# Dynamic Human-Robot Object Handover Learning from Human Feedback

Andras Kupcsik, David Hsu, Wee Sun Lee
School of Computing, National University of Singapore
{kupcsik,dyhsu,leews}@comp.nus.edu.sg

***Abstract*—Object handover is a basic and yet essential capability for robots interacting with humans in a broad range of applications, including caring for the elderly, assisting in a manufacturing workcell, etc. It appears deceptive simple, as humans perform object handover almost flawlessly. The success of humans, however, belies the complexity of object handover as a collaborative physical interaction between a robot and a human. This works addresses the problem of *dynamic* object handover, for example, when a robot hands over a water bottle to a marathon runner passing by a water station. We formulate the problem as context-aware policy search, which enables the robot to learn dynamic object handover through interaction with the human. One key challenge here is to learn the latent reward of the handover task under *noisy* human feedback. Experiments show that the robot can learn to hand over an object with very high success rate. It can also adapt to the dynamics of human motion naturally.**

## I. Introduction

In the future, robots will become efficient helpers of humans in everyday tasks. A key physical interaction channel between robots and humans is object transfer. In order for robots to become trustworthy helpers of humans, they need to master their handover skill in a wide variety of different situations. Robots today execute object transfer in a static manner by holding the object at a fixed location and wait for the human to take it. This is a very limited setting, which does not exhibit the dexterity, dynamics and generality of human to human object transfers. In this paper we address the problem of learning *dynamic*, human-like handover robot skills, while physically interacting with humans in different situations.

Humans are experts in transferring objects to one another without conscious planning of the hand and finger movements. However, programming an object handover skill for a robot is highly challenging. Firstly, a handover does not always happen in a static scenario, but in a more dynamic setting. In a static handover scenario the physical interaction during object transfer happens at a fixed location. On the other hand, in a dynamic handover situation when there is nonzero hand velocity during physical interaction, careful control of the finger and hand movements are required to ensure a robust transfer while maintaining low forces and jerks. Such a dynamic scenario is, for example, when a bottle of water is handed over to a runner. Other than the level of dynamics, a handover depends on many other factors, such as object type, human features (old – young, weak – strong, etc.), human preference, etc. Secondly, even for humans it is difficult to define what a "good" and successful handover is, and thus, it is challenging to find appropriate robot controllers for human-like handovers. Lastly, robots have a significantly lower amount of actuation and perception capability compared to humans. Moreover, the structure of the robot arm and hand is considerably different from that of humans. Thus, finding skills equivalent to that of humans (e.g. by demonstration) is not straightforward.

In this work, we address the problem of finding human-like dynamic handover skills for robots in a Policy Search [5] setting. Policy Search (PS) is a particularly successful Reinforcement Learning (RL) approach to learn skills for high DoF robots. PS algorithms learn the parameters of a task-appropriate control policy by maximizing the expected reward, which is the measure of how well the robot is performing. However, to find high quality robot skills, we need to identify the control policy to use and the reward function we aim to maximize. In this paper we discuss 1) a controller architecture for object handovers, 2) how to learn a latent reward function of the task from high level human feedback, and finally, 3) we demonstrate the learned dynamic handover skill in experiments. Overall, the learned handover is successful (fast and robust) over 95% of all experiments. Video footage of some typical experiments before and after the learning is available at http://youtu.be/QG-C9hW3YcU. More details can be found in [11].

## II. Related Work

**Human-robot handovers.** Human-robot handovers have been studied by many researchers in the past. One important question many papers targeted is how to generate human-like and legible trajectories for robots to communicate intent and adapt to human preferences [8, 1, 6]Another set of works analyze the complete handover process (pre-, during- and post-handover) and propose control algorithms for human-robot object transfers [12, 7]. In the above papers handovers were considered only in a static case, that is, the physical interaction happens at a fixed location.

An efficient object transfer also requires a robust control of grip forces. Research in human-human object handovers showed that human grip force is typically linear in the load force [2]. It was also observed that the giver overloads the taker, such that the taker has an excess of grip force, presumably to ensure a robust transfer.

**Policy Search with human feedback.** Robot skill learning by policy search has been highly successful in recent years [5]. Policy search algorithms perform an iterative update of

the parameter distribution of a control policy by maximizing the expected reward. In order to allow robot skills to adapt to different situations, contextual policy search has been proposed [9, 10]. In contextual policy search we learn a conditional policy that maps a context parameter to a controller parametrization. The context variable is task dependent, and it is used to fully describe a given situation. For example, in the dart throwing game the context may represent the target coordinates, while the parametrized control policy will generate the motion.

Policy Search algorithms assume that a reward function is given to guide the learning. However, in many learning tasks it is difficult to define an appropriate reward function. Daniel et al. use reward feedback of humans to learn manipulation skills for robot hands [4]. In our paper, as opposed to previous work, we mix both preference and absolute human feedback to learn the latent reward function to guide the learning.

## III. Problem Formulation and Approach

### A. The handover situation

In this work we are considering learning and generalizing robot dynamic object transfer skills over a wide variety of static and dynamic handover situations. We assume that both the human and the robot are aware of the handover situation and we focus on finding appropriate robot arm and hand controllers for human-like handover skills during physical interaction. We distinguish between static and dynamic handover skills based on the velocity of the human hand motion during the handover process. For example, static handovers can be considered when the human reaches out for an object, or when walking up to the robot and taking the object from it. In both of these situations the hand velocity is typically low during physical interaction, that is, until the robot releases the object. In a dynamic handover situation however, the hand velocity and acceleration of both the taker and/or the giver is nonzero and might not even have the same direction, for example, when handing over a leaflet to a pedestrian. This will generate forces and torques during the physical interaction that are not contributed to the object load. This makes the finger and arm control more challenging as opposed to the static handover scenario.

In our problem formulation we consider the robot to be the giver that aims to hold the object at a fixed location and we consider the human to be the taker. We hypothesize that the giver conditions her dynamic handover skill based on the situation, in our case the hand velocity of the taker, which is a measure of how dynamic a handover is. We assume that the handover controller of the robot giver, which ultimately encodes the finger and hand movements, is parametrized by a vector $\omega$. After choosing a controller parametrization, the deterministic control policy will generate control signals $u_t = \pi_\omega(x_t),\ t = 1,\ldots,T$, such as torques and grip forces given the state $x_t$ of the robot at time $t$.We denote the situation relevant parameters, in our case the hand velocity of the taker, with the *context* variable $s$. Our goal is to find a conditional policy $\pi(\omega|s)$ that maps contexts to controller parametrizations, such that the giver performance is optimal w.r.t. a reward function in a wide variety of dynamic handover situations.

### B. The control architecture

In our experiments we learn three different sets of parameters. As trajectory generator we solely use a linear trajectory generator that tracks the right hand of the human with constant speed. However, the hand tracking is only enabled in a certain distance between the robot and the human hand. The minimal distance is required to enable safe handover, and the maximal distance is required to avoid generating unfeasible trajectories for the robot.

For tracking the trajectory, we use Cartesian compliant control, which generates a contact force and torque $F = M\Delta\ddot{x} + D\Delta\dot{x} + P\Delta x$, where $\Delta x$ is the deviation from the reference trajectory. We choose $M$ to be the intertia of the robot at the current state and we set $D$ such that the closed loop control system is critically damped. For stiffness parameters $P$ we learn the translational stiffness and 1 parameter for all the rotational stiffness values.

In our experiments we assume a position controlled finger controller, where the grip force can only be controlled indirectly via the finger positions. For a certain object we first identify a finger position that exerts the minimal possible grip force. With the minimal finger position $p_{min}$ the robot barely holds the object and with some effort a human can take the bottle. The commanded finger position is given by $p_{finger} = (G_{bottle} - F_{human}) \times m + p_{min}$, where $G_{bottle}$ is the bottle weight and $F_{human}$ is the magnitude of the measured human force. The only parameter we learn is the slope $m$. A higher slope parameter will result in firmer grip.

We collect the parameters of the control architecture in $\omega$. That is, the minimal and maximal hand tracking distance, the stiffness parameters and the slope parameter for hand control. We hand-tune the initial policy $\pi(\omega|s)$, such that it performs well in static handover scenarios.

### C. Learning the handover skill

We consider finding the optimal conditional policy $\pi(\omega|s)$ as a contextual policy search problem [9, 5, 10]. In general, $\pi(\omega|s)$ is considered a stochastic policy to enable exploration during the learning process. The optimal policy maximizes the expected reward $R(\omega,s) \in \mathbb{R}$ over the joint distribution $\mu(s)\pi(\omega|s)$

$$\pi^* = \arg\max_\pi \int_s \mu(s) \int_\omega \pi(\omega|s) R(\omega,s), d\omega ds. \qquad (1)$$

Here $\mu(s)$ represents a distribution over possible situations, or contexts and $R(\omega,s)$ is the reward function that measures how good the controller parametrization $\omega$ is in context $s$. In practice, the policy is iteratively updated after collecting $N$ samples $\{\omega_i, s_i, R(\omega_i,s_i)\}_{i=1}^N$. Although human-human object transfers have been investigated in the past by many researchers [8, 1, 2, 12], it is still not clearly understood what is the underlying utility function humans try to maximize. This makes finding the true reward function $R(\omega,s)$ for human-robot object transfer a tedious task.

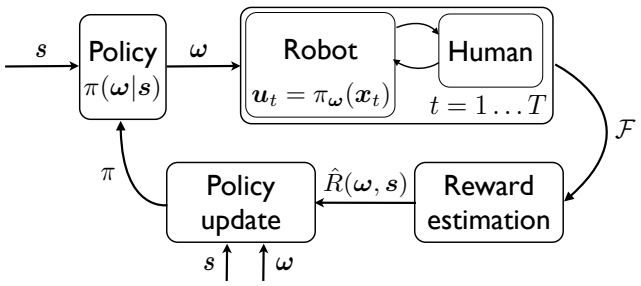


Fig. 1: The human-robot handover skill learning framework.

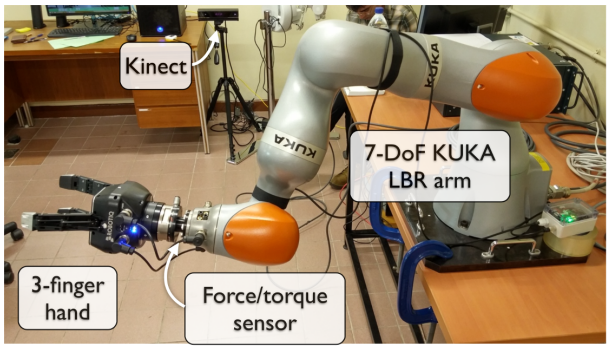


Fig. 2: The hardware setup. We use the KUKA LBR arm with the 3-finger Robotiq hand.

Instead, humans may directly give feedback to the robot on how well it is doing in a given situation. In this paper we distinguish between *absolute* and *preference* human feedback. Absolute feedback gives a direct assessment of the robot performance, while preference feedback gives a relative evaluation between two solutions. While the former has a higher information content, the latter is typically easier to assess by humans. Thus, our goal is to find a latent reward function $\hat{R}(\omega, s)$ from high level human feedback $\mathcal{F}$, which might refer to preference and/or absolute feedback.

We estimate the latent reward using a Bayesian approach. We build on the work of Chu and Ghahramani [3]. They use a Gaussian Process prior and propose a likelihood function that gives the probability of a human preferring a solution over another based on the latent reward difference. The posterior is then maximized in a convex optimization problem over the latent rewards. We extend the likelihood function of the above approach with a Gaussian model for absolute human feedback. This extension will leave the optimization problem for finding the latent reward convex. Learning a reward model for policy search has been considered in the literature for HRI settings [4], mixing preference and absolute assessment of the robot performance is novel. For learning the handover skill we consider the control and learning architecture depicted in Figure 1. As opposed to previous PS algorithms, we do not consider the sample rewards to be known, instead we estimate them as latent variables. In a toy experiment we observed that even sparse absolute feedback improves learning performance significantly and reduces the variance of the final performance.

## IV. Results

For the handover experiment we use the 7-DoF KUKA LBR arm (Fig 2). For the robot hand we use the Robotiq 3-finger hand. The fingers are position controlled, but the grip force can be indirectly adjusted by limiting the finger currents. In order for accurate measurement of external forces and torques, a wrist mounted force/torque sensor is installed.

### A. Experimental Setup

An experiment is executed as follows. First, a 1.5l water bottle is placed at a fixed location, which the robot is programmed to pick up. Subsequently, the robot moves the bottle to a predefined handover position. At this point we enable compliant arm control and we use a Kinect sensor (Fig 2) to track the hand of the human. Subsequently, the human moves toward the robot to take the bottle. While approaching the robot, we use the Kinect data to estimate the hand velocity $s$ of the human, which we assume to be constant during the approach. We only use data when the human is relatively far (above 1m) from the robot. After the context variable is estimated the robot sets its parameter by drawing a controller parametrization $\omega \sim \pi(\omega|s)$. Subsequently, the robot and the human make physical contact and the handover takes place. Finally, the human evaluates the robot performance (preference and/or absolute evaluation on a 1-10 scale, where 1 is worst 10 is best) and walks away such that the next experiment may begin.

### B. Learning Results

With the C-REPS learning algorithm we updated the policy after evaluating a set of 10 experiments. Initially we used 40 experiments to start the learning. After evaluating a batch of experiments we estimated the latent rewards from high level human feedback.

The expected latent reward of the initial policy is estimated to be around 6.8. Humans may give preference and/or absolute feedback in a 1-10 scale. However, we noticed that humans mostly gave absolute feedback for very good or bad solutions. This is expected as humans can confidently say if a handover skill feels close to that of a human, or if it does something unnatural (e.g., not releasing, or dropping the object). After evaluating the learning, roughly after 90 experiments, that is, after 6 policy updates the expected latent reward rose to the region of 8 over 5 experiments with low variance. But how did the policy and the experiments change with the learning?

After evaluating nearly 300 experiments with the robot, we had only one occasion that the robot dropped the bottle due to failed grasping. In less than 10 experiments the Kinect could not detect the human hand, which lead to a failed experiment. Overall, the learned policy provided a successful handover (fast and natural) in more than 95% of the experiments.

**Human preferences for static handovers.** For static handover tasks we observed that compliance parameters were less important for success, but a robust and quick finger control was always preferred and highly rated. A preferred solution

always maintained a low jerk and forces remained limited. Moreover, a successful handover happens relatively fast. In our experiments we observed that a high quality solution happens within 0.6 seconds and no faster than 0.4 seconds. Similar results have been reported in human-human object transfers experiments [2]. A disliked controller had low translational stiffness and a stiff finger control, resulting in the robot not releasing the object quick enough, which is considered a failure. These experiments typically lasted for 1 to 2 seconds until the bottle was released.

**Human preferences for dynamic handovers.** In dynamic handover situations contact forces and jerks were significantly higher compared to the static case. A typical preferred dynamic handover controller has lower stiffness parameters, especially in the direction of the motion during contact, and a more firm finger controller. We noticed that a physical contact time in a dynamic handover scenario is around $0.3 - 0.6$ sec. Based on the latent rewards, we noticed that there is a strong preference towards faster handovers, as opposed to the static case, where we did not observe such strong correlation in handovers within 0.6 seconds. Interestingly, we noticed that humans preferred stiffer finger controllers ($m$ is high) in dynamic handovers. We assume that this helps a safe and quick transfer of the object from giver to taker. In a dynamic handover situation vision might not provide a fast enough feedback about the handover situation, and thus, an excess of grip force would be necessary to ensure the robust transfer and to compensate for inaccurate position control.

Overall we can conclude that learning indeed improved the performance and adapted to human preferences. For static handovers a fast and smooth finger control was necessary for success, while in dynamic handover situation higher compliance and a firm finger control were preferred. Video footage of some typical experiments before and after the learning is available at http://youtu.be/QG-C9hW3YcU.

## V. Discussion and Summary

In this paper we investigated how robots can learn dynamic handover skills from humans while physically interacting with them. Our proposed learning algorithm not only improves initial performance, but is able to adapt to human preferences and can generalize to multiple situations. We demonstrated in robot experiments that the robot is able to successfully hand over a water bottle, even in highly dynamic situations.

However, our work is also limited, as we do not consider explicitly the adaptation of the human to robot performance, but rather assume the policy of the human to be fixed. On the other hand, in a realistic scenario humans (representing the taker here) may also adapt to the policy of the giver, e.g., when taking an object from a child or an elderly. Thus, future work will investigate how the policy of both the taker and the giver can be considered in a learning or planning setup.

## VI. Acknowledgement

This research was supported by NUS grant C-251-000-042-001 and A*STAR Industrial Robotics Program grant R-252-506-001-305.

## References

[1] Maya Cakmak, Siddhartha Srinivasa, Min Kyung Lee, Jodi Forlizzi, and Sara Kiesler. Human preferences for robot-human hand-over configurations. In IEEE, editor, *IEEE/RSJ IROS*, September 2011.

[2] Wesley P. Chan, Chris A. C. Parker, H. F. Machiel Van der Loos, and Elizabeth A. Croft. Grip forces and load forces in handovers: implications for designing human-robot handover controllers. In Holly A. Yanco, Aaron Steinfeld, Vanessa Evers, and Odest Chadwicke Jenkins, editors, *HRI*, pages 9–16. ACM, 2012.

[3] Wei Chu and Zoubin Ghahramani. Preference learning with gaussian processes. ICML '05, pages 137–144, New York, NY, USA, 2005. ACM.

[4] C. Daniel, M. Viering, J. Metz, O. Kroemer, and J. Peters. Active reward learning. In *Proceedings of Robotics: Science and Systems (R:SS)*, 2014.

[5] Marc Peter Deisenroth, Gerhard Neumann, and Jan Peters. A survey on policy search for robotics. *Foundations and Trends in Robotics*, 2(1-2):1–142, 2013.

[6] Anca Dragan and Siddhartha Srinivasa. Generating legible motion. In *Robotics: Science and Systems*, June 2013.

[7] Elena Corina Grigore, Kerstin Eder, Anthony G. Pipe, Chris Melhuish, and Ute Leonards. Joint action understanding improves robot-to-human object handover. In *Proceedings of IROS*, pages 4622–4629. IEEE, 2013.

[8] Markus Huber, Markus Rickert, Alois Knoll, Thomas Brandt, and Stefan Glasauer. Human-robot interaction in handing-over tasks. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pages 107–112, Munich, Germany, 2008.

[9] J. Kober, K. Mülling, O. Kroemer, C. H. Lampert, B. Schölkopf, and J. Peters. Movement Templates for Learning of Hitting and Batting. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2010.

[10] A. Kupcsik, M. P. Deisenroth, J. Peters, and G. Neumann. Data-Efficient Contextual Policy Search for Robot Movement Skills. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2013.

[11] Andras Kupcsik, David Hsu, and Wee Sun Lee. Learning dynamic human-robot object handover from human feedback. In *International Symposium of Robotics Research*, submitted.

[12] Kyle Strabala, Min Kyung Lee, Anca Dragan, Jodi Forlizzi, Siddhartha Srinivasa, Maya Cakmak, and Vincenzo Micelli. Towards seamless human-robot handovers. *Journal of Human-Robot Interaction*, 2013.