Proactive, Incremental Learning of Gesture-Action Associations For Human-Robot Collaboration

Dadhichi Shukla, Özgür Erkent, and Justus Piater

***Abstract*— Identifying an object of interest, grasping it, and handing it over are key capabilities of collaborative robots. In this context we propose a fast, supervised learning framework for learning associations between human hand gestures and the intended robotic manipulation actions. This framework enables the robot to learn associations on the fly while performing a task with the user. We consider a domestic scenario of assembling a kid's table where the role of the robot is to assist the user. To facilitate the collaboration we incorporate the robot's gaze into the framework. The proposed approach is evaluated in simulation as well as in a real environment. We study the effect of accurate gesture detection on the number of interactions required to complete the task. Moreover, our quantitative analysis shows how purposeful gaze can significantly reduce the amount of time required to achieve the goal.**

I. Introduction

The domestic scenario of a table assembly as shown in Fig. 1 is composed of sub-tasks like identifying the targeted object, grasping the object, handing the object to the user, or placing the object within reach of the user. In such scenarios different users will have different preferences for the sequence of actions the robot has to perform. The robot must account for these preferences and must address the operational flexibility expected of natural human-robot interaction (HRI). To establish a robot system with these abilities we propose a fast, supervised Proactive Incremental Learning (PIL) framework. The PIL framework is designed to learn the associations between hand gestures and the robot's manipulation actions.

In real life human-robot collaboration the robot needs to be aware of three main states: the state of the human, the state of the objects to manipulate, and its own state. In this work the state of the human is the instruction command given by the user using a static hand gesture. The state of an object is whether it is in the robot's hand or not. A manipulation action to be performed by the robot is selected based on the probability of the action given these three states. However, the presence of a human results in a large state-action space where the robot would need to explore all the possible combinations of associations. It can be time expensive and practically challenging for a human to label all the state-action associations. For example, in this work we have 4 states of the human, 2 states of the object, and 5 states of the robot then the robot will have to explore the state-action space of 200 associations.

The authors are with the Intelligent and Interactive Systems, Institute of Computer Science, University of Innsbruck, Austria. Corresponding author: dadhichi.shukla@uibk.ac.at

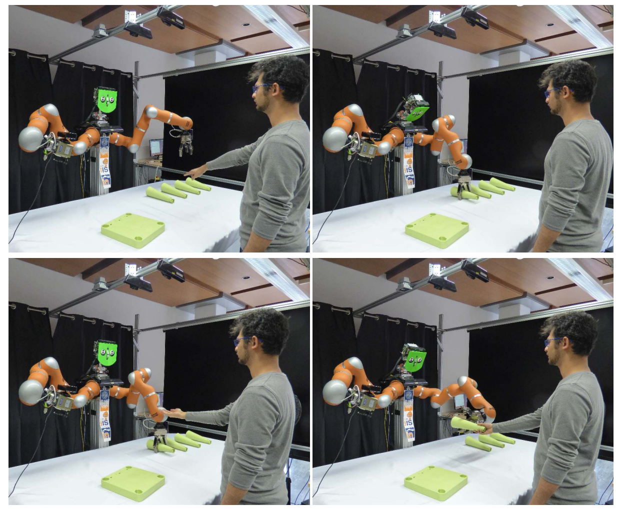

Fig. 1: Robin (the robot) assists the user in the assembly of a table. The user performs gestures like 'pointing' and 'give me' to which the robot reacts as 'grasp object' and 'hand over', respectively. It uses its gaze to indicate the action to be performed.

An advantage of the PIL framework is that it does not require prior training of the associations. Rather, the gesture-action associations are learnt on the fly while working with the user to reach the goal. Although the user sees the associations between gestures and actions, the robot actually learns associations between the state of the system (i.e., the state of the human, the state of the object, and the robot's own state) and the action performed.

Cohen et al. [1] describe *collaboration* not only as the commitment of members to achieve the goal but also – if necessary – as having a mutual belief about the state of the goal. Our framework is motivated by their joint intention theory to design a collaborative framework. The proposed PIL framework is an extension of our previous work [2]. In our previous work we studied benefits of a proactive learning over active learning which requires prior training.

A. Contributions

The table assembly task takes place in close proximity; hence it is irrelevant to observe full-body human pose. We choose to interact using hand gestures since they are more natural than a computer interface. Moreover, they inherently provide spatial information of the user's hand. For example, gestures like pointing can be used to localize objects [3]. We use gestures from the Innsbruck Multi-view Hand Gestures

(IMHG) dataset [4]. The gestures in the IMHG dataset[1] are closely related to the semantic content of verbal language and were designed based on the human-robot interaction study conducted by Jensen et al. [5].

The proposed framework is statistically driven. The robot receives feedback (positive or negative) on the action it has performed. Candidate state-action associations are scored based on the feedback, consequently learning the mapping. The novelty of our PIL framework lies in the following two attributes. First, it is *incremental*: It is a supervised, probabilistic learning approach; however, unlike traditional machine learning techniques, we posit that incremental learning provides the freedom to the user to establish the gesture-action associations at will. In other words, initially the robot is not aware of associations between the state of the system and the manipulation action. There are no distinct *training* and *testing* phases. Since associations are learnt on the fly both phases are active till the system reaches the goal. Secondly, our framework is *proactive*. A recent HRI study by Jensen et al. [6] shows that people expect a robot to act proactively during the interaction. The proactive nature of the framework requires the robot to predict the intent of the user. It becomes active once the scores of the associations begin to consolidate, i.e., after some number of interactions. Consequently, based on the learnt probabilities, the robot can decide on the most likely action.

We use hand gestures to instruct the robot, but current gesture detection systems are prone to misclassification. A misclassified gesture can result in an invalid state of the system. For example, a *grasp object* gesture may be detected instead of a *release object* while the robot is already holding an object. Invalidity of a state may be detected in two main ways: (1) A gesture is detected with a low confidence score irrespective of the state of the system; therefore it is discarded; (2) a misclassified gesture is incompatible with the state of the system irrespective of the confidence score. The PIL framework allows the robot to proactively correct the misclassified gesture based on the acquired knowledge. The state of the human as understood by the robot depends on the detection accuracy of the instructional gestures. We quantitatively study and discuss the effect of gesture detection accuracy in Section IV.

We also use the gaze of the robot for communication to establish common ground in our collaborative framework. Since our robot currently does not speak or have a screen, in order to complete the communication cycle, the robot uses its head (or gaze) to establish mutual belief regarding the action to be performed. In this work the main purpose of gaze is to indicate to the user where the robot action is going to take place. For example, the robot will look at its hand if it is going to either open it or close it, or the robot will look at the object if is going to reach for it.

To summarize, the main contributions of this paper are,

1) a fast, statistical, supervised learning framework – Proactive Incremental Learning (PIL) – for learning the associations between hand gestures and the robot's manipulation actions,
2) on-the-fly learning of these associations,
3) prediction of the intent of the user to speed up the task,
4) proactive gesture correction,
5) study of the effect of accuracy of the gesture detection system,
6) use of robot gaze to facilitate time efficient HRI.

[1] `https://iis.uibk.ac.at/downloads/start`

B. Related work

Numerous studies [7], [8], [9] demonstrate the advantage of active, human-in-the-loop interaction. Lenz et al. [10] proposed a framework to allow joint action of humans and robots for an assembly task. Their system is capable of anticipating human behaviour based on the learnt sequence, ensuring smooth collaboration. The work by Pellegrinelli et al. [11] is based on Partially Observable Markov Decision Processes (POMDPs) for shared autonomy in order to provide assistance to the user without knowing the exact goal. Myagmarjav et al. [12] proposed an incremental active-learning architecture trained with limited knowledge of the task. The robot asks questions to acquire relevant information from the user. A pertinent difference with these methods is that the PIL does not require an explicit training phase.

Thomaz et al. [13] introduced Interactive Reinforcement Learning (IRL) which is based on Q-learning. It has demonstrated that human-generated reward can be powerful and can be fast compared to classical reinforcement learning. It enables the user to provide positive and negative rewards in response to the manipulation action of the robot to train it on the task. In addition to evaluative feedback, in the work by Suay et al. [14] the user provides guidance signals to constrain the exploration towards a limited set of actions. In their method the user has to provide feedback for every action. A similar system by Najar et al. [15] learns the meaning of the guidance signals by using evaluative feedback instead of task rewards.

We propose a method in which the user can interact with the robot using hand gestures, which is a natural, unencumbered, non-contact, and prop-free mode of interaction. Though state-of-the-art methods have shown promising results in simulated and/or controlled scenarios, it is essential to study practical challenges with a real robot. Therefore in this work we study the effects of detection accuracy of an instruction command, the time taken for path planning, the time taken to detect the hand gesture, and the time taken by the robot to perform an action. We evaluate the PIL framework in a simulated as well as a real robot environment.

Various studies have demonstrated the role of the gaze of the robot and its benefits in HRI. Ruesch et al. [16] show that gaze aids users in interpreting the robot's action and can contribute to smoother interaction. Likewise, gaze also serves to indicate readiness of the robot [17]. Fischer et al. [18] specifically explore the effects of social gaze in a human-robot collaborative toolbox assembly scenario with naive users. Their analyses show that people engage significantly faster with the robot with active gaze. Our PIL framework

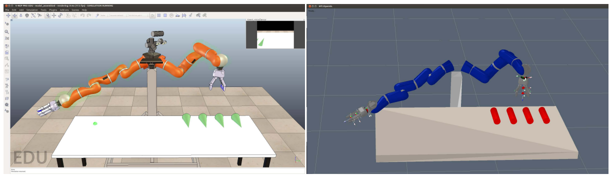

Fig. 2: *L*: Simulated scene of Robin in V-REP, *R*: KOMO motion planner to plan and execute manipulation actions of Robin.

incorporates robot gaze in ways found to be effective in these studies.

II. Human-Robot Collaboration Scenario

A. Assembly Scenario

We designed the HRI setup with 'Robin', an anthropomorphic configuration of two KUKA light-weight robot arms, two Schunk SDH-2 hands, and a KIT robotic head [19] in a simulated and a real environment. To simulate the environment we use the V-REP robot simulator[2]. We use the left arm of the robot to interact since gestures in the IMHG dataset are commonly performed with the right hand. The robot actions are planned and executed using the KOMO motion planner [20]. An illustration of the simulation and the motion planner is shown in Fig. 2.

Let $G = \{\mathsf{pointing}, \mathsf{give\ me}, \mathsf{grasp}, \mathsf{release}\}$ be the set of 4 instructional gestures. The semantics learned by the robot of the gestures in G are, *pointing* is to indicate an object or a position in the workspace, *give me* is to instruct the robot to bring the object towards the user, *grasp* is to instruct the robot to grasp the object or close its own hand, *release* is to instruct the robot to release the object or open its hand. It is to be noted that the PIL framework provides the flexibility to learn these meanings as per the user's choice. Let $F = \{\mathsf{ok}, \neg\mathsf{ok}\}$ be the set of feedback signals given by the user by performing *thumb up* and *thumb down* gestures, respectively. The feedback is in the form of a binary approval signal. It determines how a gesture-action association is scored. Since the robot learns the associations based on feedback signal, the meaning of gestures in F is known. We use probabilistic appearance-based pose estimation (PAPE) by Erkent et al. [21] to detect gestures. It is based on the probabilistic representation of the data by Teney et al. [22].

Let $A = \{\mathsf{open}, \mathsf{close}, \mathsf{object}, \mathsf{human}, \mathsf{location}\}$ be the set of 5 manipulation actions known to the robot. They are defined as, open is to open robot's hand, close is to close robot's hand, object is to go on the top of the pointed object, human is to go to the location of the user's give me gesture, and location is go to the pointed location in the workspace.

Let $E = \{\mathsf{hand}, \mathsf{object}, \mathsf{palm}, \mathsf{position}, \mathsf{face}\}$ be the set of 5 gaze directions. The robot uses these gaze directions to communicate to the user about the action it is going to perform next or that it is ready to detect the next gesture. Gaze directed at the robot's own hand indicates that action open or close will be performed next, the robot may direct its gaze at the object or workspace position pointed at by the user, palm means gazing at the user's give me gesture, and gazing at the user's face indicates that the robot is ready for the next gesture.

The state of the object mainly depends on the state of the robot's hand. Hence we define the set $H = \{\mathsf{free}, \mathsf{occupied}\}$ with the 2 states of the robot hand, representing whether or not it is currently holding an object.

[2] http://www.coppeliarobotics.com/

B. Task execution

The state s of the system at the time step t consists of three attributes. It is defined as $s_t = \langle g_t, a_t, h_t \rangle$, where $g_t \in G$ is the detected gesture, $a_t \in A$ is the state of the robot, i.e., the action performed by the robot in the previous step, and $h_t \in H$ is the state of the robot's hand. At each step t of the assembly the robot records three entities, the state s_t of the system, the action $a_{t+1} \in A$ the robot will perform, and the feedback signal $f_{t+1} \in F$ given by the user after the action. Each action a_{t+1} is preceded by a gaze movement $e_{t+1} \in E$ to inform the user about the action to be performed. Henceforth the gaze movement always precedes the action to be performed.

Consider the sequence of sextuples as shown in Table I, describing handover of legs of the table. Each block delimited by a dotted line is one gesture-action association, and each block delimited by two solid lines is one handover. For example, steps 1 to 3 is one gesture-action association, and steps 1 to 9 is one handover. Robin performs gaze face, i.e. looks at the user, after each action to indicate that the action is done and it is ready for the next command. For explanation purposes let us consider that gestures are detected accurately. Robin is at a default position with its hand open and the four legs of the table are kept within its reachable workspace.

At time $t = 1$, the user points at one of the legs of the table. The vision system detects the hand gesture as pointing. Robin then selects one of the random actions from A, here close. The gaze paired with action close is hand, i.e. to look at its hand. The user observes the robot looking at its own hand. Knowing that the hand gaze in a free hand state indicates an imminent close action, here the user immediately reacts by giving $\neg$ok feedback because close is not the desired action. Similarly, $\neg$ok is also given upon the next randomly-selected position gaze, which is paired with action location. Finally, at time $t = 3$ the user gives ok, the object gaze signaling that the robot is about to perform the intended object action.

Next, the user performs a grasp gesture to instruct the robot to grasp the object. A similar procedure as above follows, and the robot learns the association between the gesture grasp and the action close. At some point Robin will have learnt that it has to perform the object action given the open state of the robot, the free state of its hand, and the pointing state of the user. This procedure repeats till Robin has handed all legs to the user.

t	g_t	a_t	h_t	e_{t+1}	f_{t+1}	a_{t+1}
1	pointing	open	free	hand	¬ok	
2		open	free	position	¬ok	
3		open	free	object	ok	object
4	grasp	object	free	palm	¬ok	
5		object	free	hand	ok	close
6	give me	close	occupied	position	¬ok	
7		close	occupied	palm	ok	human
8	release	human	occupied	object	¬ok	
9		human	occupied	hand	ok	open
10	pointing	open	free	object	ok	object
11	grasp	object	free	hand	ok	close
12	pointing	close	occupied	hand	¬ok	
13		close	occupied	palm	¬ok	
14		close	occupied	position	ok	location
15	release	location	occupied	hand	ok	open
16	pointing	open	free	object	ok	object
17	pointing	object	free	hand	¬ok	
18		object	free	object	ok	object
19	grasp	object	free	hand	ok	close
20	pointing	close	occupied	position	ok	location
21	release	location	occupied	hand	ok	open
22	pointing	open	free	object	ok	object
23	grasp	object	free	hand	ok	close
24	pointing	close	occupied	position	ok	location
25	give me	location	occupied	object	¬ok	
26		location	occupied	palm	ok	human
27	release	human	occupied	hand	ok	open

TABLE I: An example of the robot handing over 4 table legs while learning gesture-action associations.

III. Proactive Incremental Learning

The Proactive Incremental Learning (PIL) framework is designed to reach the goal while minimizing the number of actions performed by the robot as well as reducing the number of gestures needed to be shown by the user. Let us consider the human-robot collaboration scenario of assembling a table as described in Section II. Let N be the number of legs Robin has to hand over to the user. The proactive incremental learning framework consists of two modules: (1) incremental gesture-action associations, and (2) proactive gesture prediction or correction.

A. Incremental Gesture-Action Association

The PIL framework works with an underlying assumption that the user wants to complete the task in a minimum number of interactions. The robot learns $P(a_{t+1}|s_t)$ which is the probability of an action a_{t+1} to be executed given the state of the system s_t during the assembly task. Initially, the probabilities are distributed uniformly among all the robot actions. The robot incrementally learns these probabilities at every step using the feedback from the user.

As mentioned earlier each action is paired with a gaze movement which precedes the actual action. The gaze movement is used to indicate the user in advance about the chosen action. It is a heuristic in our framework based on aforementioned gaze studies. The user then gives a feedback ok or ¬ok based on whether the robot has indicated an appropriate action to proceed towards the goal. If ok is signaled then the robot goes ahead performing the action; otherwise it will indicate another action using its gaze. In the situation that a gaze is misunderstood and an undesired action is performed then the user can give a ¬ok feedback after the action. Then the state-action score is updated with the latter feedback. The feedback f is received as a binary score η for the state-action association (s_t, a_{t+1}) as

$$\eta = \begin{cases} 1, & f_{t+1} = \mathsf{ok} \\ 0, & f_{t+1} = \neg\mathsf{ok}. \end{cases} \tag{1}$$

In the human-robot interaction studies conducted by Jensen et al. [5], the authors observed that for an appropriate action taken by the robot *no feedback* or rarely *a positive* feedback is given. On the other hand, users strongly give negative feedback for unexpected actions. Therefore in the PIL framework we consider *no feedback* too as ok.

Let T be the 4-D table storing the state-action associations at each step t. The score of each state-action association is recorded in cell $\mathsf{T}(s_t, a_{t+1})$, and is updated based on the binary score η as

$$\mathsf{T}(s_t, a_{t+1}) = \mathsf{T}(s_t, a_{t+1}) + \eta. \tag{2}$$

At $\eta = 0$ the scores of state-action associations other than the indicated action are incremented by 1. We describe this as *complement feedback* technique. Previous methods [14], [15], give either -1 or 0 for a ¬ok feedback. Such a strategy either precludes a probabilistic formulation or lessens the influence of the feedback. Contrarily the *complement feedback* technique maintains the probabilistic nature of the PIL framework as well as promotes the decisiveness of the feedback. The scores of all the possible actions $a^{\times}_{t+1} \in A$ except a_{t+1} are updated as

$$\mathsf{T}(s_t, a^{\times}_{t+1}) = \mathsf{T}(s_t, a^{\times}_{t+1}) + 1,\ \eta = 0,\ a^{\times}_{t+1} \neq a_{t+1}. \tag{3}$$

For example, at $t = 1$, in Table I $s_1 = \langle \mathsf{pointing}, \mathsf{open}, \mathsf{free} \rangle$ and $e_2 = \langle \mathsf{hand} \rangle$ i.e., $a_2 = \langle \mathsf{close} \rangle$, receives a ¬ok feedback so the score in cell $\mathsf{T}(s_1, \mathsf{close})$ remains the same. However, scores of all the other associations $\mathsf{T}(s_1, \mathsf{open})$, $\mathsf{T}(s_1, \mathsf{object})$, $\mathsf{T}(s_1, \mathsf{human})$, and $\mathsf{T}(s_1, \mathsf{location})$ are incremented by 1.

During the early steps of the task the robot performs random actions since it has not acquired any knowledge. Though in later steps of the task it has updated the score for the state-action association (s_t, a_{t+1}) and it can compute $P(a_{t+1}|s_t)$. The probability of an action a given the state s of the system is computed as

$$P(a|s) = \frac{\mathsf{T}(s, a)}{\sum_{i=1}^{|A|} \mathsf{T}(s, a_i)}. \tag{4}$$

If the score values in T are normalized as per the joint probabilities then eq. 4 represents Bayes' rule. The action a_{t+1} to perform given the state of the system is selected as

$$a^* = \underset{a_{t+1}}{\operatorname{argmax}}\, P(a_{t+1}|s_t). \tag{5}$$

B. Proactive gesture prediction or correction

1) Gesture prediction: Human intent prediction has been discussed in various studies [11], [23], [24] in the context of human-robot interaction. In the PIL framework we focus on human-robot collaboration in a shared workspace. The key aspect of predicting human state is to minimize efforts on the user and train the robot to act independently.

The robot can predict the next gesture of the user after it has progressed on the task. In other words, the prediction module is active once it has recorded the state-action associations in T. The goal of the PIL framework is to predict $g_{t+1} \in G$. The robot can refer back to the history of interactions in T to compute the probability of the next gesture g_{t+1} given the current state s_t of the system and the associated action a_{t+1}, i.e., $P(g_{t+1}|s_t, a_{t+1})$. The gesture with the highest probability $P(g_{t+1}|s_t, a_{t+1})$ is selected as the predicted gesture g_{t+1}.

Predicting the next gesture enables the robot to proactively decide on the action a_{t+2} associated with g_{t+1}. An advantage of predicting a gesture is that human does not have to make an effort in performing the gesture. However, if the user decides to deviate from the learnt sequence then a different gesture can be performed after the execution of a_{t+1}. The readiness of the robot is indicated to the user with the gaze movement face. If the user does not perform any new gesture then the robot proceeds with the task with the prediction g_{t+1}.

2) Gesture correction: Accurate detection of the gesture is vital in learning of the gesture-action associations. However, the gesture detection system may misclassify due to practical challenges like changes in the lighting conditions, differences in the appearance of gestures among users, etc. Misclassification of a gesture can evoke an invalid state of the system. To overcome this problem we incorporate a *gesture correction* module in the PIL framework.

The methodology is similar to that described for *gesture prediction* in section III-B.1. The gesture correction module too becomes active after associations are recorded in T. If the detected gesture g_t triggers an invalid state then no a_{t+1} is selected. At this point the system checks for all the state-action associations which were followed by (s_{t-1}, a_t) and were given feedback ok.

The robot computes the $P(g_t|s_{t-1}, a_t)$ for the gestures. The gesture with the highest probability is chosen as the *corrected* gesture. Based on the updated s_t it then performs gaze e_{t+1} which is paired with a_{t+1}. The user always has the freedom to provide $\neg\mathsf{ok}$ feedback. The robot then selects the next best action. If none of the previously-learnt actions are acceptable by the user then the robot explores the state-action associations with $\neg\mathsf{ok}$ feedback.

A valid state-action association can be miscategorized as $\neg\mathsf{ok}$ and an invalid state-action association can be miscategorized as ok if the gesture detection system has a poor accuracy. An advantage of the PIL framework is that at first it proactively decides to perform an action, and if it is not the one the user expects then it enables the user to choose the action as described in section III-A.

C. PIL example

From the user's point of view the interaction follows the flowchart as shown in Fig. 3 In the flowchart *Yes* and *No* decisions are ok and $\neg\mathsf{ok}$ feedbacks, respectively. The user sees it as learning of gesture-action associations, however, the state of the hand and the state of the robot are not actively perceived as part of the association by the user. Here we explain the PIL framework using the sample assembly sequence described in Table I.

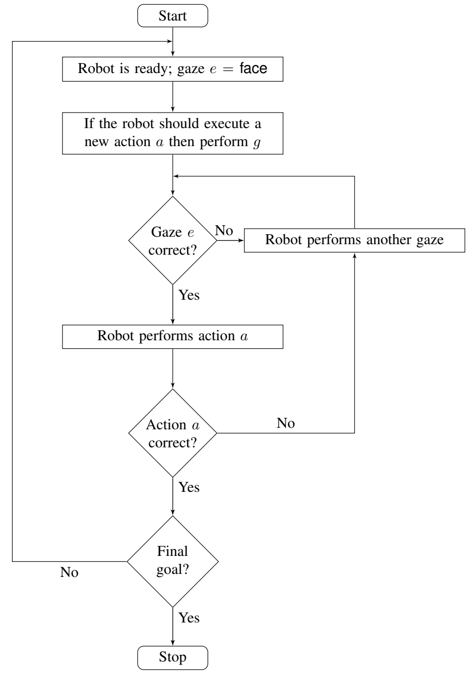

Fig. 3: Human-robot collaboration flowchart from the user's viewpoint.

At step $t = 3$ the robot learns the association between the state of the system $s_3 = \langle\mathsf{pointing}, \mathsf{open}, \mathsf{free}\rangle$ with the action $a_4 = \langle\mathsf{object}\rangle$. It can be seen at $t = 10$ that the robot chooses action object based on the learnt association instead of a random selection.

In another instance at $t = 17$ the user points at another object. From the user's perspective it may seem that the robot will perform action object. However, $s_{17} = \langle\mathsf{pointing}, \mathsf{object}, \mathsf{free}\rangle$ has never occurred previously. At this point the probability of an action given the state s_{17} is uniformly distributed among all the possible actions. Therefore, the robot opts to perform a random action.

The PIL framework enables the design of an intent prediction model to speed up the task. The robot can predict the next likely gesture based on the learnt probabilities of the state-action association. For example, at step $t = 15$, the robot predicts that the next most likely gesture at $t = 16$ succeeding $\mathsf{release}$ is $\mathsf{pointing}$. Similarly, at $t = 22$ the robot predicts that the user can perform two possible gestures at $t = 23$, either grasp or $\mathsf{pointing}$. The grasp gesture, however, has as higher prediction score compared to $\mathsf{pointing}$.

In a real environment the robot might misclassify the instruction gesture. A misclassified gesture can evoke an invalid state. For example, at $t = 21$, if $\mathsf{release}$ is misclassified as grasp, it would trigger an invalid state. The robot cannot grasp an object when the state of the hand is occupied. The system then decides that gesture correction is necessary. Based on the learnt probabilities the most likely gesture to occur after pointing given $a_t = \langle \text{location} \rangle$ and $h_t = \langle \text{occupied} \rangle$ is release. The detected gesture is corrected from grasp to release, and the robot performs action open.

IV. Results and Discussion

We conduct 2 sets of quantitative experiments to evaluate the proactive incremental learning (PIL) framework. The first experiment takes place in the simulated environment shown in Fig. 2 to compare the PIL with state-of-the-art methods. We also compare with two additional conditions to show the advantage of using the robot's gaze movements. The second experiment is with the the real robot where we compute various evaluation metrics for human-robot interaction proposed by Olsen et al. [25].

1) Experiment 1: We simulated the table assembly scenario described in section II-B. The entire interaction consists of assembling 3 tables each with 4 legs; i.e., in total the robot has to hand 12 table legs over to the user. To obtain statistically-significant data we repeat the interaction 5 times. We performed t-test on the acquired data to check if the data points are significantly different from each other. The gestures g and the feedback f are fed to the system using a computer interface.

A critical aspect in HRI is the accurate detection of the command signal (here hand gestures). To evaluate the effect of accuracy of the vision system we simulated the gesture detection rate d. We simulate 3 detection rates $d = \{0.6, 0.8, 1.0\}$. For example, if the gesture sequence (the g_t column) in Table I is used repeatedly to hand over 12 legs at $d = 0.6$, then only 32 out of 54 gestures will be detected correctly.

We quantitatively compare the PIL framework with two interactive reinforcement learning (IRL) methods. IRL methods have shown promising results in learning tasks involving human feedback. However, in proposed HRI scenario in addition to doing the task the robot also has to learn the gesture-action associations. The two IRL implementations of the table assembly scenario are as described by Suay et al. [14] (IRL1) and by Najar et al. [15] (IRL2).

We set the learning rate α and the discount factor γ for IRL1 and IRL2 to $\alpha = 0.3, \gamma = 0.75$ and $\alpha = 0.3, \gamma = 0.0$, respectively. The authors argue that $\gamma = 0.0$ is more suitable for learning from human feedback. It allows a task to be divided into a sequence of single-step tasks. However, $\gamma = 0.0$ would make reinforcement learning aspect incongruous for overall task learning. For a fair comparison we also incorporate the gaze e and consider *no feedback* as ok feedback for both methods. Nevertheless, the IRL methods do not predict the next gesture; therefore they require instructional gesture at every step.

As described earlier, all manipulation actions are preceded by a corresponding gaze. Here we evaluate its significance for human-robot collaboration. We analyse how the feedback after the gaze movement incorporated in the PIL framework works to the advantage of the user. The 2 conditions to compare with the PIL framework are,

1) *Without post-gaze feedback*: After the gaze movement no feedback is considered; feedback gestures are only detected after a manipulation action has been performed.
2) *With post-gaze feedback*: The feedback is considered after the gaze movement. It also waits for the feedback after the manipulation action is performed to guarantee that the gaze was rightly understood. However, the system considers latter feedback if it differs from the first one.

It is to be noted that in both of the above conditions the robot is not proactive, i.e., it cannot perform gesture prediction or correction.

The results in Figs. 4 and 5 show the comparison of the PIL framework with IRL1, IRL2, and the 2 conditions with respect to 2 criteria, the number of gestures performed by the user and the number of the robot's actions, respectively. The number of gestures by the user is the sum of the number of instructional gestures G and the number of feedback gestures F. The number of the robot's actions is the sum of the number of manipulation actions A and the number of gaze movements E. A general observation from our experiments is that as the detection rate increases the effort of the user and the number of robot actions reduces for all the approaches. The results of IRL methods are close to each other because the associations are scored with either -1 or 0 for a $\neg$ok feedback.

It can be seen in Fig. 4 that the number of gestures performed by the user reduces significantly with the PIL framework. Since the PIL system is able to predict the user's next gesture, it frees the user from the effort of performing a gesture. The proactive behaviour allows the robot to proceed with the task without attention from the user. The results from IRL1, IRL2, and the 2 conditions are comparable at $d = \{0.8, 1.0\}$. However, condition 2 is better than both IRL methods at $d = 0.6$.

The results in Fig. 5 shows that the PIL framework requires far fewer robot actions to hand over all legs compared to other conditions. This is because of the complement feedback technique as given by eq. 3. It gives an advantage to the PIL framework to minimize the number of robot actions. It can be seen that condition 1 requires more actions by a huge factor compared to the others. Since there is no post-gaze feedback the robot would directly perform the action without considering whether it was desired by the user or not. And in order to avoid the robot heading towards a cul-de-sac it needs to go back to a previous valid state whenever a $\neg$ok feedback is given. However, if state of the object has changed then the robot cannot reverse the state. For example, at $t = 7$, in Table I if the robot performs $a_t = \langle \text{open} \rangle$ instead of $a_t = \langle \text{human} \rangle$ when $h_t = \langle \text{occupied} \rangle$, the object will drop in the workspace. In such cases the robot will go back to default position.

The t-test is performed on the number of hand gestures and the number of robot actions for all the methods at various

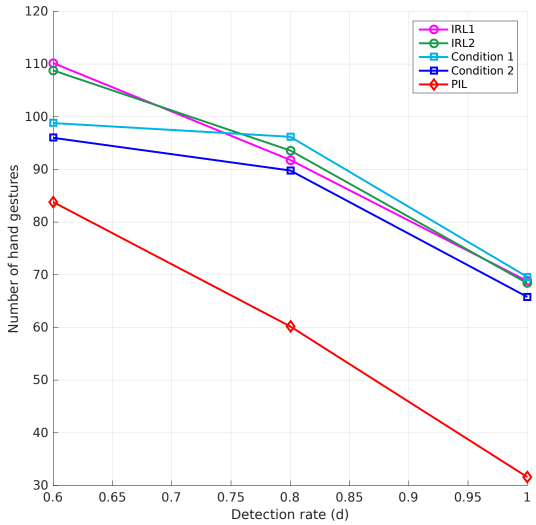

Fig. 4: Number of gestures performed by the user at various detection rates to assemble 3 tables.

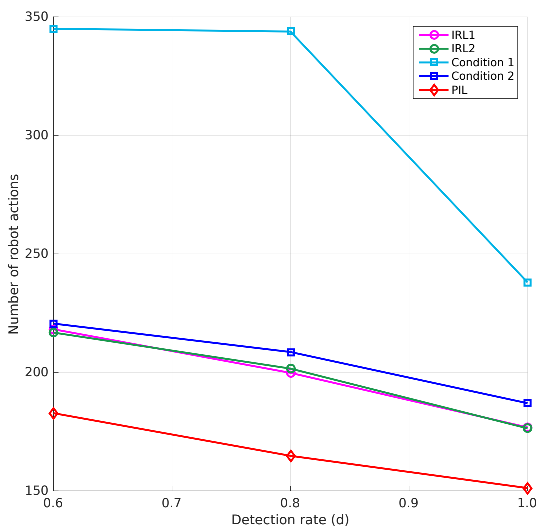

Fig. 5: Number of robot actions at various detection rates to assemble 3 tables.

detection rates. The p-values of our data indicate that the *null hypothesis* can be rejected with 1% significance level. Our simulation data is statistically significant with $p < 0.01$ having a maximum at $p_{\max} = 0.00012$.

2) Experiment 2: The experiment with the real robot consists of the same interaction setup as described in section IV-.1. Olsen et al. [25] proposed various interrelated metrics to evaluate the quality of human-robot interaction frameworks. There are a variety of metrics; however, for the purpose of the PIL framework, we focus on the overall task effectiveness.

We compute 3 evaluation metrics, Neglect Tolerance (NT), Interaction Effort (IE), and Robot Attention Demand (RAD). These are time-based metrics that attempt to maximize the speed of performance, minimize mistakes, and measure the autonomy of the robot. *Neglect tolerance* is defined as the amount of time a human can ignore a robot. It represents

	1		2		3	
HRI metrics	Cond. 2	PIL	Cond. 2	PIL	Cond. 2	PIL
IE	204.00	153.60	188.00	155.00	195.00	151.50
NT	715.29	665.38	652.04	643.57	728.26	721.19
RAD	0.22	**0.18**	0.22	**0.19**	0.21	**0.17**

TABLE II: Comparison of HRI evaluation metrics for condition 2 and PIL. Interaction effort (IE), Neglect tolerance (NT) are measured in seconds and Robot attention demand (RAD) is a unitless quantity.

tasks which a robot can perform without human supervision. An obvious goal of the framework is to increase the NT of the robot. *Interaction effort* is the amount time the user has to invest with the robot. The goal of the system is to reduce the IE and lead the robot towards proactive behaviour. *Robot attention demand* is the relation between NT and IE given by

$$\mathrm{RAD} = \frac{\mathrm{IE}}{\mathrm{IE} + \mathrm{NT}}. \tag{6}$$

It is a unitless quantity that represents the effort that the user expends interacting with the robot relative to the total robot time. A good human-robot interaction system tries to minimize RAD value. The lower the value of RAD, the better it is because then the user can focus on other tasks besides interacting with the robot.

To compute NT we take into account the time taken by various sub-tasks performed by the robot like planning the trajectory of the robot arm, executing the trajectory, opening and closing its hand, and gaze movements to indicate either the next action or its readiness. The IE is computed as the sum of the time taken by the vision system to detect user's hand gestures i.e., the instruction gestures G and the feedback gestures F. It takes 1.5-2.5 seconds for the PAPE method [21] to detect various hand gestures of the IMHG dataset [4]. Its detection accuracy ranges from 75%-95% to detect different hand gestures in the real robot scenario.

RAD measures the effort that a user has to invest in the interaction. And since the results in Fig. 4 show that condition 2 performs better than other non-proactive conditions, we compare the PIL framework only with condition 2. The evaluation metric results of the comparison of the PIL with condition 2 are shown in Table II. We assemble 3 tables 3 times with the real robot. It can be seen from the values of IE and RAD that the user has more free time interacting using the PIL framework than condition 2. In the proposed scenario a high NT does not necessarily suggest a better HRI. The value of NT can be increased by slowing down the speed of the robot. However, we would like to complete the table assembly task in the least time. The NT for condition 2 is higher because the robot has to perform more number of gaze $e = \langle \textsf{face} \rangle$ to indicate that it is ready for the next gesture.

The PIL framework can be scaled up where the state s of the system consists of more than 3 attributes. For example, in a complex task of furniture assembly the robot may have to perform sub-tasks like hold an object or a nail, insert a screw, or use a screw driver or a hammer. Additional attributes can be, the state of both hands of the user, the state of multiple objects, or the state both arms of the robot. However, this can increase the time of interaction since the system will have to detect multiple gestures and plan for both the robot arms. We ask the readers to refer to the *video attachment* with this article demonstrating the PIL framework with the real robot for assembly of the table.

V. Conclusions

We proposed a fast, supervised Proactive Incremental Learning (PIL) framework to learn the associations between human hand gestures and the robot's manipulation action. We also introduced a novel *complement feedback* technique which ensures the probabilistic nature of the PIL framework. Our quantitative analyses demonstrate that the complement feedback technique promotes the decisiveness of the feedback, consequently reducing the number of robot actions. The results from the simulated and the real environment experiments show that the PIL framework outperforms state-of-the-art methods.

Due to the incremental nature of the PIL framework, users are free to train the robot the gesture-action associations of their own choice. The proactive behaviour learns to predict the next gesture therefore reducing the interaction effort (IE) of the user. We studied that introducing feedback post-gaze facilitates the human-robot collaboration with two aspects, 1) establishing the mutual belief regarding the robot's next action, and 2) it speeds-up the interaction without executing undesired actions. We are currently working on methods to overcome the randomized selection of the robot actions based on the knowledge acquired on the fly.

Acknowledgement

The research leading to this work has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 (Specific Programme Cooperation, Theme 3, Information and Communication Technologies) under grant agreement no. 610878, 3rd HAND. The authors would like to thank Safoura Rezapour Lakani (University of Innsbruck) for discussions.

References

[1] P. R. Cohen and H. J. Levesque, "Teamwork," *Nous*, vol. 25, no. 4, pp. 487–512, 1991.

[2] D. Shukla, O. Erkent, and J. Piater, "Supervised learning of gesture-action associations for human-robot collaboration," in *1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production*, 2017, workshop at the 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG). [Online]. Available: https://iis.uibk.ac.at/public/papers/Shukla-2017-ASL4GUP.pdf

[3] ——, "Probabilistic detection of pointing directions for human robot interaction," in *International Conference on Digital Image Computing: Techniques and Applications*, 11 2015. [Online]. Available: https://iis.uibk.ac.at/public/papers/Shukla-2015-DICTA.pdf

[4] ——, "A multi-view hand gesture rgb-d dataset for human-robot interaction scenarios," in *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016.

[5] L. Jensen, K. Fischer, D. Shukla, and J. Piater, "Negotiating Instruction Strategies during Robot Action Demonstration," in *10th ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015. [Online]. Available: https://iis.uibk.ac.at/public/papers/Jensen-2015-HRI.pdf

[6] L. Jensen, K. Fischer, F. Kirstein, D. Shukla, O. Erkent, and J. Piater, "It gets worse before it gets better," in *12th ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017. [Online]. Available: https://iis.uibk.ac.at/public/papers/Jensen-2017-HRI.pdf

[7] J. Bandera, J. Rodriguez, L. Molina-Tanco, and A. Bandera, "A survey of vision-based architectures for robot learning by imitation," *International Journal of Humanoid Robotics*, 2012.

[8] W. B. Knox and P. Stone, "Reinforcement learning from human reward: Discounting in episodic tasks," in *2012 IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2012.

[9] L. Rozo, P. Jiménez, and C. Torras, "A robot learning from demonstration framework to perform force-based manipulation tasks," *Intelligent service robotics*, 2013.

[10] C. Lenz, S. Nair, M. Rickert, A. Knoll, W. Rosel, J. Gast, A. Bannat, and F. Wallhoff, "Joint-action for humans and industrial robots for assembly tasks," in *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2008.

[11] S. Pellegrinelli, H. Admoni, S. Javdani, and S. Srinivasa, "Human-robot shared workspace collaboration via hindsight optimization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[12] B. Myagmarjav and M. Sridharan, "Incremental knowledge acquisition for human-robot collaboration," in *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2015.

[13] A. L. Thomaz and C. Breazeal, "Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance," in *AAAI*, 2006.

[14] H. B. Suay and S. Chernova, "Effect of human guidance and state space size on interactive reinforcement learning," in *RO-MAN, 2011 IEEE*, 2011.

[15] A. Najar, O. Sigaud, and M. Chetouani, "Training a robot with evaluative feedback and unlabeled guidance signals," in *International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016.

[16] J. Ruesch, M. Lopes, A. Bernardino, J. Hornstein, J. Santos-Victor, and R. Pfeifer, "Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008.

[17] N. Bee, E. André, and S. Tober, "Breaking the ice in human-agent communication: Eye-gaze based initiation of contact with an embodied conversational agent," in *International Workshop on Intelligent Virtual Agents*. Springer, 2009.

[18] K. Fischer, L. Jensen, F. Kirstein, S. Stabinger, O. Erkent, D. Shukla, and J. Piater, "The Effects of Social Gaze in Human-Robot Collaborative Assembly," in *Social Robotics*, ser. LNAI, A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi, Eds. Springer, 2015. [Online]. Available: https://iis.uibk.ac.at/public/papers/Fischer-2015-ICSR.pdf

[19] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, "The Karlsruhe humanoid head," in *Humanoids 2008-8th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2008.

[20] M. Toussaint, "Newton methods for k-order markov constrained motion problems," *CoRR*, vol. abs/1407.0414, 2014. [Online]. Available: http://arxiv.org/abs/1407.0414

[21] O. Erkent, D. Shukla, and J. Piater, "Integration of probabilistic pose estimates from multiple views," in *European Conference on Computer Vision*, 2016. [Online]. Available: https://iis.uibk.ac.at/public/papers/Erkent-2016-ECCV.pdf

[22] D. Teney and J. Piater, "Multiview feature distributions for object detection and continuous pose estimation," *Computer Vision and Image Understanding*, 2014. [Online]. Available: https://iis.uibk.ac.at/public/papers/Teney-2014-CVIU.pdf

[23] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*, 2013.

[24] F. Stulp, J. Grizou, B. Busch, and M. Lopes, "Facilitating intention prediction for humans by optimizing robot motions," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, 2015.

[25] D. R. Olsen and M. A. Goodrich, "Metrics for evaluating human-robot interactions," in *Proceedings of Performance Metrics for Intelligent Systems*. Citeseer, 2003.