://youtu.be/60X74kCUTes>

agnostic demonstration-based RL.

Humans are able to master contact-rich manipulation tasks, thus it is natural to take inspiration from them when formulating the RL problem. It is well-known that humans regulate the impedance of their arms according to the requirement while performing contact-rich tasks [12]. In [13], the Variable Impedance Control in End-effector Space (VICES) is introduced as a novel action space for RL that consists of Cartesian impedance and end-effector pose. In the paper, it is shown that this action space formulation improves sample efficiency of model-free RL algorithms with respect to common formulations. Although that, none of the task-agnostic demonstration-based RL approaches uses this action space formulation. These approaches are presented in detail in the next section. This might be due to the lack of a suitable data collection system.

In this paper, we design an open system to ease and enable the collection of motion and impedance expert trajectories for contact-rich manipulation with Cartesian impedance-controlled robots. More specifically, we target tasks that require impedance adaptation to be successfully fulfilled. We gather insights from related works on data collection systems and task-agnostic demonstration-based RL approaches to understand what the requirements for such a system should be. Finally, we design and test the devised framework.

II. RELATED WORK

A. Task-agnostic Demonstration-based RL

The approaches discussed here integrate a dataset of task-agnostic interactions (i.e., state-action sequences) into RL to learn a new contact-rich task. In [6] the authors use generative modeling and representation learning to learn a model that maps a random (latent) variable to meaningful action distributions conditioned on the current state. Demonstrations come from a mix of different successful tasks, where meaningful behaviors are executed. Learning a policy in the space of the latent variable becomes easier because random actions are more likely to correspond to user behaviors. A similar approach is adopted in [7], where exploration is guided using a learned prior distribution over skills (i.e., action sequences) conditioned on the current state. The same authors in [8] improve their previous framework by adding a few task-specific demonstrations in the dataset. In [9] demonstrations of semantically meaningful behaviors coupled with data-relabeling are used to initialize goal-conditioned hierarchical policies. During RL for learning a new task, the loss is augmented to keep the policies close to the behavior encoded in the dataset.

Some works integrate demonstrations in the offline (or batch) RL framework [14]. The authors of [10] enhance the generalization capability of an offline RL algorithm, i.e., conservative Q-Learning (CQL) [15], using a dataset of prior experience unrelated to the new task. In [11] a behavior prior is learned from an imperfect dataset of multiple tasks and is used to avoid that the policy chooses actions not supported by the data.

B. Data Collection Systems

Machine-generated demonstrations are a tempting idea to generate datasets. In [11] an agent is trained with Maximum a posteriori Policy Optimization (MPO) [16] and roll-outs of the final policy are collected to build a dataset. Another way to collect machine-generated demonstrations is through the use of scripted policies, as it is done in [6] and [10]. However, humans are more suitable for collecting a wide set of semantically meaningful and natural behaviors, which is harder to guarantee for machine-generated data.

The works in [8] and [9] use MuJoCo HAPTIX [17] as data collection system. The authors of [17] propose to augment the MuJoCo simulator [18] with a system that enables teleoperation using real-time motion capture (MoCap) of arm and hand movements, coupled with a stereoscopic monitor for visual feedback. This is combined with head-tracking to show the virtual environment always from the human viewpoint. This system is employed also in [7] and [19], where the former uses a dataset created in [20] while the latter exploits behavioral cloning to augment the dataset collected in [21] with an updated version of MuJoCo HAPTIX.

The authors of [22] aim at building an inexpensive teleoperation framework that allows intuitive robotic manipulation and collection of high-quality demonstrations. They come up with a system that uses a VR headset to let the user perceive the environment through the robot’s sensor space and control it with motion-tracked VR controllers. A similar system is used also in [23], where the demonstrations are collected in a virtual environment, and the demonstrator sees a rendering of the same observations of the agent and records actions through an HRC Vive interface. In contrast, in order to use kinesthetic teaching while avoiding human presence in the robot scene, in [24] a robot kinesthetically force controlled by a human is used to teleoperate the same robot type in simulation or in the real world.

In [25] RoboTurk is developed, a crowdsourcing platform for 6 degrees-of-freedom (DoFs) trajectory-based teleoperation. It allows the users to provide task demonstrations in a virtual environment through their own mobile devices. The system has the advantages of using a ubiquitous interface, i.e., a mobile device, while being intuitive and sufficiently precise.

While the works presented above are able to provide movement demonstrations, they lack of the ability to produce impedance signals, which is one of the main goals of our work. In addition, some of them use expensive devices which might not be affordable by some research lab (e.g. VR headsets), another important requirement that we consider here. In general, none of the systems has been designed on purpose for demonstrations to be integrated in the RL frameworks considered in this work. In section III-A we further clarify these and other requirements for the design of our data collection system.

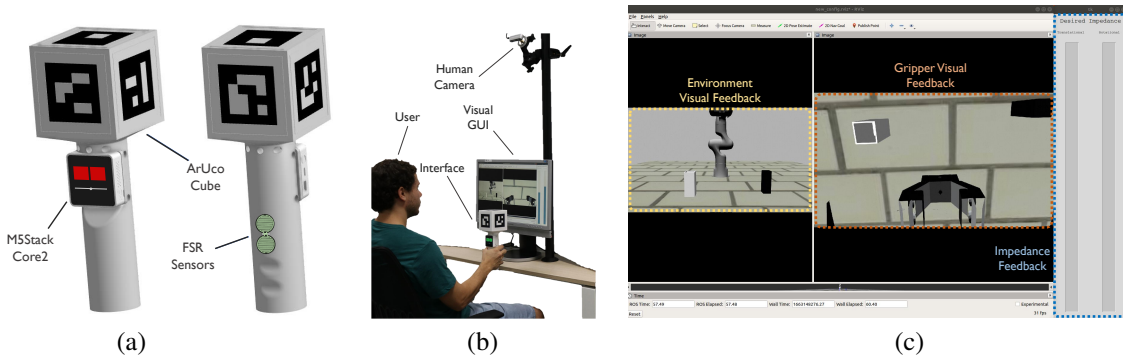


Fig. 2. Design concept of the proposed Tele-impedance System. (a) Tele-impedance interface, (b) Human side, (c) Human visual feedback composed of (from left to right) static camera view showing robot side and its workspace, wrist-mounted camera view, and feedback of the commanded translational and rotational impedance.

III. METHODOLOGY

A. Data Collection System Requirements

The problem tackled in this work can be stated as that of creating a data collection system for obtaining expert demonstrations τ_E of state-action trajectories such that i) it is as suitable as possible for task-agnostic demonstration-based RL methods, ii) the VICES action space [13] is chosen in the problem formulation, and iii) the system is usable by the widest possible range of researchers. The state in the problem formulation can be of different nature, e.g. exteroceptive, proprioceptive or a mix of the two. We consider the problem where the state (or observation) consists of current end-effector position, estimated interaction force at the end-effector and views from cameras, with humans and robots getting the same view.

To fulfill the three points aforementioned, we formulate the following requirements developed according to insights gathered from the works in Section II:

- *Simplicity.* The larger and more various are the datasets, the more the RL approaches considered will benefit from them. Thus, the system should be easy and fast to use.
- *Accessibility.* The system should be as much accessible as possible, meaning that the tools used for implementing it should be ubiquitous, low-cost, and thus easily accessible for a research lab that starts from scratch.
- *Compatibility.* RL is often cast as the problem of learning visuomotor policies, i.e. motor policies conditioned on images. In order to make the expert demonstrations usable, they should be compatible with the interactions of the agent with the environment during RL training. Thus, firstly visual artifacts should be avoided in the context of visuomotor policy learning. That is, the humans should not be present in the scene during the demonstration (e.g., in kinesthetic teaching) since they will not be during agent training. Secondly, human visual feedback and the images provided as input to the robot’s policy should coincide in order to avoid that the decisions taken by the human cannot be correlated to the input image gotten by the robot (e.g., when the robot’s view is obstructed but the human one is not).
- *Intuitiveness.* Human demonstrations represent more

natural and meaningful behaviour compared to machine-generated ones, and could determine the success of RL approaches. To retain this advantage, user intuitiveness needs to be ensured. The interface used determines how demonstrations are provided and the feedback the humans receive. Both these aspects play a key role in ensuring intuitiveness.

- *Relaxed Precision.* In the RL approaches considered, expert demonstrations can provide a coarse indication of what actions are worth to be taken in a state since they aid learning for a new task by guiding exploration. Therefore, framework precision can be slightly relaxed as long as meaningful behaviors can be still provided. In other words, the demonstrations are task-agnostic, meaning that they are not supposed to be used to learn the same task that is demonstrated, but to create behavior priors for the RL agent that is learning a new different task.
- *VICES Action Space.* The framework should be able to command desired poses and Cartesian impedances to the robot.

B. Data Collection Framework Design

Our data collection framework is depicted in Fig. 2. In particular, Fig. 2(b) shows all the components of the framework, Fig. 2(a) enlarges and highlights the components of the physical interface used by the human to teleoperate the robot and Fig. 2(c) depicts the visual feedback provided to the human.

1) *Desired Pose command:* The desired end-effector pose can be commanded to the robot through a MoCap system that tracks directly the human hand, gaining intuitiveness. Technologies based on wearable sensors, e.g., XSens based on Inertial Measurement Units (IMUs), while highly precise usually are not easily accessible (i.e. expensive). Contrarily, some camera-based technologies are open-source, have acceptable precision, and cheap. In addition, accurately placing sensors and calibrating them takes additional time, at the expense of simplicity. Therefore, we opt for tracking a polyhedron of ArUco markers similarly to the work in [26], using an RGB camera and the open-source ArUco detect

algorithm³. The polyhedron of ArUco markers is attached on top of an interface that is tightly grasped by the human during teleoperation, both are 3D printed. One of the main limitations is that the ArUco marker might be obstructed, which would not happen with previously mentioned wearable sensors. That is why we use a polyhedron of ArUcos and we mount it on top of the interface grasped by the human.

Let \mathcal{C} , \mathcal{A}_i , \mathcal{I} and \mathcal{W}_h be respectively the camera frame, the i -th ArUco marker frame ($i = 1, \dots, n$ with $n + 1$ being the number of faces of the polyhedron), the interface frame located where the human grasps it, and the world frame (fixed) on human side.

The polyhedron of ArUco markers is rigidly attached on top of the interface far enough from the human grasping point. This allows for a) making the ArUco markers more visible to the camera and b) knowing the (constant) homogeneous transformation $\mathbf{T}_{\mathcal{A}_i\mathcal{I}}$ from each ArUco frame to the interface frame. Considering the i -th ArUco marker, the ArUco detect algorithm provides an estimation of $\mathbf{T}_{\mathcal{C}\mathcal{A}_i}$, while the constant transformation $\mathbf{T}_{\mathcal{W}_h\mathcal{C}}$ is known once the camera has been fixed and calibrated. Therefore, the i -th ArUco marker provides the homogeneous transformation $\mathbf{T}_{\mathcal{W}_h\mathcal{I}}^i$ obtained as:

$$\mathbf{T}_{\mathcal{W}_h\mathcal{I}}^i = \mathbf{T}_{\mathcal{W}_h\mathcal{C}}\mathbf{T}_{\mathcal{C}\mathcal{A}_i}\mathbf{T}_{\mathcal{A}_i\mathcal{I}}. \quad (1)$$

Since only one camera is used, one ArUco marker is not sufficient to command the desired pose when the desired orientation changes. The polyhedron of ArUco markers can circumvent this problem and can make the pose tracking more robust.

Let $c\mathbf{p}_{\mathcal{C}\mathcal{A}_i}$ be the position vector that goes from the camera to the i -th ArUco marker expressed in the camera frame, and let $\mathbf{R}_{\mathcal{C}\mathcal{A}_i}$ be the rotation matrix describing the orientation of the i -th ArUco marker with respect to the camera. The third column of $\mathbf{R}_{\mathcal{C}\mathcal{A}_i}$ is $c\mathbf{e}_z^{\mathcal{A}_i}$, that is the z -axis of frame \mathcal{A}_i expressed in frame \mathcal{C} , which is the one coming out perpendicularly from the ArUco surface. In agreement with the considerations in [27] regarding the pose ambiguity problem and accordingly to the heuristic observations, we consider $\mathbf{T}_{\mathcal{C}\mathcal{A}_i}$ more reliable when the angle between $c\mathbf{p}_{\mathcal{C}\mathcal{A}_i}$ and $-c\mathbf{e}_z^{\mathcal{A}_i}$ is small. We compute the cosine similarity between $c\mathbf{p}_{\mathcal{C}\mathcal{A}_i}$ and $-c\mathbf{e}_z^{\mathcal{A}_i}$ as:

$$\alpha_i = \frac{-c\mathbf{e}_z^{\mathcal{A}_i} \cdot c\mathbf{p}_{\mathcal{C}\mathcal{A}_i}}{\| -c\mathbf{e}_z^{\mathcal{A}_i} \| \| c\mathbf{p}_{\mathcal{C}\mathcal{A}_i} \|} = \cos \theta_i, \quad (2)$$

where θ_i is the angle between the two vectors and $\alpha_i \in [0, 1]$. Then, $\mathbf{T}_{\mathcal{W}_h\mathcal{I}}$ is computed as:

$$\mathbf{T}_{\mathcal{W}_h\mathcal{I}} = \sum_{i|\alpha_i > \alpha_{thr}} \mathbf{T}_{\mathcal{W}_h\mathcal{I}}^i \alpha_i, \quad (3)$$

where α_{thr} is a threshold below which $\mathbf{T}_{\mathcal{C}\mathcal{A}_i}$ is not considered. The higher n , the more robust the pose tracking of the interface is. In this work we use $n = 5$, i.e. the polyhedron used is a cube. As we will demonstrate in Section IV this choice is enough in our case to have a sufficiently precise

pose estimation. We choose $\theta_{thr} = \arccos(\alpha_{thr}) = 50^\circ$ resulting in $\alpha_{thr} \approx 0.64$.

2) *Desired Impedance command*: Previous work has tried to estimate the human impedance to transfer it to the robot, in order to make commanding an impedance as natural, intuitive, and precise as possible [28], [29]. However, for estimating it precisely, sensors that measure human muscular activity, e.g., electromyographic sensors (EMGs), are needed. These sensors are not easily accessible, require calibration, and should be worn by humans. Instead, we decide to use resistive buttons. The pressure applied to them deforms an internal resistance and causes a voltage signal change, which in turn can be mapped to a change in the desired impedance.

Formally, we command a desired diagonal stiffness $\mathbf{K}_d = \text{diag}\{\mathbf{k}_d\} \in \mathbb{R}^{6 \times 6}$, bounded between a minimum and a maximum value, and the corresponding desired damping $\mathbf{D}_d = \text{diag}\{\mathbf{d}_d\} \in \mathbb{R}^{6 \times 6}$ is obtained through double diagonalization formula:

$$\mathbf{d}_d = 2 \cdot 0.707 \cdot \sqrt{\mathbf{k}_d}. \quad (4)$$

The buttons are connected to the interface that sends the desired Cartesian impedance to the robot controller.

In the process of deciding how many impedance dimensions the human expert should be able to command independently, which corresponds to the number of buttons to include on the physical interface, we observe that too many buttons would degrade the intuitiveness of the framework. Hence, we decide to consider a trade off between intuitiveness and complexity and included only two buttons, one for translational and one for rotational impedance.

3) *Feedback Provided to Human and Other Commands*: When learning visuomotor policies for manipulation tasks, a viable choice is to use a wrist-mounted camera and a static camera recording the whole task workspace [30]. The views of these cameras are provided as feedback to the human. This choice could degrade intuitiveness (the cameras could provide a non-functional view of the scene for the task purpose) but the benefit in terms of compatibility outbalances this negative aspect (human and robot should condition actions on the same view). Another important feedback is the interaction force of the robot end-effector with the environment and the commanded impedance. The former is provided by means of vibrotactile feedback generated by the interface while the latter through visual feedback of progress bars on the screen.

Additionally, the interface provides a touch screen where we implement:

- A button to command the gripper (open/close).
- A button to deactivate and activate the teleoperation.
- A slider to command the scaling factor between the leader and the follower movements.

C. Tele-Impedance Framework Components

According to the aforementioned design description, the tele-impedance framework shown in Fig. 2 is implemented. The diagram of Fig. 3 illustrates how all the final components of the framework interact with each other. The interface uses

³http://wiki.ros.org/aruco_detect

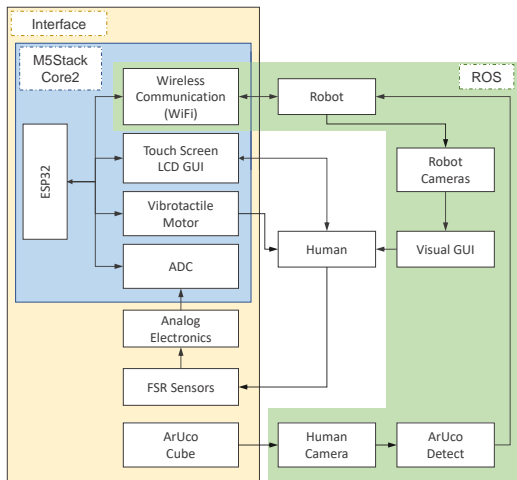


Fig. 3. Block diagram of the tele-impedance framework. The human-robot communication is obtained thanks to the tele-impedance interface (see Fig. 2) and visual GUI. The communication between the different hardware elements according to the connections in the figure is done in ROS.

an M5Stack Core2⁴ as a human-robot communication unit. An ESP32 microprocessor manages the hardware components of the interface and communicates with the robot via WiFi. The human can regulate the impedance of the robot with two Force Sensor Resistor (FSR). A module formed by two voltage dividers with resistors of $3.3k\Omega$ is used as the analog electronics of the FSR. The human can also communicate with the robot through the M5Stack Core2 touch screen. In addition, the human receives feedback of the task through the visual GUI, and vibrotactile feedback of the force measure at the end-effector of the robot. More details about the hardware components and assembly instructions are included in the open repository of the framework.

D. Data Logging and Simulation Environment

An important feature for data collection systems for learning applications is the autonomous data logging. We include such feature in the code we open source, so that states-action trajectories can be readily available once the demonstration is terminated. In addition, we provide the simulation environment used for the experiment presented in the next section.

IV. EXPERIMENTS AND RESULTS

In order to validate the designed framework, we demonstrate that a human expert is able to fulfill a highly interactive task having well-defined impedance requirements when using our system. Since our framework is thought for collecting data for learning applications, we perform the experiment in simulation (gazebo⁵) [31]. We opted for simulation because training of RL agents for these tasks is usually performed in simulation. The reason is that algorithms remain too sample inefficient for training directly in real world, other than the problem of safety and resetting of the environment. Then, a sim-to-real transfer phase is usually needed afterwards. In

any case, the proposed framework gives the user the freedom to use it both in simulation or in real world.

A. Experiment

The experimental setup can be viewed in Fig. 2(b)-(c). A human expert is asked to perform a peg-in-hole task fulfilling specific impedance requirements. We use simulated models of the RGB-D Realsense d435 cameras, one mounted on the wrist of the robot and one fixed in front of the robot workspace using OpenVico [32]. Peg and hole are parallelepipeds of dimensions $50 \times 50 \times 150$ mm and $56 \times 56 \times 150$ mm (width \times depth \times length), respectively. The impedance requirements are as follows:

- 1) Keep a low impedance when reaching the peg in order to reduce the impact due to possible collisions.
- 2) When carrying the peg, increase the impedance in order to reject tracking errors due to the object's weight.
- 3) When carrying the peg close to the hole, keep a low impedance to favor alignment and to avoid the arising of high contact forces due to contact between the peg's and the hole's sides.
- 4) When inserting the peg inside the hole, increase the translational impedance to overcome friction and enable insertion while keeping low rotational impedance to avoid rotating the peg when it is inside the hole.

ROS bags⁶ are used for data logging, which have been widely used from researchers to record datasets, visualize, label, and store them for future use.

B. Results

Fig. 4 depicts the results of the experiment. The human expert completes the task successfully, as shown by the hole and peg poses. In addition, the impedance requirements are fulfilled. Despite the misalignment between the peg and the hole at the beginning of the insertion, the interaction force does not grow too high thanks to the low impedance. Later, the high translational impedance enables the completion of the insertion by easing the sliding of the peg into the hole. Since the translational impedance is increased along all directions, the contact force between the peg and the hole grows.

V. CONCLUSION

We have shown that our framework allows us to readily provide the successful demonstration of a peg-in-hole task including impedance requirements. The successful completion of the task implies that the tracking performance of our system is sufficiently precise for teleoperating the robot, recalling that the human is able to compensate for small tracking errors if provided with visual feedback. Datasets of these demonstrations have the potential to merge the RL approaches considered in this paper with an impedance-based action space formulation like the VICES proposed in [13].

We have made our system accessible to the research community hoping that it can advance research in the field

⁴<https://docs.m5stack.com/en/core/core2>

⁵<https://gazebo.org/home>

⁶<http://wiki.ros.org/Bags>

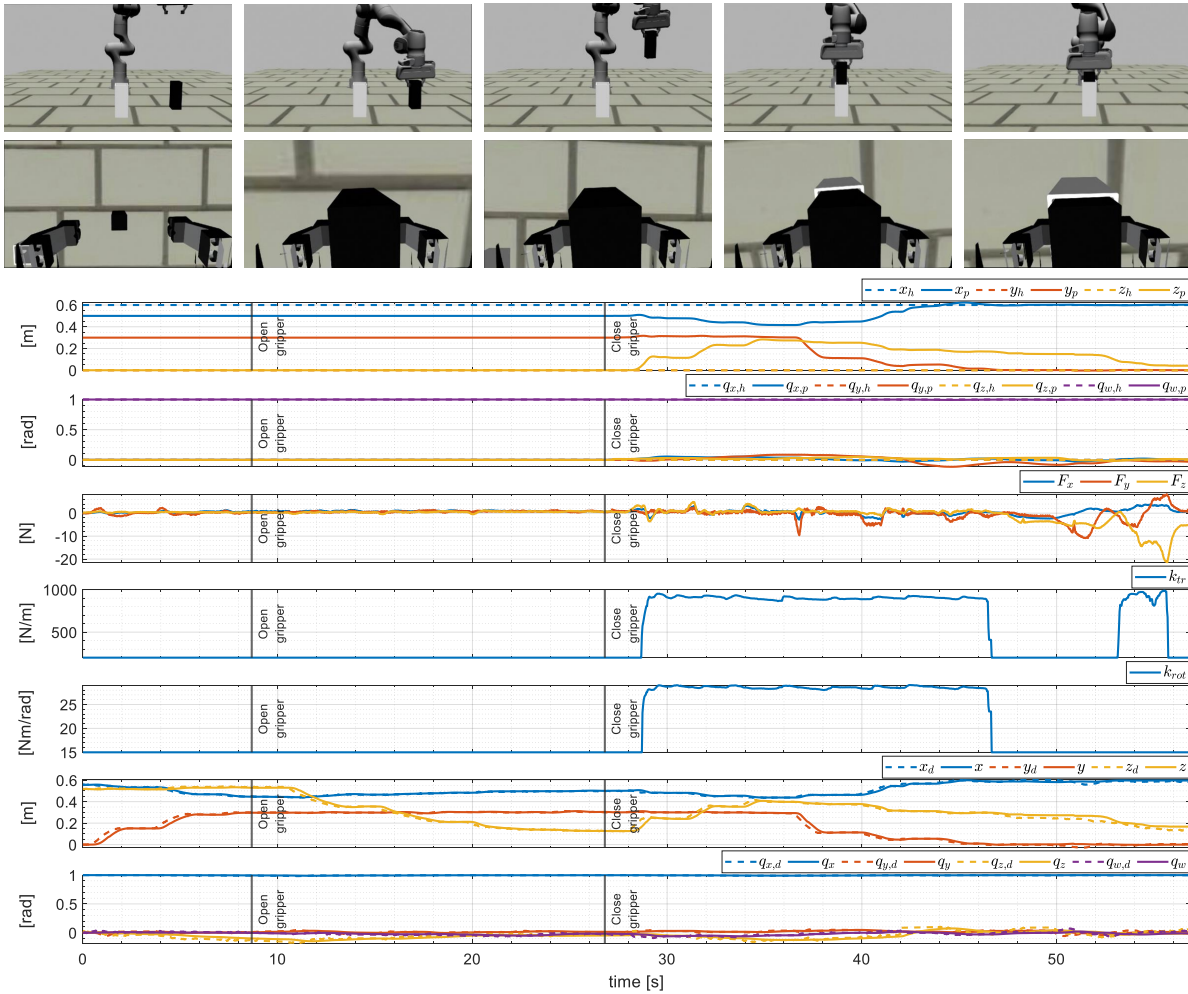


Fig. 4. Top: excerpts of the visual feedback gotten by the human expert during the experiment. Bottom: plot depicting (from top to bottom) hole and peg positions, hole and peg orientations, estimated external forces at the robot end-effector, translational Cartesian stiffness, rotational Cartesian stiffness, desired and current end-effector position, and, desired and current end-effector orientation. In the first 27 seconds the end-effector is moved above the peg with low impedance and the gripper is closed to grasp the peg. Then, the peg is carried above the hole with high impedance to command precise motions despite the weight of the peg, and to reach a good peg-hole alignment before insertion. When the peg contacts the hole, the impedance is reduced to avoid high contact forces due to small misalignments, and to obtain a natural alignment thanks to the high compliance. Finally, only the translational impedance is increased to overcome the friction during insertion, while the rotational impedance is kept low to maintain alignment and avoid the generation of excessive torques while the peg is into the hole.

of robotics for learning contact-rich manipulation tasks. The cost of our system ranges between 200\$ and 400\$ depending on the type of camera used (which is not required to provide depth information) and excluding the cost of the computer. Thus, it is much more cost effective and affordable than other commercial MoCap systems ($\approx 30000\$$) or VR devices ($\approx 3000\$$). In future works, we plan to use our framework to collect expert demonstrations to learn contact-rich tasks using the RL approaches considered in this work.

REFERENCES

- [1] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, “Imitation learning: A survey of learning methods,” *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, pp. 1–35, 2017.
- [2] M. Bain and C. Sammut, “A framework for behavioural cloning,” in *Machine Intelligence 15*, 1995, pp. 103–129.
- [3] S. Ross, G. Gordon, and D. Bagnell, “A reduction of imitation learning and structured prediction to no-regret online learning,” in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 627–635.
- [4] A. Y. Ng, S. J. Russell, *et al.*, “Algorithms for inverse reinforcement learning,” in *Icml*, vol. 1, 2000, p. 2.
- [5] A. Y. Ng, D. Harada, and S. Russell, “Policy invariance under reward transformations: Theory and application to reward shaping,” in *Icml*, vol. 99, 1999, pp. 278–287.
- [6] A. Singh, H. Liu, G. Zhou, A. Yu, N. Rhinehart, and S. Levine, “Parrot: Data-driven behavioral priors for reinforcement learning,” *arXiv preprint arXiv:2011.10024*, 2020.
- [7] K. Pertsch, Y. Lee, and J. J. Lim, “Accelerating reinforcement learning with learned skill priors,” *arXiv preprint arXiv:2010.11944*, 2020.
- [8] K. Pertsch, Y. Lee, Y. Wu, and J. J. Lim, “Guided reinforcement learning with learned skills,” *arXiv preprint arXiv:2107.10253*, 2021.
- [9] A. Gupta, V. Kumar, C. Lynch, S. Levine, and K. Hausman, “Relay policy learning: Solving long-horizon tasks via imitation and reinforcement learning,” *arXiv preprint arXiv:1910.11956*, 2019.
- [10] A. Singh, A. Yu, J. Yang, J. Zhang, A. Kumar, and S. Levine, “Cog: Connecting new skills to past experience with offline reinforcement learning,” *arXiv preprint arXiv:2010.14500*, 2020.
- [11] N. Y. Siegel, J. T. Springenberg, F. Berkenkamp, A. Abdolmaleki, M. Neunert, T. Lampe, R. Hafner, N. Heess, and M. Riedmiller, “Keep doing what worked: Behavioral modelling priors for offline

- reinforcement learning,” *arXiv preprint arXiv:2002.08396*, 2020.
- [12] A. Ajoudani, N. Tsagarakis, and A. Bicchi, “Tele-impedance: Teleoperation with impedance regulation using a body-machine interface,” *The International Journal of Robotics Research*, vol. 31, no. 13, pp. 1642–1656, 2012.
- [13] R. Martín-Martín, M. A. Lee, R. Gardner, S. Savarese, J. Bohg, and A. Garg, “Variable impedance control in end-effector space: An action space for reinforcement learning in contact-rich tasks,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1010–1017.
- [14] S. Levine, A. Kumar, G. Tucker, and J. Fu, “Offline reinforcement learning: Tutorial, review, and perspectives on open problems,” *arXiv preprint arXiv:2005.01643*, 2020.
- [15] A. Kumar, A. Zhou, G. Tucker, and S. Levine, “Conservative q-learning for offline reinforcement learning,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 1179–1191, 2020.
- [16] A. Abdolmaleki, J. T. Springenberg, Y. Tassa, R. Munos, N. Heess, and M. Riedmiller, “Maximum a posteriori policy optimisation,” *arXiv preprint arXiv:1806.06920*, 2018.
- [17] V. Kumar and E. Todorov, “Mujoco haptic: A virtual reality system for hand manipulation,” in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 657–663.
- [18] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.
- [19] A. Nair, M. Dalal, A. Gupta, and S. Levine, “Accelerating online reinforcement learning with offline datasets,” *arXiv preprint arXiv:2006.09359*, 2020.
- [20] J. Fu, A. Kumar, O. Nachum, G. Tucker, and S. Levine, “D4rl: Datasets for deep data-driven reinforcement learning,” *arXiv preprint arXiv:2004.07219*, 2020.
- [21] A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine, “Learning complex dexterous manipulation with deep reinforcement learning and demonstrations,” *arXiv preprint arXiv:1709.10087*, 2017.
- [22] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel, “Deep imitation learning for complex manipulation tasks from virtual reality teleoperation,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 5628–5635.
- [23] A. Nair, B. McGrew, M. Andrychowicz, W. Zaremba, and P. Abbeel, “Overcoming exploration in reinforcement learning with demonstrations,” in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 6292–6299.
- [24] M. Vecerik, T. Hester, J. Scholz, F. Wang, O. Pietquin, B. Piot, N. Heess, T. Rothörl, T. Lampe, and M. Riedmiller, “Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards,” *arXiv preprint arXiv:1707.08817*, 2017.
- [25] A. Mandlkar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay, *et al.*, “Roboturk: A crowdsourcing platform for robotic skill learning through imitation,” in *Conference on Robot Learning*. PMLR, 2018, pp. 879–893.
- [26] P.-C. Wu, R. Wang, K. Kin, C. Twigg, S. Han, M.-H. Yang, and S.-Y. Chien, “Dodecapen: Accurate 6dof tracking of a passive stylus,” in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017, pp. 365–374.
- [27] G. Schweighofer and A. Pinz, “Robust pose estimation from a planar target,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2024–2030, 2006.
- [28] A. Ajoudani, N. G. Tsagarakis, and A. Bicchi, “Tele-impedance: Towards transferring human impedance regulation skills to robots,” in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 382–388.
- [29] Y. Wu, F. Zhao, W. Kim, and A. Ajoudani, “An intuitive formulation of the human arm active endpoint stiffness,” *Sensors*, vol. 20, no. 18, p. 5357, 2020.
- [30] A. Mandlkar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, “What matters in learning from offline human demonstrations for robot manipulation,” *arXiv preprint arXiv:2108.03298*, 2021.
- [31] H. Choi, C. Crump, C. Duriez, A. Elmquist, G. Hager, D. Han, F. Hearl, J. Hodgins, A. Jain, F. Leve, *et al.*, “On the use of simulation in robotics: Opportunities, challenges, and suggestions for moving forward,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 1, p. e1907856118, 2021.
- [32] L. Fortini, M. Leonori, J. M. Gandarias, and A. Ajoudani, “Open-vico: An open-source gazebo toolkit for multi-camera-based skeleton tracking in human-robot collaboration,” *arXiv preprint arXiv:2203.14733*, 2022.