Towards Planning Human-Robot Interactive Manipulation Tasks: Task Dependent and Human Oriented Autonomous Selection of Grasp and Placement

Amit Kumar Pandey¹,², Jean-Philippe Sa¹,²ut, Daniel Sidobre¹,³ and Rachid Alami¹,²

Abstract—In a typical Human-Robot Interaction (HRI) scenario, the robot needs to perform various tasks for the human, hence should take into account human oriented constraints. In this context it is not sufficient that the robot selects grasp and placement of the object from the stability point of view only. Motivated from human behavioral psychology, in this paper we emphasize on the mutually depended nature of grasp and placement selections, which is further constrained by the task, the environment and the human’s perspective. We will explore essential human oriented constraints on grasp and placement selections and present a framework to incorporate them in synthesizing key configurations of planning basic interactive manipulation tasks.

I. INTRODUCTION

HRI requires the robot to be capable of performing pick-and-place operations for a variety of tasks such as show, give, hide, make accessible, etc. The nature of the task and the presence of human germinate additional constraints on the grasp and placement selections, which demand reasoning beyond the shape and stability of the object.

A. Motivation from Human Behavior Psychology

How we plan: Theory and studies such as [1] suggest that before planning to reach, we, the humans, first find a single target-posture, by constraint-based elimination, then a movement is planned from the current to the target posture. And the target-posture requires us to choose a target-grasp. [2] argues that how we grasp objects depends upon what we plan to do with them. Further it has been shown that initial grasp configuration depends upon the target location from the aspect of task [3], end state comfort [4], object initial and goal positions [5]. Further while deciding the goal position of the object, we take into account various aspects, including the perspective of the person we are interacting with. E.g. Fig. 1 shows two different ways to grasp and hold an object to show it to someone. In both cases, the grasp is valid and the placement in space is visible to the other human, but in Fig. 1(a) the object will be barely recognized by the other person, because the selected grasp to pick the object and the selected orientation to hold the object are not good for this task. We would rather prefer to grasp and hold the object in a way, which makes it significantly visible and also tries to maintain the notions of top and front from other person’s perspective, as shown in Fig. 1(b), which is also supported by HRI studies such as [6]. This suggests three main points: (i) A target-posture should be found before any movement. (ii) It is important to plan pick-and-place as one task, instead of planning and executing them separately. (iii) It is important to take into account the perspective of the human for whom the task is being performed. In this paper we will explore pick-and-place tasks for Human Robot Interactive Manipulation by incorporating these discovered aspects.

B. Related Works in Pick-and-Place Tasks in Robotics

Planning of pick-and-place tasks has been long studied in robotics, such as dynamics simulator [7]; Handey [8] which could back-propagates the constraints for grasp, In [9] constraint on grasping is learnt for the tasks of hand-over, pouring and tool used. In the context of HRI manipulation, it is assumed that either the grasp or to place position and orientation are fixed or known for a particular task, [10], [11]. In addition for human to grasps the object at the same time, robot’s grasp site is just shifted [12] or just enough space is left [9]. These approaches do not synthesize simultaneous grasps by the human and the robot for object of any shape. Further they do not reason from the human’s perspective for reachability, visibility and on effort levels. Also the set of tasks is limited: hand-over or to place, [6], [13].

In this paper first we will identify the key constraints for basic HRI tasks. Then we will present a generic framework, which addresses above mentioned issues, and could plan for basic HRI tasks by incorporating various constraints. It can autonomously decide upon the grasp, the position to place and the placement orientation of the object, depending upon the task, the environment and the human’s perspective while ensuring least feasible effort of the human. We will show its generality by planning tasks of different natures: cooperative tasks such as show, make accessible, and give an object to human, and competitive task to hide an object from human.
II. PROBLEM STATEMENT FROM HRI PERSPECTIVE

We define a task \( T \) to be pick-and-place type as:

\[
\forall T : \text{pick\_and\_place}(T) \Rightarrow (\text{consists\_of}[\text{reach, grasp, carry, place}] : place = \text{put\_on\_support} \lor \text{hold\_in\_space})
\]

Where [ ] is an ordered list. Hence we assimilate ‘holding an object in space’ also as a placement. Fig. 2 shows different decisional components of planning pick-and-place tasks. We recognize two complementary aspects, in coherence with finding (i) of motivation: (i) Pose & Config: First synthesize the configurations \( C \) of the robot's pose and configurations \( O \) of the object. (ii) Traj: Then plan a trajectory between two configurations.

Grasp-Placement interdependency: As shown in Fig. 2, \( C^\text{grasp} \), i.e. how to grasp restricts \( C^\text{robot} \), \( O^\text{object} \), \( P^\text{object} \), i.e. how and where the robot could place the object and vice versa. Hence from the perspective of robot task planning also, pick-and-place should be planned as one task, thus coherent with finding (ii) of motivation.

Further as shown in Fig. 2, we have directly incorporated the finding (iii) of the motivation that the robot should take into account various constraints including the restrictions from the human’s perspective (object’s visibility, reachability, etc.,) affordances (e.g., minimizing human effort), environmental constraints (collision, etc.), task specific requirements (simultaneous grasp, placing on an object, etc.,). We further identify that to-place an object involves: (i) \( O^\text{object} \) i.e. where to place and (ii) \( O^\text{object} \), \( P^\text{object} \), i.e. what should be the orientation of the object. In [14], we have presented a framework for extracting a feasible ‘where’ point to perform place tasks. However there are complementary aspects to [14]: incorporate different possibilities to grasp, different orientations to place and maintain the least feasible effort of the human. The key contribution of this paper addresses these complementary aspects and enables the robot to explicitly takes into account its own constraints as well as the constraints, preferences and effort of the human partner and to plan for both to autonomously synthesize a feasible instance of \( C^\text{robot} \), \( C^\text{grasp} \), \( O^\text{object} \), \( P^\text{object} \). Then we can use any trajectory planner to plan the path between these feasible configurations, such as [15] to obtain a “smooth” trajectory.

III. METHODOLOGY

First we will identify various constraints for pick-and-place task. Then the framework will be presented followed by instantiation through different tasks. \( C \), \( O \) and \( P \) stand for Configuration, Orientation and Position (see Fig. 2).

A. Object Property

The robot maintains geometric information, for each object \( \text{obj} \), it encounters in its lifetime, in the form of tuple:

\[
\text{obj}_\text{prop} = (\text{id}, \text{name}, 3\text{D}_\text{mesh}\_\text{model}, V_F, V_T, \{G^\text{hand}_h\}, O^\text{obj}_\text{place})
\]

Where \( V_F \) and \( V_T \) are manually-provided vectors associated to the symbolic front and top of the object. \( G^\text{hand}_h \) is the set of the possible grasps for hand type \( h \) for \( \text{obj} \). Currently \( h \in \{\text{gripper}_\text{robot} : \text{rg}, \text{hand}_\text{anthropomorphic} : \text{ah}\} \), hence \( n = 2 \). In [16] we have presented our approach to compute grasps of object of any shape (Fig. 3(a)). And \( O^\text{obj}_\text{place} \in \{O^\text{obj}_\text{plane}, O^\text{obj}_\text{space}\} \) is described below.

B. Set of To Place in space orientations \( O^\text{obj}_\text{space}_\text{place} \)

For an arbitrary point in space, the set of object’s orientations are computed by rotating it around its axes.

C. Set of To Place in plane orientations \( O^\text{obj}_\text{plane}_\text{place} \)

The robot can find stable placement of any object on any planar top. The robot generates and stores a set of stable orientations of the object on an imaginary support plane, which is further filtered by the shape of real support during planning. As the object’s shape is modeled as a polyhedron, the stable placement is defined if the projection of object’s center of mass is strictly inside the contact facet. Contact facet \( f \) is a facet of the convex hull of the object, as drawn in blue in Fig. 3(b). This is ‘a’ placement orientation \( O_f \) based on ‘a’ contact facet \( f \). Fig. 3(b) shows different placement orientations with different contact facets. The robot further enriches a particular \( O_f \) by rotating the object along the vertical to get \( O^\text{obj}_\text{plane}_\text{place}_f \). Finally the robot generates the set of all the stable placement orientations for all the \( f \), denoted as \( O^\text{obj}_\text{plane}_\text{place} = \{O^\text{obj}_\text{plane}_\text{place}_f : f_i \in [1, \text{number\_of\_contact\_facets}]\} \).

D. Extracting Simultaneous Compatible Grasps (CG\text{h2,obj}_n)

To facilitate the object hand-over tasks, the robot should be able to reason on how to grasp so that the human could also grasp simultaneously. A grasp pair \( (g_{h_1} \in G_{h_1}, g_{h_2} \in G_{h_2}) \) is simultaneous compatible \( SC \) (Fig. 3(c)) if:

\[
\text{SC}(g_{h_1}, g_{h_2}, \text{obj}) \Rightarrow (\text{apply}(g_{h_1}, \text{obj}) \land \text{apply}(g_{h_2}, \text{obj}) \land (\text{collision}(\text{hand}(h_1), \text{hand}(h_2)) = \emptyset))
\]
E. Constraints based ‘To Place’ positions ($P^{obj}_i\text{Cnts}_{\text{place}}$)

This is to find the positions to put or hold the object. In [14], we have presented the concept of Mighlability Maps (MM), which facilitate 3D grid based multi-state visuospatial perspective taking. The idea is to analyze the various abilities, $A_b \in \{\text{See, Reach}\}$, of an agent not only from her/his/its current state, but also from a set of states, which the agent might attain from the current state. We have further associated an effort level as:

$$E_{\text{see|reach}} \in \{\text{NoEffort, (Head|Arm)Effort, TorsoEffort, Whole_BodyEffort, DisplacementEffort}\}$$

For example if the agent $Ag$ is currently sitting and if $Ag$ has to just lean or turn, it is torso effort, if $Ag$ has to stand up it is whole body effort. As $Ag$ has to move, it is displacement effort. By combining Mighlability Maps with effort levels the robot estimates set of places as: $P^{obj}_i\text{Cnts}_{\text{place}} = \{p_j| p_j = (x,y,z) \land j = 1 \ldots n \land (p_j, \text{holds|view}_c \in \text{Cnts})\}$, $n$ is the number of places. The set of effort constraints $\text{Cnts}_i = \{c_i : i = 1 \ldots m\}$ consists of tuple ($m$ is number of constraints):

$c_i = \langle \text{ability: } A_b, \text{agent: } Ag, \text{effort: } E_{A_b} = (true|false)\rangle$

This enables the robot to find the commonly reachable and visible places for hand-over task, places to put object for hide task, etc. with particular effort levels of the agents.

F. Alignment Constraints on object’s ‘To Place’ orientations from human’s perspective ($AC_{Ag}^{obj,\Phi,\theta}$)

The set of possible orientations to place an object at a particular position $p$ is also restricted based on the visibility of the symbolic features of the object from human’s perspective. Fig. 4(g) shows human-object relative situation. Blue and Green frames represent human’s eye and the object. Frame $F_T$ of the object defines $V_F$ as front direction and $V_T$ as top vector. An object is completely aligned to the agent’s view if: (i) object’s front vector, $V_F$, points towards origin of the human’s eye frame and (ii) object’s top vector, $V_T$, is parallel to human’s eye $H_e$-vector, as shown. Deviation in this alignment could be represented by two parameters $\Phi$ and $\theta$, where $\pm \Phi$ is the angle to rotate the object about $V_T$ of $F_P$ followed by $\pm \theta$, the angle to rotate about $V_F$. The constraint on allowed deviations of the object’s front and top from agent $Ag$ perspective is represented as $AC_{Ag}^{obj,\Phi,\theta}$. The resultant set of orientations at a particular position $p$ after applying alignment constraints is denoted as $Q^{obj,p}_{\text{grasp|place}}$.

G. Alignment constraint of robot’s wrist from human’s perspective ($AC_{Ag}^{\Phi,\theta}$)

We define a tuple $T_{obj}$ for object as: $T^{obj} = \langle \text{grasp: } g, \text{position: } p, \text{orientation: } o\rangle$

The position $p$ to place the object, orientation $o$ of the object at $p$ and the selected grasp $g$ for the object, all together define the wrist orientation of the robot. Similar to alignment constraint on object, constraints on robot wrist alignment from the human’s perspective is used, denoted as $AC_{Ag}^{\Phi,\theta}$.

H. Generating robot’s configurations ($Q^{\text{robot|grasp|place}}$)

For a particular instance of $T^{obj}$ presented above, an inverse kinematics (IK) solver gives the collision-free configuration to grasp or place an object, which is denoted a $Q^{\text{robot}} : (g \rightarrow obj)^p$ read robot’s config after applying grasp $g$ on object $obj$ placed at $p$ with orientation $o$.

I. Constraints on quantitative visibility $VS_{obj} : [\text{min}, \text{max}]$

The robot calculates a visibility score $VS$ of an object $obj$ from an agent $Ag$ perspective as: $VS_{Ag}^{obj} = \frac{N_{obj}}{N_{FOV}}$. $N_{obj}$ is number of pixels of the object in the image of agent’s field of view and $N_{FOV}$ is total number of pixels in that image. Acceptable range of $VS$ is given as $[\text{min}, \text{max}]$.

J. Planning Pick-and-Place tasks: Constraint Hierarchy based approach

The key feature of our planning approach is: introduce right constraint at the right stage. This is also supported by the posture based motion planning model of humans [1], which suggests that candidate postures are evaluated and eliminated by prioritized list of requirements called constraint hierarchy. This elimination by aspect method [17] has been shown to be effective in modeling flexible decision making with multiple constraints [18]. This serves another important purpose: instead of introducing all the constraints at once initially, in the large search space, this approach holds the constraints to be introduced successively at appropriate stages of planning; hence significantly reducing the search spaces before introducing expensive constraints.

We have carefully chosen the constraint hierarchy by taking into account the importance of each constraint, their computation complexity and contribution on the reduction of the search space. Highest priority was given to the human’s effort level (Fig. 5). The planner extracts candidate list of grasps $GL$, to-place positions $PL$ and to-place orientations $OL$.
Fig. 6. Part of the presented generic planner, showing the 4 aspects: (i) How the different candidate lists of Fig. 5 are extracted in blocks 1-A, 4-A and 5-A. (ii) How the candidate triplet $\langle$ grasp: $g$, orientation: $o$, position: $p$ $\rangle$ (blocks 6), are extracted, which in fact could lead to a feasible solution. (iii) Constraint hierarchy: different constraints are introduced at different stages of planning where the search spaces have been reduced significantly. (iv) All the Pose & Config components required for planning a pick-and-place task shown in Fig. 2 have been synthesized, as summarized in block 8.

OL starting with the human’s least effort. Then successively introduces various environment-, planning-, human- and task-oriented constraints at different stages (Fig. 6).

Fig. 6 details the inner block of Fig. 5 and illustrates how different candidate lists GL (block 1-A), PL (block 4-A) and PO (block 5-A) are extracted. It also shows how a particular instance of $T^{obj}$ for picking (block 2-A, 2-B) and for placing (block 6) the object are synthesized. In each green block, if the content at the end sub-block is $\emptyset$, only then the control flows to the next green block, otherwise it iterates appropriately as shown in Fig. 5. This successive introduction of constraint significantly reduces the search spaces at each step. In Block 7 further more expensive constraints are introduced on a particular instance of $T^{obj}$.

The object visibility score at candidate place from the human’s perspective and feasibility of the arm path between the current candidate grasp configuration obtained in block 2-C, and the current candidate place configuration obtained in block 7-A are checked. If the planner succeeds to find the path, it returns with the current candidate Pose & Config, otherwise it iterates appropriately as shown in Fig. 5. In the current implementation, the planner uses [19] to find collision-free paths, blocks 2-D and 7-B. The presented planner is generic in the sense it can find solution for basic human robot interactive manipulation tasks of different natures, when represented in terms of various constraints. Next we will explore such tasks, which are building blocks for planning complex HRI tasks.

### IV. Instantiation for Basic Tasks

Most of the constraints related to IK, collision, human least effort, etc. are common for the HRI tasks. We discuss below some task specific constraints, provided to the presented framework to get a feasible solution.

**Show an object to the human** This task requires grasping an object and holding it in a way so that the human can see it with least feasible effort. But it is not sufficient to hold the object in any orientation. As shown in Fig. 4, showing the toy horse by placing it in the ways shown in (a), or the red bottle shown in (c) and (e) do not reveal much symbolic information from human’s perspective about the object as compared to the one shown in (b) and (d). So, for the task of showing, the constraints on placement are: (i) Front should be visible to the human. (ii) Object should maintain its top upward from human’s perspective. (iii) Maximal parts of the objects should be visible.

These constraints could be imposed to the system by providing appropriate parameters of the object’s alignment constraint $AC_{obj,\phi,\theta}^{ag}$; by allowing a deviation by setting $\theta$ to be $60^\circ$ and then ranking the orientations based on their visibility scores. This value has been chosen arbitrarily to avoid the system to be over-constrained as well as to satisfy the requirements. Fig. 4(c) shows the accepted range of object’s orientations $O^{obj,place}_{p}$ from human’s perspective by using these thresholds, if placed at a particular position $p$. Note that in all these orientations the front is visible and the top is maintained upward from the human’s perspective. The orientations similar to the one shown in Fig. 4(b)
Fig. 7. **Show Object task:** (a) (b) Maximally visible orientation, maintaining object’s front and top: PR2 showing an object, marked in (b), in an orientation to ensure its maximal part is visible, while maintaining the front and top of the object from the human’s perspective. (c) Initial scenario. (d) Selected grasp of higher stability for the case: the constraint on the visibility score of the object at final placement was relaxed. (e) View from the human’s perspective, the object is placed just based on the visible position in the space. The final configuration of the robot itself hides the object from the human. (f) With the constraint on visibility score and to maintain the top upright from the human’s perspective. The planner selected a different feasible grasp. (g)-(j): Views from the human’s perspective. (g) Final placement for case (f). (h) Final placement with the additional constraint of maintaining the object’s front towards the human. (i) Final placement when the constraint to maintain the top was relaxed up to a greater extent. (j) Final placement when the constraints to maintain the top as well as the front were relaxed up to greater extent. Hence constraints and their values restrict the final placements, which influence the initial grasp.

automatically get higher ranking because of visibility of relatively larger part of the object to the human. Similarly to avoid the system to be over constrained we allow a deviation of $\pm 75^\circ$ for the wrist alignment.

**Make an object accessible to the human** The goal is to place an object, which is currently hidden and/or unreachable to the human, on some support plane so that the human can see and take it with least feasible effort. Additional constraint on object orientation to maintain the top upright from the human’s perspective is imposed for this task.

**Give an object to the human** In addition to the constraints of show an object task, the hand-over task imposes the constraint of the simultaneous compatible grasps and reachability by the human with least feasible effort.

**Hide an object from the human** The task is to place the object somewhere on a support plane, so that the human cannot see it, with a particular effort level. There will be no constraint about maintaining the object upright or reachability by the human.

V. **Experimental Results and Analysis**

The system has been tested in simulation and on two real robots of different structures: JIDO and PR2. Objects are identified and localized by stereovision-based tag identification system. The human is tracked by Kinect motion sensor. The human’s gaze is simplified to head orientation obtained through markers-based motion-capture system. **Show Task:**

In Fig. 7(a), PR2 shows an initially-hidden object to the human. The selected grasp and orientation show the inclusion of the constraints of visibility of object’s front while ensuring maximal visibility of the object. Fig. 7(c)-(j) show effect of parameter variation in a different scenario with JIDO. Observe that final placements are avoiding exact alignment of the front or back of the toy horse towards the human, as due to the constraint of maximal visibility such orientations are ranked lower. **Give Task:** Fig. 8(a) shows PR2 giving an object to the human by maintaining the front of the object and the wrist towards the human. Fig. 8(b) shows a different scenario with JIDO. Fig. 8(c) shows the robot’s final configuration to hand over the object to the human.
planner to synthesize the configuration, orientation and position: the key elements for trajectory planning. The problem statement and the constraint hierarchy based planning approach are motivated from human behavioral psychology. The planner circumvents the necessity to provide initial grasp and final placement to the robot, instead autonomously synthesizes those based on the task and the associated constraints. The framework can be used for a variety of tasks by adapting, relaxing or varying the parameters or constraints. It is a step towards incorporating human factors in manipulation planning and developing complex socio-cognitive HRI manipulation behaviors.

**REFERENCES**


