Disagreement-Aware Variable Impedance Control for Online Learning of Physical Human-Robot Cooperation Tasks

van der Spaa, L.F.; Franzese, G.; Kober, J.; Gienger, Michael

Publication date
2022

Document Version
Final published version

Citation (APA)
van der Spaa, L. F., Franzese, G., Kober, J., & Gienger, M. (2022). *Disagreement-Aware Variable Impedance Control for Online Learning of Physical Human-Robot Cooperation Tasks*. Paper presented at Workshop "Shared Autonomy in Physical Human-Robot Interaction: Adaptability and Trust".

Disagreement-Aware Variable Impedance Control for Online Learning of Physical Human-Robot Cooperation Tasks

Linda van der Spaa[1,2], Giovanni Franzese[1], Jens Kober[1], Michael Gienger[2]

***Abstract*— In order to make the coexistence between humans and robots a reality, we must understand how they may cooperate more effectively. Modern robots, empowered with reliable controls and advanced machine learning reasoning can face this challenge. In this article, we presented a Disagreement-Aware Variable Impedance (DAVI) Controller, where the robot stiffness is regulated as a function of the perceived disagreement with the human cooperator. We tested the algorithm on a 7 DoF Franka Emika Panda robot performing the learning of a pick&place task with continuous adaptation of the goal location and the via-points with human interactive corrections, triggered by our proposed approach. A validation study was conducted with 5 users in order to understand the reliability of the method.**

I. Introduction

The strength of a team depends very much on the ability of its members to cooperate. Humans and robots have different strengths and weaknesses. Potentially, teaming up a robot with a human would allow the partners in the team to complement each other to the benefit of both. However, the actual benefit depends on the cooperation skills of both the human and the robot. In case these are lacking, the attempted cooperation may instead inadvertently lead to reduced performance.

Successful cooperation requires partner-awareness, and the ability to communicate and negotiate on personal preferences (how to do something), intentions (what to achieve) and constraints. Factors like preferences depend, at least partially, on the partners in the team and the cooperation between them. Therefore, we argue that most effective cooperation can be achieved when learned online; i.e. while trying to cooperate, the agents update their behavior to improve on the overall result. As cooperation skills grow over time, preferences may change as well as what cooperative behavior may be optimal. Life-long learning is required in order to keep adapting accordingly.

Our specific interest is in physical human-robot cooperation (pHRC) tasks with prolonged physical interaction continuing over a sequence of dependent actions. In such tasks, haptic communication has been shown effective in integrating intentions in shared decision making [1], and is actually able to lead to faster optimal decisions than explicit communication [2]. We focus on tasks where human and robot move an object to a new location in space, only communicating intuitively through the interaction forces (see Fig. 1). The robot has to learn how the human prefers to do the task in order to provide appropriate support.

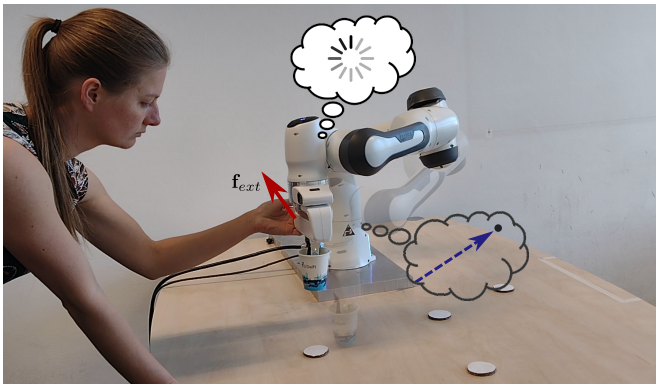

Fig. 1. Scenario for (cooperatively) moving a cup to one of the available coasters. Initially, the robot does not know where or how to move the cup. The robot can be guided even when the human moves the cup without touching the robot.

Starting from an initial solution which allows the human to finish the task together with the robot, our objective is to have the robot learn to adhere to the human partner's personal preferences, learning from feedback the human implicitly provides in the interaction. Here, we propose a learning and control framework (Algorithm 1) that learns high level target policies $\pi(\boldsymbol{x})$ and allows a modulation of the robot stiffness as a function of the disagreement with the human. The control of the task is traded between the human and the robot and when the robot is passive is interactively updating the desired policy.

The variable admittance/impedance of the robot for a safer human robot interaction was already proposed in the literature. For example, in [3], the robot admittance is increased when the human applies sufficient positive work to the system, effectively and smoothly changing the robot behavior to that of a passive follower but without interactively improving or modifying the desired execution of the task. Alternatively, [4] proposes to decay the stiffness as a function of the epistemic uncertainty of the policy encoded with a Gaussian Process. We show equally smooth transitioning for impedance controlled trajectory tracking, ramping down the impedance upon detection of significant interaction force.

We tested our framework in a cooperative pick-and-place scenario with a 7 DoF Franka Emika Panda robot.

[1]Delft University of Technology, Mekelweg 2, 2628CD The Netherlands {l.f.vanderspaa, g.franzese, j.kober}@tudelft.nl

[2]Honda Research Institute Europe, Offenbach, Germany michael.gienger@honda-ri.de

Algorithm 1: DAVI controller framework

1 Initialize $\{\mathcal{S}, \pi(\mathbf{x})\}$
2 **while** *Episode* **do**
3 $\quad RobotInControl = (K > 0)$
4 $\quad$ **if** $\dot{\mathbf{x}} = 0 \ \wedge$!*RobotInControl* **then**
5 $\quad\quad \mathcal{S} = \mathcal{S} \cup \mathbf{x}$
6 $\quad\quad \pi(\mathbf{x}_0) = x$
7 $\quad$ **if** $((RobotInControl \ \wedge \ \|\mathbf{x} - \mathbf{x}_g\| < \varepsilon) \vee$ $(!RobotInControl \ \wedge \ \mathbf{x} \in \mathcal{S})) \ \wedge \ \pi(\mathbf{x}) \neq \{\}$ **then**
8 $\quad\quad RobotInControl = True$
9 $\quad\quad \mathbf{x}_g = \pi(\mathbf{x})$
10 $\quad\quad \mathbf{x}_0 = \mathbf{x}$
11 $\quad$ **if** *RobotInControl* **then**
12 $\quad\quad \hat{\mathbf{x}} = f(\mathbf{x}, \mathbf{x}_g, \mathbf{x}_0)$ Eq. (2)
13 $\quad\quad \gamma = \text{Disagreement}(\mathbf{f}_{\text{ext}})$
14 $\quad\quad \dot{K} = \text{ImpedanceModulation}(\gamma)$ Eq. (4)
15 $\quad$ ImpedanceControl$(\hat{\mathbf{x}}, K)$

II. DAVI Controller

Our DAVI controller allows an incremental learning of the policy while being user friendly on the interaction with the human. The Algorithm 1 gives a hint on what happens during every learning episode. We can recognize:

- an initial set of states $\mathcal{S}$ (which could be empty)
- a policy $\pi(\boldsymbol{x})$ that learns the desired next goal of the robot
- an impedance control low defined by the attractor $\hat{x}$ and the stiffness K, allowing safe and comfortable human-robot interaction during action execution
- a disagreement detection based on the sensed external force
- the possible switching of the controller and move the robot in a gravity compensation mode letting the human to provide kinesthetic teaching.

A. Cartesian Impedance Control

As a base control layer, we use a Cartesian impedance controller. Briefly, in Cartesian impedance control [5], the end-effector dynamics are modeled in the form of a mass-spring-damper system

$$\mathbf{\Lambda}(\mathbf{q})\ddot{\mathbf{x}} = \mathbf{K}\Delta\mathbf{x} - \mathbf{D}\dot{\mathbf{x}} + \mathbf{f}_{\text{ext}}, \tag{1}$$

where $\mathbf{\Lambda}(\mathbf{q})$ is the physical system's Cartesian inertia matrix, $\mathbf{K}$ is a diagonal matrix with the desired stiffness in the principal directions, $\mathbf{D}$ is the corresponding critical damping matrix, and $\mathbf{f}_{\text{ext}}$ are the external forces. The external forces are estimated using the provided model of the robot (mass matrix, Coriolis, gravity) and the estimated joint friction provided in [6].

We distinguish between active and passive mode. In active mode, the end-effector is controlled to follow a trajectory. In passive mode, the end-effector stiffness and damping are set to zero. In this circumstance the robot is still gravity compensated. At any time during active mode, a detected disagreement would trigger a transition to passive mode; this allow the human to kinesthetically demonstrate the new desired behavior. At all times, the robot records the states and actions it observes during interactive task execution so it can learn from them.

B. State and Action Learning from Interactions

We define our set of states $\mathcal{S}$ as the points through which the human may want the robot to pass when doing the task. Here, we just consider a 3D end-effector workspace with a fixed end-effector orientation, not caring about the robot's joint configuration. Initially, the state space only contains the starting position, and no desired actions are known. A new state is added when the robot is stopped in an unknown state for 0.25 seconds, line 5 of Al. 1. The data aggregation is communicated to the user by a haptic vibration of the end-effector. The next goal $\mathbf{x}_g$ of the robot is considered as the high-level action coming from the policy, line 9 of Al. 1.

During active mode, if a known state is visited, the robot will remember where to go next. However, if the robot is corrected to go to a different subsequent state, which may also be a known state, the policy is update, line 6 of Al. 1. If the robot stays at a state for at least 5 seconds, the state is flagged as a final goal state. Before the state is added, the robot hand signals with a countdown of three vibrations to alert the human. This would allow them to continue the demonstration if the state was not their intended final goal. During passive mode, the robot is considered to be in a state if the end-effector is within 2 cm distance of the point defining the state. To help the human feel where the states are, we let the robot transition to active attraction to the state point if it is within 10 cm distance.

We connect the current state and the next desired one with a straight-line trajectory, assuming the absence of obstacles. For trajectory tracking, we employ a relatively low impedance (of max. 600 N/m), allowing safe physical interaction. We apply online attractor distance modulation [7] to allow a reactive following with a limit on the force exerted by the robot according to

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \alpha(\mathbf{x}_g - \mathbf{x}_0) \qquad 0 \leq \alpha \leq 1 \tag{2}$$

$$\dot{\alpha} = \frac{v_{\text{ref}}}{\|\mathbf{x}_g - \mathbf{x}_0\|} \frac{1}{1 + \|\mathbf{x} - \hat{\mathbf{x}}\|/l} \tag{3}$$

where α determines the progress of the linear trajectory, l the equivalent tracking error that makes the progress rate to drop to half, and v_{ref} (of 0.3 m/s) is the desired Cartesian linear velocity.

C. Disagreement Detection

Intuitively, disagreement can be detected based on interaction force/torque, or deviations from the expected trajectory. The two are coupled by the set robot impedance(s). Alternatively, we can detect disagreement based on the human *virtual* work, the work they would do if the robot would not exert a force. In contrast to reacting to the actual work the human does [3], this also detects disagreement when the human keeps the robot from moving.

Looking at Eq. (1) and neglecting, for simplicity, the robot accelerations and velocity, the external force is displacing the robot according to $\mathbf{f}_{\text{ext}} = -\mathbf{K}\Delta\mathbf{x}$. This means that we can estimate the virtual work done by the external forces as $E_{\text{ext}} = -\mathbf{f}_{\text{ext}}\Delta\mathbf{x} = \mathbf{f}_{\text{ext}}^T\mathbf{K}^{-1}\mathbf{f}_{\text{ext}}$. This equation is meaningful as long as the stiffness is positive. Considering that the injected external energy can only be estimated as a function of the norm of the external force (and controlled stiffness), we assign a negative value to our disagreement constant γ every time the external force is beyond a safety threshold $\mathbf{f}_{\text{ext}}^{\text{th}}$ and positive otherwise. The stiffness changes according to

$$\dot{K} = \text{sign}(\gamma)K_{\max}/\Delta t_{\text{transition}}. \tag{4}$$

The stiffness value will saturate when it goes beyond the set max limit. The hyperparameter $\Delta t_{\text{transition}}$ regulates the desired stiffness rate during the negotiation phase on whom has fully control of the task. If an external force was applied unintentionally, as long as the interaction was not longer than $\Delta t_{\text{transition}}$, then the impedance has not dropped entirely to zero and hence the passive mode is not activated. When the force drops again below the safety threshold, positive γ of Eq. (4) will ramp the stiffness back up to the maximum. This hysteresis time band helps to prevent unintentional switching from robot to human control [8]. Once the impedance on the trajectory the robot was following has become zero, the robot changes to passive mode. From now on, it keeps track of its proximity to the states it has stored in its model. The robot transitions back to active mode, i.e., γ becomes positive, when it detects itself in a state (other than the one it just came from) where it knows what action to take.

III. Experimental Validation

We test our general framework on the pick&place task shown in Fig. 1. The cup can be moved to one of the other coasters, but our robot has no information on them or any prior on how it might move. For a parameterized behavior, knowledge of the environment, such as where the coasters are, would improve generalizability. But just for showing the use of interactive learning with disagreement-awareness, we test in a fixed environment, only using the end-effector position and external force data.

We asked five people to teach the same task of pick&place of Fig. 1. Their expertise in robotics ranged from beginner to expert. The goal was to challenge the algorithm robustness with all possible interactions, from under to over-confident. The participants first showed the robot to place the cup on one of the other coasters, with an arbitrary number of intermediate states. Next, they altered the trajectory to pass through at least one additional or alternative state. At least once, they were asked to steer the robot to another coaster, a new goal state. Each participant was asked to disagree with the robot at least once, moving the robot to a different point in space, unknown and known, in each of the following ways:

- moving the robot in a different direction w.r.t. the trajectory followed initially,
- stopping the robot on the trajectory it is executing,

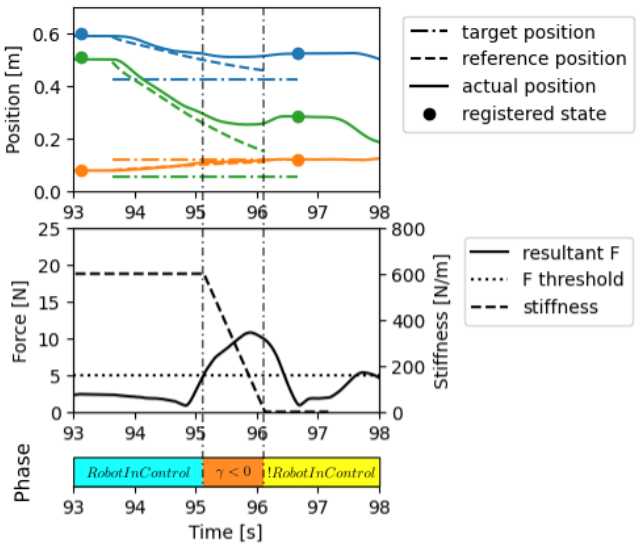

Fig. 2. Position, resultant force and stiffness of the end-effector during an action that is corrected to a new position in space on the trajectory the robot was executing. The colors in the position plots indicate x (blue), y (orange) and z (green) respectively. The phase bar shows first the robot is in control. Upon detecting disagreement, the control transferred to the human.

- making the robot move over a state without stopping there, teaching it to move to a further lying state instead.

Figure 2 shows the position and force result of a disagreement case. A new state is learned on the executed trajectory. It is a typical force profile for all cases of disagreement. Cases in which a state is added in a different region in space generally only show a higher force peak. After the disagreement phase, the human is free to teach the robot a new state, which it registers when its motion is stopped. This is observed at $t = 96.7$s. After that, we see the the force on the end-effector increase again. Since the robot has arrived in a state it had not seen before, the user is performing a kinesthetic demonstration to show the robot what it should do the next time it arrives in the state that was just observed.

Less experienced users struggled considerably more with deciding on their preferences and remembering them. While they could teach the robot the same things, they experienced increased difficulty. They tended to be more surprised when the robot would activate to start moving towards a state it had recognized as close. We let that be the robot's way of asking the human: "is this where you want to go?" But less experienced users reflexively let go, or at least did not immediately resist the robot, which the robot would interpret as confirmation, until the human would actively disagree again. This led to some confusion, stiffness going up and down and some additional interaction forces. However, when the users understood they were basically negotiating with the robot, they could successfully push their point and make the robot understand.

Figure 3 shows a state set that is learned with an inexperienced user. At the start, the robot only knows the state marked "t=0.0". Each state is marked with the time it was added to the robot's state set. The recorded end-

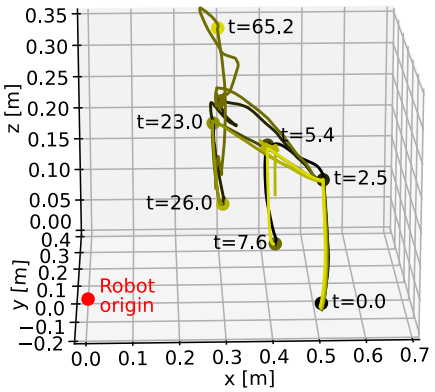

Fig. 3. States and trajectories from which they were observed colored in order of learning from dark to light. Time stamps show when a state was added to the robot state space.

effector trajectories are also shown in the figure. Both states and trajectories are color coded a lighter shade for each newly observed one. The figure shows the new states learned on the demonstrated trajectories. Changes in preferred state sequence could be demonstrated with smooth trajectories made possible by the smooth mode transitions, and once it was recognized that new behavior was being demonstrated, the robot lets the human demonstrate without interference. A video of the experiment, as well as our code, can be found in our GitHub repository[1].

IV. Discussion

By responding to interaction forces by changing the impedance and sending haptic cues on model updates, there is two-way communication between the human and the robot in the physical interaction. This communication allows intuitive mode switching between human and robot control without taking the human attention off the physical task, the way pressing a button would do. We expected this to make the interaction intuitive for users. Some of the users who tested our framework agreed. On the other hand, we also received the remark that switching at a button-press better disambiguates for the human when the robot is accepting the demonstration. A future study is necessary to compare and evaluate the intuitiveness of our implicit mode switching w.r.t. explicit switching, with a more representative group of subjects.

The comfort experienced with the mode switching and reactivation on model recognition varies with people's expectations and preferences. When and how fast (or slow) the robot responds currently depends on a number of preset variables. Ideally, these variables, or a more general model, defining the interaction and learning dynamics should be learned to match people's individual preferences.

In the current setting, the robot remembers every state it has seen from the moment we set it to start learning. At every point in its state space, it has stored a corresponding subsequent state it was shown to go to. For some participants, it was harder to remember the states and sequences they had taught the robot. Indeed, it may not always be desirable for the robot to remember all it has seen in the past. How and what to selectively forget is out of the scope of this study.

The disagreement detection in the presented implementation is based on a force threshold. Hence is would also trigger if a user is pushing in the direction the robot is going, e.g., to speed up the movement. This issue can be resolved by additionally considering the direction of the force. Similarly, for tasks requiring applying a force to an object or the environment, the disagreement detection will need to be modified to consider the difference to the expected force.

V. Conclusion and Future Work

With the presented framework, we showed how to smoothly transition between letting the robot execute the task and demonstrating alternative behaviors. By actively recognizing when the human is demonstrating, in contrast to only letting the human take the lead during execution [3], the task execution can be interactively corrected using the given human feedback.

For generalization, states can be parameterized with respect to objects' reference frames. That introduces the additional complexity of solving the possible ambiguity in the selection of the right frame for the given goal [9]. Furthermore, the linear trajectory assumption defined in Eq. (2) can be relaxed and a nonlinear trajectory or a dynamical system can be learned during the kinesthetic teaching interaction [4]. For this reason, we believe that the presented framework opens up many further directions of future work, as it allows online learning of (parameterized) high-level pHRC policies on real hardware with real (non-expert) users in the loop.

[1] https:\github.com/franzesegiovanni/franka_human_friendly_controllers

References

[1] R. Groten, D. Feth, R. L. Klatzky, and A. Peer, "The role of haptic feedback for the integration of intentions in shared task execution," *IEEE Transactions on Haptics*, vol. 6, no. 1, pp. 94–105, 2012.

[2] G. Pezzulo, L. Roche, and L. Saint-Bauzel, "Haptic communication optimises joint decisions and affords implicit confidence sharing," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[3] M. Khoramshahi and A. Billard, "A dynamical system approach for detection and reaction to human guidance in physical human–robot interaction," *Autonomous Robots*, vol. 44, no. 8, pp. 1411–1429, 2020.

[4] G. Franzese, A. Mészáros, L. Peternel, and J. Kober, "ILoSA : Interactive learning of stiffness and attractors," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[5] N. Hogan, "Impedance control: An approach to manipulation," in *American Control Conference*, 1984.

[6] C. Gaz, M. Cognetti, A. Oliva, P. R. Giordano, and A. De Luca, "Dynamic identification of the franka emika panda robot with retrieval of feasible parameters using penalty-based optimization," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4147–4154, 2019.

[7] A. Gams, A. J. Ijspeert, S. Schaal, and J. Lenarčič, "On-line learning and modulation of periodic movements with nonlinear dynamical systems," *Autonomous Robots*, vol. 27, no. 1, pp. 3–23, 2009.

[8] R. Hoque, A. Balakrishna, C. Putterman, M. Luo, D. S. Brown, D. Seita, B. Thananjeyan, E. Novoseller, and K. Goldberg, "LazyDAgger: Reducing context switching in interactive imitation learning," in *IEEE 17th International Conference on Automation Science and Engineering (CASE)*, 2021.

[9] G. Franzese, C. Celemin, and J. Kober, "Learning interactively to resolve ambiguity in reference frame selection," in *Conference on Robot Learning (CoRL)*, 2020.