2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Human Grasp Classification for Reactive Human-to-Robot Handovers

Wei Yang*[1], Chris Paxton*[1], Maya Cakmak[1,2], and Dieter Fox[1,2]

***Abstract*— Transfer of objects between humans and robots is a critical capability for collaborative robots. Although there has been a recent surge of interest in human-robot handovers, most prior research focus on robot-to-human handovers. Further, work on the equally critical human-to-robot handovers often assumes humans can place the object in the robot's gripper. In this paper, we propose an approach for human-to-robot handovers in which the robot meets the human halfway, by classifying the human's grasp of the object and quickly planning a trajectory accordingly to take the object from the human's hand according to their intent. To do this, we collect a human grasp dataset which covers typical ways of holding objects with various hand shapes and poses, and learn a deep model on this dataset to classify the hand grasps into one of these categories. We present a planning and execution approach that takes the object from the human hand according to the detected grasp and hand position, and replans as necessary when the handover is interrupted. Through a systematic evaluation, we demonstrate that our system results in more fluent handovers versus two baselines. We also present findings from a user study ($N = 9$) demonstrating the effectiveness and usability of our approach with naive users in different scenarios. More information can be found at `http://wyang.me/handovers`.**

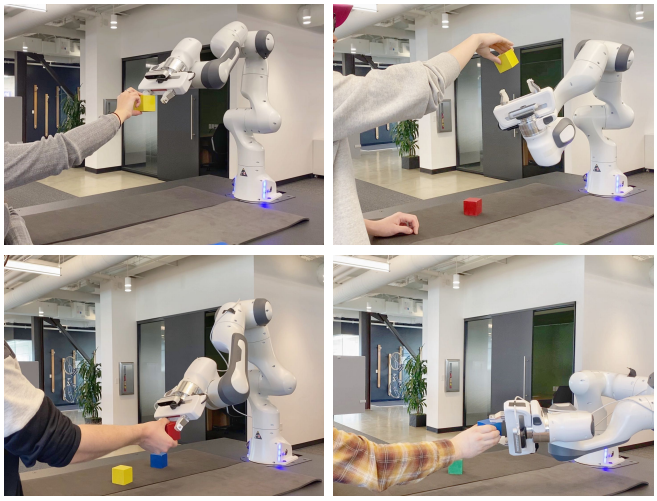

Fig. 1: Humans hand objects over in different ways. They can present the object on their palm or use a pinch grasp and present the object in different orientations. Our system can determine which grasp a human is using and adapt accordingly, enabling a reactive human-robot handover.

I. INTRODUCTION

Giving and taking objects to and from humans are fundamental capabilities for collaborative robots across applications from manufacturing to physical assistance in the home. A growing community of researchers in robotics has been studying the problem of enabling fluent human-robot handovers. Most work focuses on transfer of objects from the robot to the human, assuming the human can just place the object in the robot's gripper for the reverse. This approach is not feasible in scenarios where the human needs to pay attention to their task at hand, such as performing a surgery, or where the human has limited mobility and arm movement due to an impairment. Such scenarios require more reactive handovers that can adapt to the way that the human is presenting the object to the robot and meet them half way to take the object.

One of the key challenges in making human-to-robot handovers reactive is the reliable and continuous perception of the object and the human. One strategy is to estimate the human hand pose as well as the 6D object pose by borrowing off-the-shelf methods from the computer vision community. However, state-of-the-art methods for hand pose estimation [1], [2], [3] and object pose estimation [4] focus on only the hand or the objects independently. Although few recent methods jointly estimate hand and object poses while the hand is interacting with the object [5], [6], [7], their accuracy is limited when the object and the hand are occluded by each other.

In this paper, we propose to address the problem of perception for human-to-robot handovers by formulating it as a hand grasp classification problem. Specifically, we discretize the ways in which humans can hold small objects into several categories (Fig. 1) and collect a dataset to learn a deep model that classifies a given human hand holding an object into one of those grasp categories. The handover task is modeled as a Robust Logical-Dynamical System following on previous work [8], which generates motion plans that avoid contact between the gripper and the human hand given the human grasp classification. We compare our system with two baseline methods, one without inferring the human hand pose and the other relying on independent hand and object pose estimation, demonstrating higher success rate and time efficiency of our approach over the two baselines. We also present a user study ($N = 9$) demonstrating the effectiveness of our approach with naive users, both while they are attentive to the robot and while they are focused on a secondary task. Participants agreed that our system is collaborative, trustworthy and aware of the humans' actions.

The main contributions of this paper are: (1) hand-object interaction reasoning for handovers posed as a classification problem, via a dataset that covers a wide range of hand shapes and poses; (2) a system that adaptively plans robot grasps for taking the object from the human, so that the

* Equal Contribution
[1] NVIDIA, USA {weiy, cpaxton, dieterf}@nvidia.com
[2] University of Washington, USA mcakmak@cs.washington.edu

robot can respond to the human fluidly and naturally; and (3) experimental results demonstrating improvements over the baseline methods, and a user study validating our approach with naive users.

II. RELATED WORK

Human-robot handovers has recently become a popular topic within human-robot collaboration [9] across a multitude of application areas from collaborative manufacturing [10], [11], [12] to assistance in the home [13], [14], [15], [16]. A large majority of this work focuses on robot-to-human handovers in which the robot starts with an object in hand and transfers it to the human. A key challenge is choosing parameters of the robot's actions to optimize for a fluent handover. This includes the choice of object pose and robot's grasp on the object, taking into account user comfort [17], preferences based on subjective feedback [18], affordances and intended use of the objects after the handover [19], [20], [21], [22], [23], motion constraints of the human [13], social role of the human [24], and configuration of the object when being grasped before the handover [25]. Other work emphasizes parameters of the trajectory to reach the handover pose, exploring the approach angle [11], starting pose of trajectory in contrast to the handover pose [15], motion smoothness [26], object release time [27], estimated human wrist pose [28], [29], relative timing of handover phases [30], and ergonomic preferences of humans [31]. While some work focuses on offline computation of handover parameters, most recent work involves perception of the human to enable reactive handovers [32], [28], [33], [34].

A number of user studies have been conducted to validate different handover approaches and provide empirical evidence, such as people's preference among alternative ways of handing objects [15], [18], impact of robot behaviors such as gaze [35], [36], or difference between novice and experienced users [37]. Some work has explored human-human handovers to characterize movement properties [38], [39], [40], grip force patterns [41], use of social cues [42], or failure recovery strategies [43].

Although less frequent, some work has explored human-to-robot handovers, *i.e.*, how robots may take objects from humans [44], [17]. Pan et al. explored the problem of detecting handover intent by the human based on skeleton tracking data obtained in human-human handovers [45]. Other work enabled human-to-robot handovers via wearable sensing on the human [12]. Vogt et al. proposed to learn a controller to both give and receive objects from a single demonstration of handovers between two humans [46]. Most closely related to our research, Marturi et al. investigated the grasping of moving objects to enable human-to-robot handovers [47].

Given our focus on perception of humans to enable reactive handovers, prior works on human hand pose estimation are also highly relevant. Though 3D human hand pose estimation is being actively studied in computer vision, most of the existing work focuses on monocular RGB images [2], [3], which results in insufficient 3D localization for handovers. Some [48], [1], [48] use depth information for more precise hand pose estimation. However, they are mostly trained on data with a bare hand only due to the difficulty to collect data of hands interacting with objects, and are tend to fail in circumstances that the object is with close proximity with the hand. Instead of understanding hand pose and object pose in isolation, some recent work estimates hand-object manipulations [6], [5], [7]. While promising, these methods either trained on synthetic data [5] which requires to bridge the sim-to-real gap, or on sensor data within a close range which is not suitable for distant hand recognition for handovers [6], [7].

Our system for task execution is based on Robust Logical-Dynamical Systems [8], an approach for automatically creating reactive task plans for robots. The idea is to constantly identify the present logical state and reactively replan to handle uncertainty and changes in logical state, an approach that's been proven useful for dealing with partially-observable environments (*e.g.*, [49]). For our purposes, the task model can be thought of in a similar way to Behavior Trees [50], a method for representing complex tasks that has previously been shown useful for human-robot collaboration [51], [52].

III. HUMAN GRASP CLASSIFICATION

When the robot takes an object from humans, the motion should be adjusted according to the way that the object is grasped by the human hand. Otherwise the robot could behave in a nonintuitive way or even grasp human fingers. As illustrated in Fig. 2, our proposed handover framework addresses this issue by taking the point cloud centered around the human hand detected by the Azure Body Tracking SDK [54], and then estimating the hand grasp class based on how the block is grasped by the human hand. Our task model will then adaptively plan robot grasps.

In this section, we first define a discrete set of human grasps which describes the way that the object is grasped by the human hand for the task of handover. Then we present how we train a deep neural network to predict the human grasp categories based on the point cloud. Finally, we discuss how we adjust the orientation of robot grasps according to the human grasps.

A. Human Grasp Definition

Inspired by the study on the human grasp taxonomy [55], we discretize the common human grasps for the task of human-robot handover into seven categories, as shown in Fig. 3: If the hand is grasping a block, then the hand pose can be categorized as *on-open-palm, pinch-bottom, pinch-top, pinch-side*, or *lifting*. If the hand is not holding anything, it could be either *waiting* for the robot to handover an object or just doing nothing specific (*others*).

B. Human Grasp Dataset

To learn a model to classify the human grasps, we create a dataset which covers eight subjects with various hand shapes and poses by using an Azure Kinect RGBD camera. Specifically, an example image of a hand grasp is demonstrated to

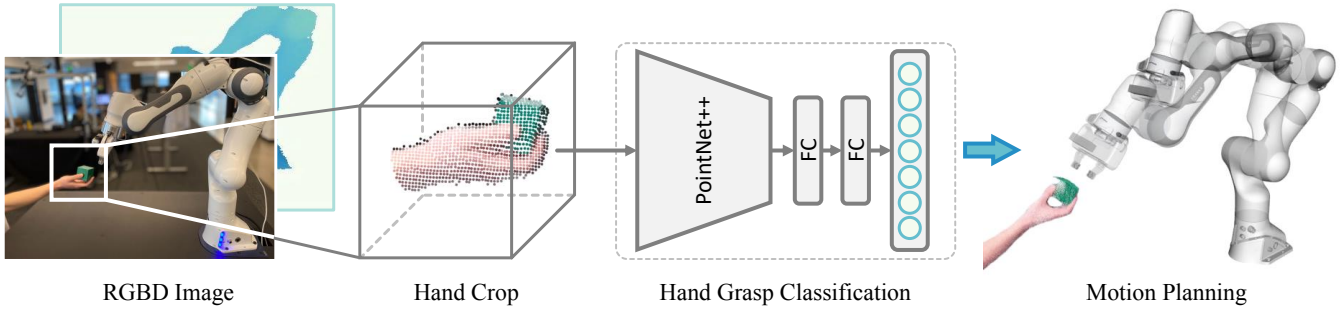

Fig. 2: An overview of our handover framework. The framework takes the point cloud centered around the hand detection, and then uses a model inspired by PointNet++ [53] to classify it as one of seven grasp types which cover various ways objects tend to be grasped by the human user. Out task model will then plan the robot grasps adaptively.

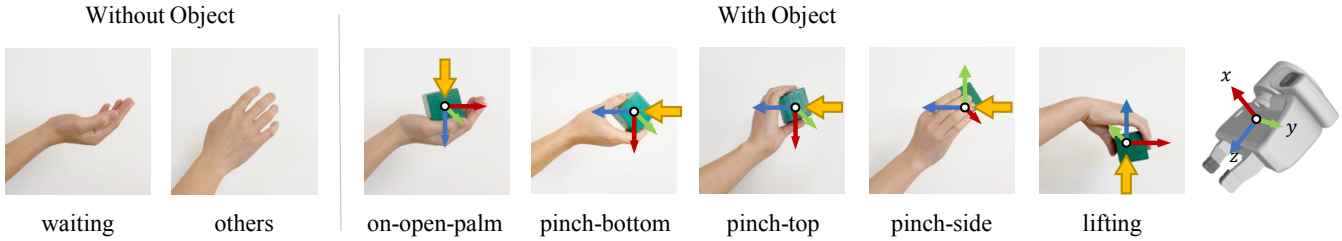

Fig. 3: Five human grasp types with two empty hand types which cover various ways objects tend to be grasped by the human user. These are associated with different robot canonical grasp directions in order to minimize human's efforts during handovers (illustrated by the coordinate system and the yellow arrow). Best viewed in color.

the subject. We record the subject performing similar poses for 20 to 60 seconds. The whole sequence of images are therefore labeled as the corresponding human grasp category. During the recording, the subject can move his/her body and hand to different position to diversify the camera viewpoints. We record both left and right hands for each subject. In total, our dataset includes around 150K images.

C. Human Grasp Classification Model

Instead of learning deep features with ConvNets on depth images, we adopt the recently developed PointNet++ [53] on point clouds for human grasp classification due to its efficiency and its success on many robotics applications such as markerless teleoperation system [48] and grasps generation [56]. Our backbone network consists of four set-abstraction layers to learn point features and a three-layer perceptron with batch normalization, ReLu and Dropout for global feature learning and human grasp classification. Given a point cloud cropped around the hand, the network classifies it into one of the defined grasp categories, which would be used for further robot grasp planning.

D. Canonical Robot Grasp Directions

We associate each human grasp type with a canonical robot grasp direction in order to minimize human's effort during the human-to-robot handovers. As shown in Fig. 3, the coordinates denote the canonical robot grasp frames in the camera frame. The motivation is to reduce the chance for the robot to grab human's hand while keeping its motion and trajectory as natural and smooth as possible.

IV. TASK MODEL

Our task model is based on Robust Logical-Dynamical Systems [8]. It represents tasks as a list of reactively-executed operators o with certain properties. Each operator is a tuple $o = \{L_P, L_R, L_E, \pi\}$, where L_P is a set of logical preconditions on entering o, L_R is a set of run conditions that must hold while execution of o is ongoing, and L_E is the set of logical effects that will be true. The operator is associated with a policy π which will generate the necessary controls to achieve effects L_E. The policy and predicates can be learned from data [57], but in our case they are specified manually. Given a plan, we choose the highest-priority operator whose preconditions are met, checking conditions at 10 Hz so we can quickly respond to changes.

Fig. 4 gives an overview of the different steps in our final task plan. The system has to adapt to different possible grasps, reactively choosing the correct way to approach the human user and take the object from them. Until it gets a stable estimate of how the human wants to present the block, it stays in a "home" position and waits.

Instead of just using reactive local planning, we found we needed to plan and make intelligent decisions based on a large number of possible grasps in order to find the one that would be the most natural to the human user, as discussed in prior work [51], [52]. We describe the extra predicates and operators involved in this reactive planning process below. Table I shows the task plan, in order of descending priority.

Wait for human. We compute several predicates determining how the robot should interact with the hand: `stable`, `hand_over_table`, and `hand_has_obj`, and `too_close_to_hand`. The `hand_over_table` predicate corresponds to whether or not these observations are in

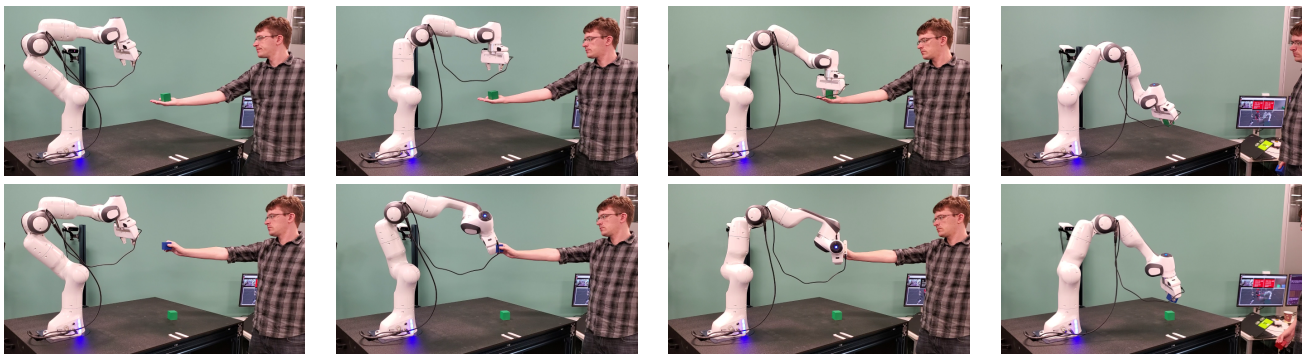

Fig. 4: Examples of task plan execution with the given system. The important operators are (1) Waiting at the "home" position for the human to enter the workspace and for position estimates to stabilize, (2) choosing a safe grasp plan, (3) moving to the grasp position and taking the block from the human, and (4) dropping the object on the table. These steps can be interrupted by the human. Descriptions of policies and preconditions are in Sec. IV.

Operator	Preconditions
Open gripper	`has_obj` $\wedge$ `gripper_fully_closed`
Wait for human	$\neg$(`stable` $\wedge$ `hand_over_table` $\wedge$ `hand_has_obj`)
Avoid human	`too_close_to_hand` $\wedge$ $\neg$`in_approach_region`
Drop object	`at_drop_position` $\wedge$ `has_obj`
Go to drop	`has_obj`
Grasp object	`in_approach_region` $\wedge$ `has_goal` $\wedge$ `is_goal_valid`
Follow plan	`has_goal` $\wedge$ `is_goal_valid`
Find plan	`has_feasible_goals`
Find feasible goals	`stable` $\wedge$ `hand_over_table` $\wedge$ `hand_has_obj`

TABLE I: Operators and corresponding preconditions L_P for task execution and reactive execution. Operators are listed in descending order of priority; if all the preconditions are true, we execute the associated operator regardless of what the previously executed operator was.

a specified volume over the table depicted in Fig. 4, and `stable()` is true if the hand is not moving and the hand has been observed for at least 5 timesteps (0.5 seconds). We defined this based on the velocity:

$$\texttt{stable()} = \|x_{t\ \ 1} - x_t\|_2 < \lambda,$$

with position x and time t, for a threshold λ. The robot will wait at the home position if these conditions are not true.

Avoid human. If `too_close_to_hand()` is true for either hand and the robot is not in the approach region corresponding to a particular grasp, the robot will attempt to avoid the hand and will move back to the home positions. We define `too_close_to_hand()` to be true if the Euclidean distance between the end effector and the hand is less than 20 cm.

Find feasible goals. In order to ensure that the robot's motions are safe, instead of the purely reactive policies used in prior work [8], we plan whole trajectories for execution. If `stable()`, `hand_over_table()`, and `hand_has_obj()`, then the robot will attempt to take the object from the hand, using the canonical grasp pose shown in Fig. 3.

In order to find a valid trajectory ξ, the robot must first find a valid grasp pose, so we add the `has_goal` and `is_goal_valid` predicates. If either of these is false, we search for a reasonable goal pose.

The planner will create a list of goal pose candidates and associated standoff positions. There are ten options, at rotations of $\theta_y \in \{-\pi/4, -\pi/8, 0, \pi/8, \pi/4\}$ around the y-axis in Fig. 3, and $\theta_z \in \{0, \pi\}$ around the z-axis. Any feasible grasp poses (*i.e.*, grasp poses with a corresponding inverse kinematics solution). Both grasp and standoff position must be collision-free and have a valid IK solution in order to be considered feasible goal options. We also add a constraint that the robot should never occlude its view of the object when determining if states are valid.

Find plan. If the planner has a list of goal options, it will then sort them according to their distance from the current joint configuration and attempt to find a motion plan to the standoff position using RRT-Connect [58]. If the system can both find a grasp pose and a motion plan, the robot will execute a sub-policy to follow this motion plan.

However, a human might move their hand or change how they are holding an object in their hand. A goal is only considered valid (as per the `is_goal_valid` predicate) if it has an associated motion plan, and if the object has not moved within some threshold of where it was first observed. If the object moves too much, the robot must stop, and the task model will instantly transition back to finding a new grasp.

Grasp Object. Once a motion plan has completed, the robot should be at a standoff pose and have an associated goal pose – the expected position of the object in the human hand. These two poses define an *approach region* – a conical volume within which the robot can move to approach the object, as described in our prior work [8]. Once the gripper closes, if the robot is at its goal pose, the `has_obj` predicate is set to true. The grasp operator may occlude the object, so we execute this as our only blocking, open-loop action.

Open Gripper. If `has_obj` is true, indicating that the robot believes it is holding an object, we may still be wrong because the object moved or the pose estimate was off. We

add a `gripper_fully_closed` predicate, saying that the gripper closed all the way. If both conditions are true, we set `has_obj` to false, and the robot will revert to a different state.

Move to drop and `Drop object.` The drop position is a single joint-space position; our robot will find a safe, collision-free motion plan. If it is at the drop position, it will open the gripper and put the object on the table.

V. SYSTEMATIC EVALUATION

In this section, we perform a systematic evaluation, where we compare our method to several baseline approaches with multiple metrics. We also report the performance of our human grasp classification model.

A. Experimental Setup

We performed a systematic of the entire system, including the classification model described in Sec. III and the task model described in Sec. V-B, on a range of different hand positions and the grasps shown in Fig. 3. We used two different Franka Panda robots, mounted on identical tables in different locations, as shown in Fig. 4. A human user handed four colored blocks over to the robot, one at a time. During the systematic evaluation, we tested each of the three approaches for determining which grasp pose to use for taking an object from the human:

Simple Baseline: waits until it sees the block in a human hand and takes it from the hand, using a fixed grasp orientation. The human hand is detected via an off-the-shelf system; in our case, the Microsoft Azure body tracker [54].

Hand Pose Estimation: A state estimation-based version of the system, in which we use the human hand pose from the Azure body tracker to infer grasp direction.

Ours: The proposed system, classifying human grasps based on depth information as described in Sec. III.

All variants executed the same task model, as described in Sec. IV. The order in which these three test cases were provided was randomized. Users used their right hand to present blocks to the robot.

B. Evaluation Metrics

We evaluate the system performance with a set of metrics computed during trials. These were computed automatically and logged while users were performing the task.

Planning Success Rate: the number of times the `follow_plan` operator was able to execute successfully, bringing the robot to its standoff pose, and measures certainty both of the human and the system.

Grasp Success Rate: how often the robot was able to successfully take the object from the human, versus the total number of grasp attempts that it made.

Action execution Time: tracks the amount of time it took to execute a single planned trajectory, grasp a block, and place it on the table. This is higher if the robot must take a longer path to grasp the block from the human.

Total Execution Time: the amount of time it took to execute all planned paths, including replanning because the human moved or because of the changing of the way of grasp.

Trial Duration: Time since human hand was first detected until the trial was complete.

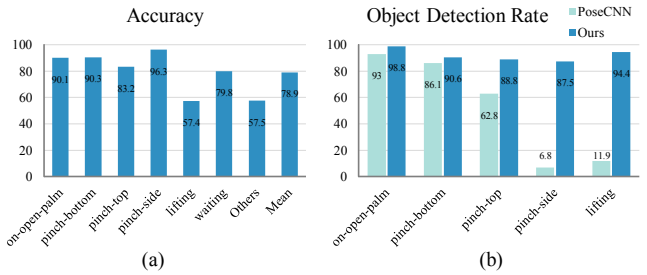

Fig. 5: (a) The accuracy of the human hand grasp classification. (b) The comparison of the object miss-detection rate between our hand states classification and PoseCNN [4]. In many cases the hand occludes the object, meaning that it is very difficult to get an accurate pose estimate.

C. Results

Table II shows results on each of our main metrics during the systematic evaluation. Our method consistently improves the success rate and reduces the total execution time and the trial duration compared with the other two baseline methods, which proves the efficacy and the reliability of our method.

The only exception is the *Action Execution Time*, where the *simple baseline* is often faster. This is because the simple baseline does not plan as adaptively as the others; it would not try to attempt an unusual grasp. This means that time from a successful approach to dropping the object is, on average, notably lower.

Evaluation of Human Grasp Classification: We evaluate our hand grasp classification model on a validation set collected with a subject which is unseen during the training procedure. The classification accuracy is reported in Fig 5 (a), which demonstrates the good generalization ability of our model on unseen subjects.

In addition, we conduct an experiment to evaluate the detection rate, *i.e.*, whether there is an object in the hand, to give us an insight on how robust the handover system is against the occlusion. We compare the detection rate of our hand grasp classification model (with/without object) to that of a state-of-the-art object detection method [4]. The result is reported in Fig. 5(b). We can see that our human grasp model achieves higher detection rate and is more robust compared with [4] especially when heavy occlusion occurs (*e.g.*, 87.5% vs. 6.8% for *pinch-side* and 94.4% vs. 11.9% for *lifting*).

VI. USER STUDY

We also performed a user study in order to determine if our system allowed for fluid human-robot collaboration. We recruited nine users, ages 20 to 36. Of these, two were female and seven were male. The average age was 30.44 ± 4.74 years. The study consisted of three rounds:

Freeform: first, users were given four blocks and instructed to stand in front of the table and hand the blocks over to the robot one at a time. They were instructed that the

TABLE II: Results for handover performance on our quantitative metrics. Planning success rate indicates how often the system needed to replan its approach, versus grasp success rate as the number of times the system successfully took the object.

	Planning Success Rate	Grasp Success Rate	Action Execution Time (s)	Total Execution Time (s)	Trial Duration (s)
Simple Baseline	42.1%	66.7%	11.37	20.93	21.59
Hand Pose Estimation	29.6%	80.0%	15.10	36.34	36.46
Ours	64.3%	100%	13.20	17.34	18.31

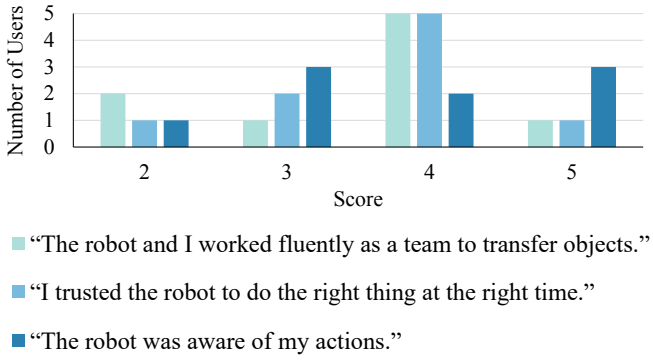

Fig. 6: Results form the Likert scale questionaire given to users during the usability study. Users thought they worked together fluently with the robot, although there were several issues. Users were asked questions at the end of the study.

robot would only take blocks if their hand is still, but they could hold the blocks any way they liked.

Attentive: Next we demonstrated the set of five human grasps shown in Fig. 3: *pinch-top, pinch-bottom, pinch-side, lifting*, and *on-open-palm*. We then told the participants to hand over four blocks again. We encouraged them to try the predefined hand grasps, but they were able to use any others.

Distracted: Finally, we tested user performance in the presence of a distraction. Users watched a music video on YouTube [1] and counted the number of faces that appeared, while handing over all four blocks to the robot.

In addition to the metrics described in Sec. V-B, we also counted the following statistics during the user study: (a) number of times robot gripper contacted human fingers, (b) number of times users changed the grasp they were using, and (c) number of times they changed their hand position. After each trial, participants were asked to describe any problems they experienced while handing the blocks to the robot. After all three trials were done, we asked them to fill a Likert scale questionnaire and explain their answers.

A. Results

Fig. 6 shows the results of the questionnaire given to our study participants. There was a range of responses, but users said that they worked fluently with the robot and trusted it to do the right thing, although they noted several common issues when asked for feedback. They also believed that the robot was aware of their actions.

We also computed our quantitative metrics on user data, as seen in Table III. Approaches and grasps were less successful when users were distracted, but times are similar. Users counted an average of 12.88 ± 3.48 faces in the music video,

[1] https://youtu.be/4jd6dNrJRh4

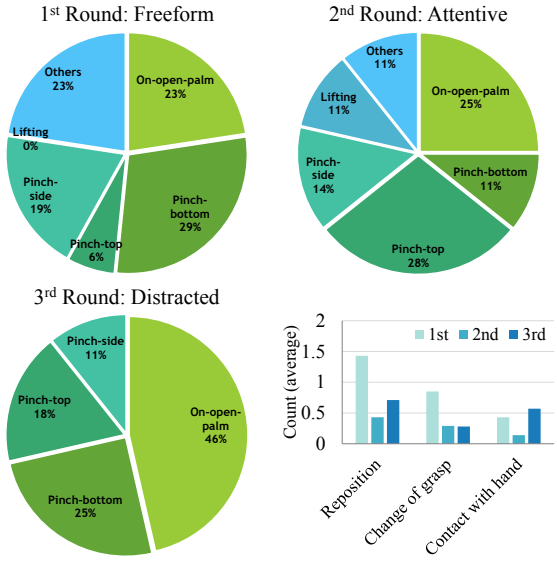

Fig. 7: Coverage of human grasps during the usability study. *Others* denotes the user grasps that are not included in our human grasp dataset. Our human grasp dataset covers 77% of the user grasps even before the users were informed the ways of grasp included in our system (see *Freeform*). We also report the average times that the users repositioned their hand, changed the way of grasp, and were contacted by the gripper.

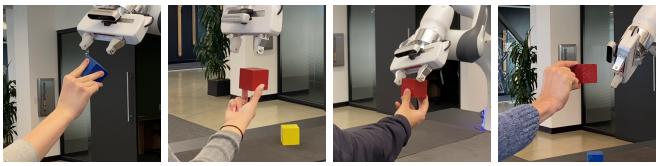

Fig. 8: Outliers grasps which do not appear in our training dataset, and are examples of the types of grasps where our system showed higher uncertainty, leading to slightly worse handover performance.

when the correct number was 13. This implies many of them felt a certain level of confidence of the handover system and were paying a good amount of attention to the video.

B. Discussion

First, we report the coverage of different ways of grasp during the usability study in Fig. 7. In general, our definition of human grasps covers 77% of the user grasps even before they know the ways of grasps defined in our system (see *Freeform* in Fig. 7, which proves a good coverage of our human grasp design. While our system can deal with most of the unseen human grasps, they tend to lead to higher uncertainty and sometimes would cause the robot to backoff and replan. Some of these unseen grasps are shown in Fig. 8. This suggests directions for future research; ideally we would be able to handle a wider range of grasps that a human might want to use.

TABLE III: Quantitative results from the user study. Users were able to complete tasks quickly even when they were distracted and had to concentrate on a different scenario.

	Planning Success Rate	Grasp Success Rate	Action Execution Time (s)	Total Execution Time (s)	Trial Duration (s)
Freeform	32.7%	67.3%	13.21	25.99	26.92
Attentive	40.0%	90.0%	14.85	23.84	24.75
Distracted	29.8%	67.9%	11.08	26.08	27.02
Overall	33.6%	73.6%	13.05	25.31	26.24

We also report the average number of times that users repositioned their hand, changed their way of grasp, and were contacted by the gripper in Fig. 7. During the final, distracted test, users had to reposition or change their ways of grasp more often compared with the first two rounds. Several complained about their fingers being pinched, or saw the robot fail to grasp objects. One specifically "chose to use the palm-facing-up hand pose" to minimize risk of failure; another "had to look at the robot every 10s or so."

This issue came up for several reasons. One of the most important is that many of the grasp poses are very hard for the robot to reach – they are at the edges of its configuration space, and may require some complex motions to get there. In the future, we should strive to make the whole system more legible, indicating which blocks the robot wants to move to and how it wants to get there [59].

In general, our users quickly noticed that the robot was trying to grab the block in an unobtrusive way as possible. They also noted a slight inaccuracy during the robot's grasps and approaches, but this doesn't seem to have been a major issue. One said about the robot "it may require my assistance in slightly moving towards where it expected the goal to be." In other words, even though the robot's grasp pose might be slightly off due to an inaccurate hand pose or occluded object, the robot and human together were reliably able to execute the handoff.

From a usability perspective, our trajectories weren't always totally legible. The underlying motion planner used in Sec. IV was based on RRT-connect [58], and sometimes it would make surprising choices to reach grasp positions. For example, one user said after the first experiment, "There was one of the trajectories that had a small detour." Another said "Holding the block in such a way that the robot needed to rotate its grasp 90 degrees seemed to cause problems." This is because the robot would often be unclear about what the position and orientation of the hand was, and would end up being uncertain. In the end, the motion plans could be hard to interpret – "I was still unsure of what the actual behavior will be like."

Users did notice that after the second round of experiments, when they were shown how to grasp the objects, that the system was more reliable and easier to work with. One users said "the estimate of the object/hand position seemed more precise" after they were taught these grasps.

During the final, distracted test, users were more nervous. Several complained about their fingers being pinched, or saw the robot fail to grasp objects. One specifically "chose to use the palm-facing-up hand pose" to minimize risk of failure; another "had to look at the robot every 10s or so."

This issue came up for several reasons. First, many of the grasp poses are very hard for the robot to reach – they are at the edges of its configuration space, and may require some complex motions to get there. Second, the users could also reposition their hands unintentionally while being distracted. In the future, we should strive to make the whole system more legible, indicating which blocks the robot wants to move to and how it wants to get there [59].

VII. CONCLUSIONS

We described a system for enabling fluid human-robot handovers via classifying different types of grasp. In the future, we will make the planning system more flexible and support more grasp types. We believe the same approach could also be applied to many other types of human-robot collaboration. The main limitation of our approach is that it applies only to a single set of grasp types, so additionally we plan to learn the correct grasp poses for different grasp types from data instead of using manually-specified rules. Based on user feedback, we also plan to make robot motions more legible and friendly.

REFERENCES

[1] L. Ge, Z. Ren, and J. Yuan, "Point-to-point regression pointnet for 3d hand pose estimation," in *ECCV*, 2018.
[2] U. Iqbal, P. Molchanov, T. Breuel Juergen Gall, and J. Kautz, "Hand pose estimation via latent 2.5 d heatmap regression," in *ECCV*, 2018.
[3] L. Ge, Z. Ren, Y. Li, Z. Xue, Y. Wang, J. Cai, and J. Yuan, "3d hand shape and pose estimation from a single rgb image," in *CVPR*, 2019.
[4] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "PoseCNN: A convolutional neural network for 6d object pose estimation in cluttered scenes," in *Robotics: Science and Systems*, 2017.
[5] Y. Hasson, G. Varol, D. Tzionas, I. Kalevatykh, M. J. Black, I. Laptev, and C. Schmid, "Learning joint reconstruction of hands and manipulated objects," in *CVPR*, 2019.
[6] C. Zimmermann, D. Ceylan, J. Yang, B. Russell, M. Argus, and T. Brox, "FreiHAND: A dataset for markerless capture of hand pose and shape from single rgb images," in *ICCV*, 2019.
[7] S. Hampali, M. Rad, M. Oberweger, and V. Lepetit, "Honnotate: A method for 3d annotation of hand and object poses," in *CVPR*, 2020.
[8] C. Paxton, N. Ratliff, C. Eppner, and D. Fox, "Representing robot task plans as robust logical-dynamical systems," in *IROS*, 2019.
[9] A. Bauer, D. Wollherr, and M. Buss, "Human–robot collaboration: a survey," *International Journal of Humanoid Robotics*, 2008.
[10] A. Koene, A. Remazeilles, M. Prada, A. Garzo, M. Puerto, S. Endo, and A. M. Wing, "Relative importance of spatial and temporal precision for user satisfaction in human-robot object handover interactions," in *Third International Symposium on New Frontiers in Human-Robot Interaction*, 2014.
[11] V. V. Unhelkar, H. C. Siu, and J. A. Shah, "Comparative performance of human and mobile robotic assistants in collaborative fetch-and-deliver tasks," in *HRI*. IEEE, 2014.
[12] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling object hand-over in human–robot collaboration via natural wearable sensing," *IEEE Transactions on Human-Machine Systems*, 2018.
[13] J. Mainprice, M. Gharbi, T. Siméon, and R. Alami, "Sharing effort in planning human-robot handover tasks," in *RO-MAN*. IEEE, 2012.

[14] C.-M. Huang, M. Cakmak, and B. Mutlu, "Adaptive coordination strategies for human-robot handovers," in *Robotics: Science and Systems*, 2015.

[15] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human handovers," in *HRI*, 2011.

[16] E. C. Grigore, K. Eder, A. G. Pipe, C. Melhuish, and U. Leonards, "Joint action understanding improves robot-to-human object handover," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013.

[17] J. Aleotti, V. Micelli, and S. Caselli, "Comfortable robot to human object hand-over," in *RO-MAN*. IEEE, 2012.

[18] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler, "Human preferences for robot-human hand-over configurations," in *IROS*. IEEE, 2011.

[19] J. Aleotti, V. Micelli, and S. Caselli, "An affordance sensitive system for robot to human object handover," *International Journal of Social Robotics*, 2014.

[20] W. P. Chan, Y. Kakiuchi, K. Okada, and M. Inaba, "Determining proper grasp configurations for handovers through observation of object movement patterns and inter-object interactions during usage," in *IROS*. IEEE, 2014.

[21] A. Bestick, R. Bajcsy, and A. D. Dragan, "Implicitly assisting humans to choose good grasps in robot to human handovers," in *International Symposium on Experimental Robotics*. Springer, 2016.

[22] W. P. Chan, M. K. Pan, E. A. Croft, and M. Inaba, "An affordance and distance minimization based method for computing object orientations for robot human handovers," *International Journal of Social Robotics*, 2019.

[23] F. Cini, V. Ortenzi, P. Corke, and M. Controzzi, "On the choice of grasp type and location when handing over an object," *Science Robotics*, 2019.

[24] S. Kato, N. Yamanobe, G. Venture, E. Yoshida, and G. Ganesh, "The where of handovers by humans: Effect of partner characteristics, distance and visual feedback," *PloS one*, vol. 14, no. 6, 2019.

[25] P. Ardón, È. Pairet, S. Ramamoorthy, and K. S. Lohan, "Towards robust grasps: Using the environment semantics for robotic object affordances," in *Proceedings of the AAAI Fall Symposium on Reasoning and Learning in Real-World Systems for Long-Term Autonomy*, 2018.

[26] E. De Momi, L. Kranendonk, M. Valenti, N. Enayati, and G. Ferrigno, "A neural network-based approach for trajectory planning in robot–human handover tasks," *Frontiers in Robotics and AI*, 2016.

[27] Z. Han and H. Yanco, "The effects of proactive release behaviors during human-robot handovers," in *HRI*. IEEE, 2019.

[28] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, 2017.

[29] A. Sidiropoulos, E. Psomopoulou, and Z. Doulgeri, "A human inspired handover policy using gaussian mixture models and haptic cues," *Autonomous Robots*, 2019.

[30] A. Kshirsagar, H. Kress-Gazit, and G. Hoffman, "Specifying and synthesizing human-robot handovers," in *IROS*. IEEE, 2019.

[31] S. Parastegari, B. Abbasi, E. Noohi, and M. Zefran, "Modeling human reaching phase in human-human object handover with application in robot-human handover," in *IROS*. IEEE, 2017.

[32] L. Peternel, W. Kim, J. Babič, and A. Ajoudani, "Towards ergonomic control of human-robot co-manipulation and handover," in *IROS*. IEEE, 2017.

[33] A. Kupcsik, D. Hsu, and W. S. Lee, "Learning dynamic robot-to-human object handover from human feedback," in *Robotics research*. Springer, 2018.

[34] T. Zhou and J. P. Wachs, "Early prediction for physical human robot collaboration in the operating room," *Autonomous Robots*, 2018.

[35] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *HRI*, 2014.

[36] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft, "Meet me where I'm gazing: how shared attention gaze affects human-robot handover timing," in *HRI*, 2014.

[37] S. M. zu Borgsen, J. Bernotat, and S. Wachsmuth, "Hand in hand with robots: differences between experienced and naive users in human-robot handover scenarios," in *International Conference on Social Robotics*. Springer, 2017.

[38] C. Becchio, L. Sartori, and U. Castiello, "Toward you: The social side of actions," *Current Directions in Psychological Science*, 2010.

[39] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, "Human-robot interaction in handing-over tasks," in *RO-MAN*. IEEE, 2008.

[40] S. Shibata, B. M. Sahbi, K. Tanaka, and A. Shimizu, "An analysis of the process of handing over an object and its application to robot motions," in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*. IEEE, 1997.

[41] W. P. Chan, C. A. Parker, H. M. Van der Loos, and E. A. Croft, "Grip forces and load forces in handovers: implications for designing human-robot handover controllers," in *HRI*, 2012.

[42] C. Shi, M. Shiomi, C. Smith, T. Kanda, and H. Ishiguro, "A model of distributional handing interaction for a mobile robot." in *Robotics: science and systems*, 2013.

[43] S. Parastegari, E. Noohi, B. Abbasi, and M. Žefran, "Failure recovery in robot–human object handover," *IEEE Transactions on Robotics*, vol. 34, no. 3, 2018.

[44] A. Edsinger and C. C. Kemp, "Human-robot interaction for cooperative manipulation: Handing objects to one another," in *RO-MAN*. IEEE, 2007.

[45] M. K. Pan, V. Skjervøy, W. P. Chan, M. Inaba, and E. A. Croft, "Automated detection of handovers using kinematic features," *IJRR*, 2017.

[46] D. Vogt, S. Stepputtis, B. Jung, and H. B. Amor, "One-shot learning of human–robot handovers with triadic interaction meshes," *Autonomous Robots*, 2018.

[47] N. Marturi, M. Kopicki, A. Rastegarpanah, V. Rajasekaran, M. Adjigble, R. Stolkin, A. Leonardis, and Y. Bekiroglu, "Dynamic grasp and trajectory planning for moving objects," *Autonomous Robots*, 2019.

[48] A. Handa, K. Van Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and D. Fox, "Dexpilot: Vision based teleoperation of dexterous robotic hand-arm system," in *ICRA*, 2020, to appear.

[49] C. R. Garrett, C. Paxton, T. Lozano-Pérez, L. P. Kaelbling, and D. Fox, "Online replanning in belief space for partially observable task and motion problems," in *ICRA*, 2019.

[50] M. Colledanchise and P. Ögren, "Behavior trees in robotics and AI: An introduction," 2018.

[51] C. Paxton, A. Hundt, F. Jonathan, K. Guerin, and G. D. Hager, "Costar: Instructing collaborative robots with behavior trees and vision," in *ICRA*. IEEE, 2017.

[52] C. Paxton, F. Jonathan, A. Hundt, B. Mutlu, and G. D. Hager, "Evaluating methods for end-user creation of robot task plans," in *IROS*. IEEE, 2018.

[53] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *NeurIPS*, 2017.

[54] "Azure kinect dk," https://docs.microsoft.com/en-us/azure/kinect-dk/, accessed: 2020-02-28.

[55] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on Human-Machine Systems*, 2015.

[56] A. Murali, A. Mousavian, C. Eppner, C.Paxton, and D. Fox, "6-dof grasping for target-driven object manipulation in clutter," in *ICRA*, 2020, to appear.

[57] K. Kase, C. Paxton, H. Mazhar, T. Ogata, and D. Fox, "Transferable task execution from pixels through deep planning domain learning," in *ICRA*. IEEE, 2020, to appear.

[58] J. J. Kuffner and S. M. LaValle, "RRT-connect: An efficient approach to single-query path planning," in *ICRA*, vol. 2. IEEE, 2000.

[59] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *HRI*. IEEE, 2013.