







Uncertainty-Aware Manipulation Planning Using Gravity and Environment Geometry

Felix von Drigalski , Kazumi Kasaura , *Member, IEEE*, Cristian C. Beltran-Hernandez ,
Masashi Hamaya , *Member, IEEE*, Kazutoshi Tanaka , *Member, IEEE*,
and Takamitsu Matsubara , *Member, IEEE*

Abstract—Factory automation robot systems often depend on specially-made jigs that precisely position each part, which increases the system’s cost and limits flexibility. We propose a method to determine the 3D pose of an object with high precision and confidence, using only parallel robotic grippers and no parts-specific jigs. Our method automatically generates a sequence of actions that ensures that the real-world position of the physical object matches the system’s assumed pose to sub-mm precision. Furthermore, we propose the use of “extrinsic” actions, which use gravity, the environment and the gripper geometry to significantly reduce or even eliminate the uncertainty about an object’s pose. We show in simulated and real-robot experiments that our method outperforms our previous work, at success rates over 95%.

Index Terms—Factory automation, manipulation planning, planning under uncertainty, intelligent and flexible manufacturing.

I. INTRODUCTION

ROBOT systems in assembly automation have been used most frequently in high-volume environments such as automobile manufacturing lines, while they are rarely seen in low-volume, high-mix environments, such as small and medium-sized enterprises [1]. One reason for this is the fact that assembly operations require very high precision, due to the tight tolerances that can be smaller than the robots’ accuracy [2]. Currently, the most common way to implement repeatable robotic assembly procedures is by using parts-specific jigs which position the parts with very high precision, and thus allow teach-and-playback programming.

Manuscript received 24 February 2022; accepted 12 August 2022. Date of publication 19 September 2022; date of current version 30 September 2022. This letter was recommended for publication by Associate Editor H. Wang and Editor H. Liu upon evaluation of the reviewers’ comments. (*Corresponding author: Felix von Drigalski.*)

Felix von Drigalski is with the Mujin Inc., Koto, Tokyo 135-0053, Japan, and also with the OMRON SINIC X Corporation, Bunkyo, Tokyo 113-0033, Japan (e-mail: FvDrigalski@gmail.com).

Kazumi Kasaura, Masashi Hamaya, and Kazutoshi Tanaka are with the OMRON SINIC X Corporation, Bunkyo, Tokyo 113-0033, Japan (e-mail: kazumi.kasaura@sinicx.com; masashi.hamaya@sinicx.com; kazutoshi.tanaka@sinicx.com).

Cristian C. Beltran-Hernandez is with the OMRON SINIC X Corporation, Bunkyo, Tokyo 113-0033, Japan, and also with the Robotic Manipulation Lab, Osaka University, Suita, Osaka 565-0871, Japan (e-mail: cristianbehe@gmail.com).

Takamitsu Matsubara is with the Robot Learning Laboratory, Nara Institute of Science and Technology, Nara 630-0192, Japan (e-mail: takam-m@is.naist.jp).

The source code was made public at github.com/omron-sinicx.

This letter has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2022.3207565>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2022.3207565

However, these jigs are expensive, to the point that their cost including engineering and system integration can make up over 50% of the total cost [3]. Their versatility is also limited, since only few jigs can be placed within reach of each robot arm and the changeover to a different product line incurs additional costs. It would be highly desirable for robot systems to handle a large variety of parts without the need to change jigs.

Programming robots without jigs poses difficulties, such as noisy pose estimation from parts detection which can introduce misalignment of the grasped parts and result in task failure. For the World Robot Summit Assembly Challenge 2018 and 2020, we developed a robotic system composed of multiple robot arms and generic grippers aiming to realize jig-less assembly [4]. Although our system won 4th and 3rd place in 2018 and 2021 respectively, our robots (and those of the other winning teams) frequently failed insertion tasks due to position uncertainty.

Previously, we developed an in-hand pose estimation method in which the robot touches the environment with the grasped part to reduce the uncertainty of the part’s pose to sub-mm precision [5] (the *Touch action*). We extended this by also evaluating the effect of a camera view on the object pose uncertainty (the *Look action*) and allowing a combination of both actions [6]. However, these methods were limited: the user had to define the action sequence manually, and the actions required either slow motions and a force sensor (for the *Touch action*), or careful camera calibration (for the *Look action*).

In this work, we solve both of those problems by first extending our method with three new actions (*Grasp*, *Place*, *Push*) which require no force sensor or camera calibration, and which are efficient and fast as they take advantage of both gravity and the environment geometry. Furthermore, we formulate a planner that selects action sequences which reduce or even eliminate object pose uncertainty, thus enabling reliable manipulation planning for tasks that require high precision, such as industrial assembly.

We show that our proposed method performs reliably in simulation for 13 parts used in a model assembly task, as well as in a real robot system. Using the three new actions, the uncertainty of many objects can be eliminated without the use of any sensors. The only constraints on the object are that the robot needs to be able to grasp it stably, that it may not topple or roll during pushing, and that it may not roll away when placed (either by using a placing surface of appropriate softness or by disallowing spheres and cylinders).

Thus, our contributions are as follows:

- The formulation of five actions that reduce the pose uncertainty of an object
- The calculation of the uncertainty that results from each action
- The planner that minimizes uncertainty using these actions.

II. RELATED WORKS

A. Uncertainty-Aware Manipulation

Creating “robust” manipulation plans that succeed even under uncertainty (also called “open-loop,” “oblivious” or “sensorless” manipulation plans) has a long history. Mason described the “Place” action as early as 1982 [7], and investigated the effect of pushing actions. Erdmann and Mason [8] showed that sensorless manipulation can arrange planar objects into a known state. Goldberg [9] described an algorithm to obtain a sequence of grasps that orients a planar polygonal object without knowledge of the initial orientation. They assume infinitely large grippers and no friction. Zhou et al. [10] formulated a planner for planar grasping sequences that allow arbitrary shapes with bounded uncertainty to be arranged into a known state, by modelling friction and the contact between gripper and object. In contrast to the works above, our method is not restricted to 2D, but works with 3D meshes that are widely used in motion planning and collision checking libraries. Allowed grasps can be defined anywhere on the object, and the uncertainty is minimized for 3D poses.

Lozano-Pérez et al. [11] proposed “pre-image backchaining,” an algorithm to obtain a robust manipulation (“fine-motion”) plan by defining permissible regions of space backwards from the goal position, from which a motion will certainly achieve its goal. To avoid having to quantify these regions and because the search space is small and can be exhausted easily, our method simply applies a breadth-first search and samples actions from the initial pose and uncertainty.

More recent approaches propose machine learning methods to handle uncertainty. Kahn et al. proposed learning a collision probability using a predictive model for collision avoidance of mobile robots [12]. Kou et al. developed a model-based reinforcement learning approach, where the maximum control input value was scaled down with high model uncertainty [13]. Lee et al. proposed to learn switching a model-based and learning-based controller with low and high uncertainty, respectively [14]. Unlike these approaches, our method does not need to collect data or train machine learning models. This is very desirable for users.

B. Active Perception

Many studies have explored active perception approaches where robots interact with objects to determine the objects’ location or properties [15]. In tactile exploration, a probabilistic model has been learned for the target object’s surface which efficiently identified the shape by touching to reduce the model uncertainty [16], [17], [18]. Koval et al. and Wirnshofer et al. proposed particle filter-based approaches to estimate object pose

by active contacts [19], [20]. We also use particle sampling in some of our actions.

Paolini and Mason learned a statistical model for an object by pick and place regrasping [21]. However, our approach requires no data collection. Chavan-Dafle and Rodriguez demonstrated other actions utilizing the environment [22] as well as stable regrasping without fixtures by sliding the object in the gripper [23]. These approaches are the most similar to our actions, but they do not consider the uncertainty around the object, and they assume that an object can slide within the gripper. Our method explicitly tracks uncertainty, and uses gravity to obtain a stable pose for the object, which is more generally applicable than in-hand sliding.

While most approaches have focused on a single action or several objects, our method combines 5 different actions that apply to a wide range of parts.

C. Object Pose Estimation Using Multi-Modal Sensors

Numerous computer vision and pattern recognition works have developed not only 2D [24] but also 6D pose estimation techniques using RGB-D information [25]. CAD information and deep learning have been used to improve robustness [26], [27]. Although vision-based methods are helpful to determine object poses, they are affected by lighting conditions and materials. Our method is particularly robust in the context of assembly tasks, where metallic parts with reflection may degrade the estimation performance.

Other sensors or sensor fusion for object pose estimation have also been used. For example, tactile sensors have been shown to improve grasped object localization [28], and in-hand object localization approaches have combined vision, force-torque, and joint angles sensors to improve estimation performance [29], [30]. However, while these multi-modal methods improve pose estimation, they do not quantify the measurement uncertainty of the observations. Our method reduces the remaining uncertainty by tracking and minimizing it explicitly.

III. PROPOSED METHOD

In our proposed method, the robot reduces the object pose uncertainty by executing a sequence of actions obtained by a planner (Subsection E). We use five type of actions: *Touch*, *Look*, *Place*, *Grasp* and *Push*, of which the last three are newly added in this letter. The pose belief is represented by a Gauss distribution (Subsection A). The planner’s objective is to minimize its size (Subsection D). To produce effective sequences, the planner must calculate the effect of each action on the pose belief. To calculate the last three types of actions, we sample particles from the distribution to obtain discrete poses without uncertainties (Subsection B), for which the effect can be calculated geometrically (Subsection C).

The first two actions (*Touch* and *Look*) are described in [5], [6]. The *Look* action requires either a calibrated camera or a calibration geometry in the image, while the *Touch* action requires the position of at least one calibrated support surface (and optionally one edge) in the environment. The remaining three are *extrinsic* manipulations using the interaction between the object and the gripper geometry and/or supporting surface.

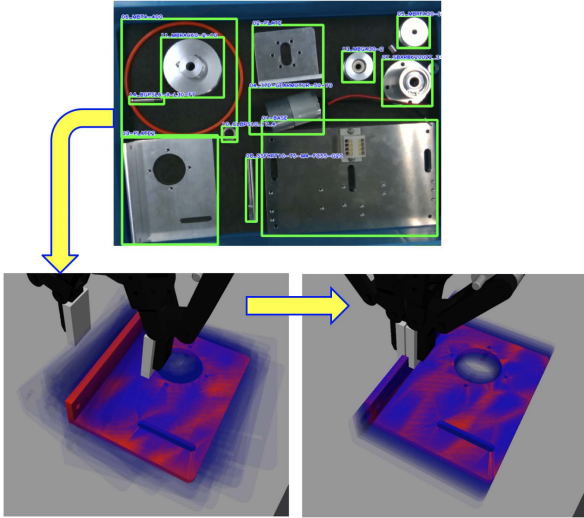


Fig. 1. A robot grasping an L-plate based on a noisy vision measurement. Our method calculates actions that reduce the uncertainty (blue) around the object pose estimate (red). **Top:** The vision input. **Bottom left:** The estimated pose (red) with uncertainty (blue) representing alternate possibilities. **Bottom right:** The grasp action restricts the possible locations. All remaining possibilities are restricted by the gripper.

They require only the support surface and no sensor. The *Place* action consists of placing the grasped object on a support surface with known height. *Grasp* actions consist of grasping an object resting on a support surface, as shown in Fig. 1. *Push* actions consist of pushing the object using the gripper geometry.

The method assumes that 1) the geometry is provided as a mesh or collection of bodies, 2) friction can be neglected until the object is grasped, and 3) the object does not topple or roll during pushing.

A. Representation of Pose Beliefs

To model the belief about the object pose, we follow the method described in [31]. We represent a pose $T \in SO(3)$ with uncertainty as a mean pose $\bar{T} \in SO(3)$ with a small perturbation by a random variable $\xi \in \mathbb{R}^6$:

$$T = \exp(\xi^\wedge) \bar{T}, \quad (1)$$

where $\wedge : \mathbb{R}^6 \rightarrow \mathfrak{se}(3)$ is defined by

$$\xi = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} \mapsto \begin{pmatrix} 0 & -x_6 & x_5 & x_1 \\ x_6 & 0 & -x_4 & x_2 \\ -x_5 & x_4 & 0 & x_3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Let $\vee : \mathfrak{se}(3) \rightarrow \mathbb{R}^6$ be the inverse of \wedge .

We assume that the variable ξ follows the zero-mean Gaussian distribution $\mathcal{N}(\mathbf{0}, \Sigma)$ where Σ is the covariance matrix. Thus, we represent a pose belief using the mean pose \bar{T} and the covariance matrix Σ of the perturbation variable.

B. Calculating Pose Uncertainty

Each action reduces the pose uncertainty in different ways. The calculation of the Touch and Look action is described in our previous work [5], [6]. For the three extrinsic manipulations, we assume that the pose distribution after the action can be approximated by the representation in §III-A. For this assumption, it is required that the function from poses before the action to poses after the action is continuous in the region of uncertainty, but not necessarily C^1 -class. Note that it may be discontinuous when an extrinsic action is performed badly (for example, placing a pole on its end in an unstable position), but our planner (§III-E) avoids such actions.

We calculate the effect of the extrinsic manipulations on pose beliefs as follows. We first generate the pose particles T_1, \dots, T_N by (1) where N is the number of particles. Then, we calculate the poses T'_1, \dots, T'_N after the manipulation for each particle using the methods described in the next subsection. Finally, we estimate the new pose belief (\bar{T}', Σ') as a distribution fitting the samples T'_1, \dots, T'_N .

To calculate \bar{T}' , we solve the equation

$$\sum_{i=1}^N \ln(T'_i (\bar{T}')^{-1})^\vee = \mathbf{0} \quad (2)$$

using the Newton method. More precisely, let \bar{T}'_s be the current approximation of the solution of (2) and let $\xi'_{s,i} = \ln(T'_i (\bar{T}'_s)^{-1})^\vee$. Then we can let $\bar{T}'_{s+1} = \exp(\hat{\eta}'_s) \bar{T}'_s$ be the next approximation, where

$$\eta'_s = -\mathcal{J}_s^{-1} \sum_{i=1}^N \xi'_{s,i} \quad (3)$$

and \mathcal{J}_s is the Jacobian of the map

$$\eta \mapsto \sum_{i=1}^N \ln(\exp((\xi'_{s,i})^\wedge) \exp(-\eta^\wedge))^\vee,$$

which can be calculated using (33) and (34) in [31].

After calculation of \bar{T}' , Σ' is calculated by

$$\Sigma' = \frac{N}{N-1} \sum_{i=1}^N \xi'_i (\xi'_i)^\top, \quad (4)$$

where $\xi'_i = \ln(T'_i (\bar{T}')^{-1})^\vee$.

C. Calculating Poses After Extrinsic Manipulations

Apart from the uncertainty, we must also calculate the mean pose resulting from the manipulation. For all extrinsic actions, we assume that the grasped object is a polyhedron and its center of gravity is known. We ignore friction between both the object and the gripper, and the object and the support surface.

In the following sections, the object's phases of rotation correspond to Fig. 2.

1) *Place Action:* Let C be the projected point of the center of gravity of the object onto the support surface. Since we ignore friction between the object and the support surface, the object cannot topple and the position of C in the support polygon does

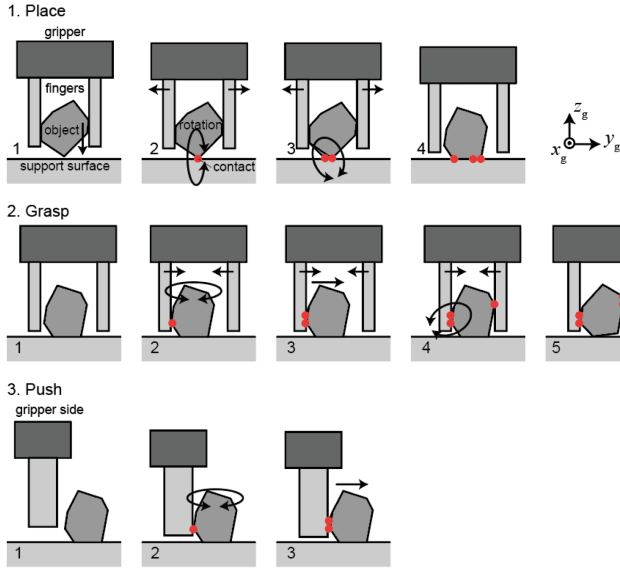


Fig. 2. Rotation phases of an object induced by each extrinsic manipulation.

not change. Thus, it is sufficient to calculate the rotation of the object as it is being placed.

The rotation of the object while it is being placed on the support surface is divided into the following four phases:

- 1) When no vertex of the object is on the support surface, it descends vertically without rotating.
- 2) When only one vertex of the object is on the support surface, the object rotates around the line on the support surface which is orthogonal to the line connecting the vertex and C .
- 3) When two vertices of the object are on the support surface and there exists a perpendicular line from C to the segment connecting the two vertices, the object is rotated around the line connecting the two vertices.
- 4) When more than three vertices are on the support surface, and if C is inside of the convex hull of the vertices of the object on the surface, the object is placed stably. Otherwise, this place action is regarded as unstable.

Note that the condition for the third case is always satisfied when the object is placed on the face of an acute-angle triangle or a rectangle. As the rotation can be approximated by the described method even when the condition does not apply, we use the calculation for all cases.

2) *Grasp Action:* We assume a parallel gripper with two flat gripper pads (“fingers”). We consider the fingers as parts of two parallel planes which are orthogonal to the support surface and move with the direction of their normal lines. We define three orthogonal axes x_g, y_g, z_g as displayed in Fig. 2, where x_g is parallel to both the support surface and the fingers, y_g is orthogonal to the fingers, and z_g points upwards, away from the support surface. When the initial pose of the object is near the stable pose and the center of gravity is not very far from the grippers, the object’s change of position with respect to x_g is negligible. The position in y_g is fixed by the fingers and the position in z_g is determined by the support surface. Thus, we only need to determine the rotation of the object during grasping.

The rotation is divided into the following phases, as the gripper closes:

- 1) When no finger touches the object, it does not rotate.
- 2) When one vertex touches one of the fingers or two vertices touch one finger each, it rotates around z_g .
- 3) When two vertices touch one of the fingers but no vertex touches the other pad, it is only pushed and does not rotate.
- 4) When two vertices touch one of the fingers and one vertex touches the other finger, it rotates around the line connecting the two vertices on the same side.
- 5) When three or more vertices touch one finger and at least one vertex touches the other finger or two vertices touch each finger, the object is grasped stably if there is an intersection of the convex hull of the vertices touching one finger and that of the other finger. Otherwise the grasp is considered unstable.

Note that the third phase is not always passed. Note also that, in the conditions of the second, third, and fourth phases, vertices with the same x - and y -coordinate are considered as one vertex.

3) *Push Action:* We use the same coordinates as for the grasp action. When the push action is executed, the fingers are closed and the object is pushed within the direction of the x_g -axis. The reasoning of the grasping case applies analogously, so we only need to calculate the rotation.

The rotation is divided into the following phases:

- 1) When the fingers do not touch the object, it does not rotate.
- 2) When at most one vertex touches the fingers, the object rotates around z_g .
- 3) When two vertices touch the fingers, the object is only pushed and does not rotate.

Note that, as in the grasp action, vertices with the same x -coordinate and y -coordinate are considered as one vertex.

D. Quantification of Uncertainty

To allow the planner to minimize the uncertainty, we need to quantify it as a scalar. We use the alternative form of (1):

$$T = \bar{T} \exp(\xi'^{\wedge}). \quad (5)$$

Let Σ' be the covariance matrix of ξ' . The transformation from Σ to Σ' can be calculated by the coordinate change formula (26) in [31]. We define the amount of uncertainty by

$$\text{score}(\bar{T}, \Sigma) = \sum_{i=1}^6 \sum_{j=1}^6 c_{i,j} \Sigma'_{i,j}, \quad (6)$$

where $c_{i,j}$ are the fixed coefficients.

The reason we transform Σ to Σ' is that Σ' is associated with the object’s own coordinate system. The first three diagonal components of Σ' correspond to the object’s position uncertainty in x , y and z , the second three to the orientation uncertainty. The non-diagonal components correspond to the covariances between them.

Some objects have rotational symmetry around an axis, around which uncertainty cannot be eliminated. In these cases, we set the coefficients $c_{i,j}$ such that unavoidable uncertainty is ignored and the plan succeeds in minimizing the rest. If we

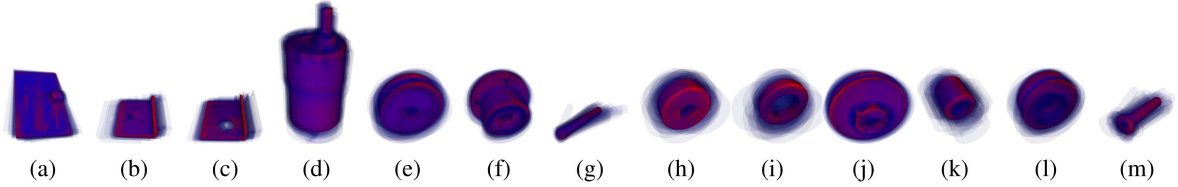


Fig. 3. All parts with initial uncertainties for the planning experiment. (a) base plate, (b) panel motor, (c) panel bearing, (d) motor, (e) motor pulley, (f) bearing, (g) shaft, (h) end cap, (i) bearing spacer, (j) output pulley, (k) idler spacer, (l) idler pulley, (m) idler pin. The red shape represents the pose belief, the blue shapes random samples from the probability distribution.

did not use Σ' instead of Σ , this would be impossible, as the symmetry axis depends on \bar{T} .

E. Planning

To obtain the action sequence from the planner, we first enumerate the possible actions.

For the Place action, we enumerate all faces of the three dimensional convex hull of the object as possible placement orientations. For the Grasp action, we define the grasp poses for each object by hand for simplicity and calculation speed, but any grasp-generating algorithm could be applied. For the Push action, we enumerate the edges of the projected object on the support surface as pushing directions. For round objects, we randomly choose a lower number of directions when there are too many.

We limit the number of possible actions to around 30, and the lengths of action sequences to 3 because this is sufficient for the majority of objects in our experiments. As the size of search space is $O(N^d)$ where N is the upper bound of the number of possible actions and d is the depth of search, it is small enough that we can use breadth first search. At each step, we select the next action and calculate the resulting uncertainty. For the Touch and Look action, we assume that the object is observed at the expected mean pose, and that the uncertainty decreases according to the selected action.

IV. PLANNING EXPERIMENTS

To verify that A) our planner gives valid results and B) the performance of our newly proposed actions (Place, Grasp and Push) compared to the previous ones (Touch and Look), we generated action sequences for a variety of objects and evaluated the resulting uncertainty.

A. Setup

We evaluated the resulting plans for the following conditions:

- We created plans for 13 different objects that were used in the WRS 2020. See Fig. 3, which shows the parts with the initial uncertainties.
- We defined three different sets of actions which the planner can perform, to compare the new and previously proposed actions:
 - 1) All five actions
 - 2) Extrinsic actions, i.e., Place, Grasp and Push actions
 - 3) Static actions, i.e., Touch and Look actions

- The action sequences contain between 0 and 3 actions. At the end, the object must be grasped.

The initial pose with uncertainty is generated as follows: First, the object has pose I (identity matrix) with uncertainty Σ such that

$$\Sigma_{i,j} = \begin{cases} 0.00001[\text{m}^2] & (i,j) = (1,1), (2,2), (3,3) \\ 0.001 & (i,j) = (4,4), (5,5), (6,6) \\ 0 & \text{otherwise.} \end{cases}$$

These values represent a standard deviation of $\sqrt{\Sigma_{i,j}} = 3.16 \text{ mm}$ in translation and 0.0316 radians in rotation. This means that 99% (three standard deviations) of sampled particles will be within 9.5 mm and 0.95 radians (5.4 degrees) of the reference pose. Then, the object is grasped and the uncertainty is calculated. This grasped pose with uncertainty is the initial pose.

For objects which have no rotation symmetry (base, panel motor, panel bearing, motor), the amount of uncertainty is defined according to (5) with the coefficients:

$$c_{i,j} = \begin{cases} 10000[\text{m}^{-2}] & (i,j) = (1,1), (2,2), (3,3) \\ 100 & (i,j) = (4,4), (5,5), (6,6) \\ 0 & \text{otherwise.} \end{cases}$$

For objects which are symmetric around the x-axis (motor pulley, bearing, shaft, end cap, bearing spacer, output pulley, idler spacer, idler pulley and idler pin), we ignore the rotation uncertainty around the x-axis by defining the uncertainty as in (5) with the coefficients:

$$c_{i,j} = \begin{cases} 10000[\text{m}^{-2}] & (i,j) = (1,1), (2,2), (3,3) \\ 100 & (i,j) = (5,5), (6,6) \\ 0 & \text{otherwise.} \end{cases}$$

Note that although the output pulley (Fig. 3(e)) is not strictly symmetric, it is close enough that we choose to disregard its rotation uncertainty.

B. Results

The result of the planning experiments are displayed in Table I. The box-plots showing the amount of remaining uncertainty (after 3 actions) are displayed in Fig. 4

When planning with all actions, the results are best, and uncertainty is eliminated or severely reduced, generally to sub-mm levels.

TABLE I
THE REMAINING UNCERTAINTY AFTER DIFFERENT ACTION SEQUENCES

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(m)	(m)
All	0.4983	1.1948	2.2824	1.6638	0.3440	0.3896	0.3793	0.3379	0.3407	2.1745	0.3289	0.3466	0.3122
	0.2507	0.7935	1.4756	1.2526	0.2952	0.3574	0.2824	0.2934	0.3004	0.6000	0.2830	0.3144	0.3050
	0.0388	0.0362	0.0549	0.9321	0.0003	0.0004	0.2000	0.0007	0.0013	0.0294	0.0003	0.0001	0.0002
	0.0035	0.0000	0.0000	0.6867	0.0000	0.0000	0.1461	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000
Extrinsic	0.4983	1.1948	2.2824	1.6638	0.3440	0.3896	0.3793	0.3379	0.3407	2.1745	0.3289	0.3466	0.3122
	-	-	-	-	-	-	-	-	-	-	-	-	-
	0.0388	0.0362	0.0549	-	0.0003	0.0004	-	0.0007	0.0013	0.0294	0.0003	0.0001	0.0002
	0.0216	0.0000	0.0000	-	0.0000	0.0000	-	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000
Static	0.4983	1.1948	2.2824	1.6638	0.3440	0.3896	0.3793	0.3379	0.3407	2.1745	0.3289	0.3466	0.3122
	0.2507	0.2812	1.4756	1.2526	0.2952	0.3574	0.2824	0.2934	0.3004	0.6000	0.2830	0.3144	0.3050
	0.1544	0.1899	0.9638	0.9321	0.2976	0.3380	0.2000	0.2696	0.2588	0.3979	0.2220	0.2778	0.3259
	0.1104	0.1294	0.6532	0.6867	0.2802	0.3419	0.1461	0.2168	0.1944	0.0796	0.2008	0.2678	0.3260

“All” means all five actions. “Extrinsic” means place, grasp, and push actions. “Static” means touch and look actions. The values correspond to actions number 0, 1, 2, 3 (top to bottom) and represent the amount of remaining uncertainty as defined in (5). The columns correspond to the 13 parts and the rows to the set of actions. Hyphens mean that no plan can be made. Note that since the object must be grasped at the end of the action sequence, valid action sequences may only be possible for e.g. 1 and 3 actions but not 2.

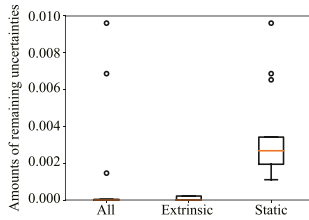


Fig. 4. Box plot showing the amount of remaining uncertainty for each set of actions, for 13 different objects.

The results show that the extrinsic manipulations are responsible for reducing the uncertainty to zero, as this occurs for 8 out of 13 parts (panel motor, panel bearing, motor pulley, bearing, end cap, bearing spacer, idler spacer, idler pulley and idler pin) even when planning only with extrinsic manipulations. For two parts (motor and shaft) the planner could not find a valid plan using only the extrinsic manipulations, as the parts have no faces to place them stably. For the base, the reason why uncertainty remains is that while a face to place it exists, it is so thin that the placing can fail. For the output pulley, the push action can sometimes fail because due to its protruding part, the center of mass and the center of the circle in the outline do not coincide, so some uncertainty remains.

When planning only with static manipulations, uncertainty is reduced for all parts, but never reduced to zero.

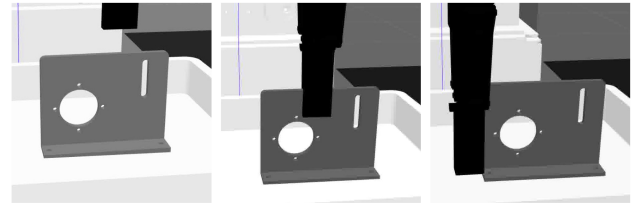
In conclusion, the extrinsic manipulations are powerful but not usable for all objects, while the static manipulations are effective regardless of the grasped object.

V. SIMULATION

To verify that the action sequences are valid and can be executed, we first conducted simulation experiments.

We used Gazebo as the simulation environment and simulated the same robot that was used in the real-robot experiment. We choose the bearing plate (see Fig. 3(c) as the target object. Friction parameters μ and μ_2 were set to 0.5 for the bearing plate and 1.0 for the gripper fingers and support surface.

Both in simulation and the real robot experiments, we use gripper fingers with a width of 17 mm and a thickness of 6 mm.



(a) Initial pose

(b) Grasp

(c) Push

Fig. 5. An action sequence in the Gazebo simulator. After one grasp and push, the uncertainty has been reduced to zero and the object pose is known.

A. Setup

For this experiment, we obtained an action sequence from our planner which satisfies the following conditions:

- At the initial state, the panel is placed with uncertainty as shown in Fig. 5(a). The uncertainty is represented by a covariance matrix Σ such that

$$\Sigma_{i,j} = \begin{cases} 0.0001[\text{m}^2] & (i,j) = (1,1), (2,2) \\ 0.0001 & (i,j) = (6,6) \\ 0 & \text{otherwise.} \end{cases}$$

- At the final state, the panel must be placed in the same orientation and have no uncertainty.

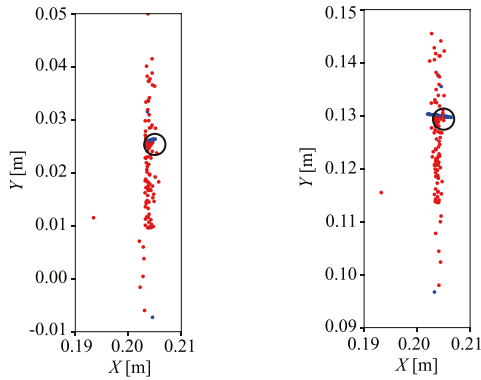
The sequence consists of the following actions: 1) Grasp the plate as shown in Fig. 5(b), 2) Place the plate at the same position, and 3) Push the plate horizontally as shown in Fig. 5(c).

As a control, we also simulate a manually created action sequence that leaves a dimension uncertain: 1) Same initial state as above, 2) Grasp the plate, and 3) Place the plate at the same position as the above sequence.

After the above action sequences, we evaluate the resulting object positions.

B. Results

Fig. 6 compares the final position of the two screw holes at the bottom of the bearing plate, which are critical for task success. If both hole centers are inside of the expected circles, the position



(a) Position of the left hole (b) Position of the right hole

Fig. 6. Results from simulation: final position of the two screw holes at the bottom of the placed plate. Blue points show the planned sequence and red points the control sequence (grasp-and-push). The control sequence shows remaining uncertainty, while the planned sequence is within the target circle in 97% of cases. The circle radius is 2 mm.

of the panel is considered acceptable. For the planned action sequence, placement was successful 97% of the time, while the control sequence only succeeded in 15% of cases.

As expected, both the planned sequence and the control sequence decrease the scattering in x to a few millimeters, but only the planned sequence removes it in y .

Some noise remains even after executing the planned sequence, especially for the right hole. This is likely due to numerical instabilities in the gripper model and the contact between the gripper and the plate. Without this effect, we would expect a 100% success rate in simulation.

VI. REAL ROBOT EXPERIMENT

To confirm the effectiveness of our method and the new actions (Push, Place, Grasp) that have been added to the Touch and Look actions, we conducted a real-robot experiment based on the actual assembly task at the World Robot Summit Assembly Challenge 2020. The experiment recreates the placement of the two L-shaped panels, the bearing, and the shaft (shown in Fig. 3(b), 3(c), 3(f), and 3(g), respectively). We evaluated the precise placement of the through-holes of the L-shaped plates with the screw holes of the base, and the alignment of the bearing and shaft.

A. Experimental Setup

We recorded a total of 480 trials. At the start of each trial, the parts were placed at an initial position, defined by a set position P , translated by a random offset within ± 10 mm in each horizontal axis and rotated by a random angle within ± 15 degrees around the vertical axis. The random offset and rotation represent a significant amount of noise in an object detection pipeline.

As seen in Fig. 5, for the motor panel and bearing panel, the robot first performs a grasp action centering and then a push action. For the bearing and shaft, the robot performs several grasp actions. As this sequence should leave the object pose

well defined without any remaining uncertainty, the robot then picks up the part and places it at the target location. The hole alignments are then recorded with an RGB camera, either from above or from the side, and the distance from the reference position estimated manually from the image. The robot then places the parts at a new random location to prepare for the next trial.

B. Results

Of 120 trials each, the motor panel, bearing panel, bearing, and shaft were placed successfully within 1 mm of the target location 106, 114, 115, and 120 times, respectively. The bearing panel and part were placed three times, the motor panel 15 times within 3 mm of the target location but at more than 1 mm distance. We recorded these offsets as a failure due to the high precision requirements for these applications, although this error can be permissible. Three and two trials for the bearing panel and bearing were not counted due to unrelated issues (e.g., motion planning issues or latency). In total, the parts were successfully placed at the target location in 95.7% of the trials (455 of 475).

All of the failures were due to mechanical problems that can be mitigated, such as the gripper surfaces sticking to the part and pulling it away from the assumed position as the gripper opened.

VII. DISCUSSION

The simulated and experimental results show that our method produces feasible action sequences that reduce or eliminate the uncertainty of an object pose, and that they lead to repeatable results on a real robot preparing a representative model assembly task.

In almost all cases, the Extrinsic actions we proposed (*Grasp*, *Place*, *Push*) reduce uncertainty more than the *Touch* and *Look* actions. This makes intuitive sense, as aligning an object with a surface (table or gripper pads) constrains multiple degrees of freedom, while the *Touch* action constrains only a single point. They are also easier to calibrate than the *Look* action, as they require only the position of a flat surface instead of complete camera calibration.

Due to size and measurement hardware constraints, we could only perform real robot experiments for a limited amount of objects and action combinations. However, previous experiments [5], [6] have shown that even using only *Touch* and *Look* actions, uncertainty can be reduced to sub-mm levels, thus confirming the simulation results.

Our method has some limitations. We make a number of simplifications, most importantly that friction has no significant effect. For the gripper, this can be reproduced by adding a linear guide to one of the fingers, or simply starting close to a stable grasp's position. For the *Push* action, the motion needs to be slow enough to be quasi-static, and the part may not topple. The *Push* action also assumes that the pushing part of the gripper is a flat surface, which is permissible for many but not all industrial grippers.

We have omitted grasp generation and stability evaluation for the *Push* and *Grasp* actions, as for industrial use it is often preferable to explicitly define permissible interactions (predictable

behavior, easier tuning than e.g. friction parameters). However, automatic grasp generation and push stability evaluation can easily be added on top of our method, by using them to define the permissible grasps and pushes.

A limitation of the *Place* action is that it assumes a stable placement pose, which does not apply for meta-stable objects which can roll, such as cylinders or balls. In practice, a layer of foam can inhibit the rolling enough for the action to work.

VIII. CONCLUSION

We presented a method to generate a sequence of actions that reduces or eliminates the uncertainty about an object pose. In real-robot experiments, the sequence resulted in sub-mm precision over 95% of the time, and in <3 mm precision 99% of the time. In simulation, we showed that the “extrinsic” actions proposed in this letter (which use gravity, the environment and the gripper geometry) improve the uncertainty reduction significantly over our previous work, and can eliminate uncertainty entirely. Our method can be used to precisely position objects without custom jigs, or it can be implemented into existing Task and Motion Planning frameworks to increase repeatability and resilience to noise.

The code has been released at github.com/omron-sinix/uncertainty-aware-manipulation-planning.

REFERENCES

- [1] F. Yang, K. Gao, I. W. Simon, Y. Zhu, and R. Su, “Decomposition methods for manufacturing system scheduling: A survey,” *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 389–400, Mar. 2018.
- [2] M. Suomalainen, Y. Karayiannidis, and V. Kyriki, “A survey of robot manipulation in contact,” *Robot. Auton. Syst.*, vol. 156, 2022, Art. no. 104224.
- [3] Y. Yokokohji et al., “Assembly challenge: A robot competition of the industrial robotics category, world robot summit—summary of the pre-competition in 2018,” *Adv. Robot.*, vol. 33, no. 17, pp. 876–899, 2019.
- [4] F. Von Drigalski et al., “Team O2AS at the world robot summit 2018: An approach to robotic kitting and assembly tasks using general purpose grippers and tools,” *Adv. Robot.*, vol. 34, no. 7/8, pp. 514–530, 2020.
- [5] F. von Drigalski et al., “Contact-based in-hand pose estimation using bayesian state estimation and particle filtering,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 7294–7299.
- [6] F. von Drigalski et al., “Precise multi-modal in-hand pose estimation using low-precision sensors for robotic assembly,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 968–974.
- [7] M. T. Mason, “Manipulator grasping and pushing operations,” *AI Tech. Rep.*, Massachusetts Institute of Technology, Cambridge, MA, USA, Tech. Rep. AITR-690, 1982.
- [8] M. A. Erdmann and M. T. Mason, “An exploration of sensorless manipulation,” *IEEE J. Robot. Automat.*, vol. 4, no. 4, pp. 369–379, Aug. 1988.
- [9] K. Y. Goldberg, “Orienting polygonal parts without sensors,” *Algorithmica*, vol. 10, no. 2, pp. 201–225, 1993.
- [10] J. Zhou, R. Paolini, A. M. Johnson, J. A. Bagnell, and M. T. Mason, “A probabilistic planning framework for planar grasping under uncertainty,” *IEEE Robot. Automat. Lett.*, vol. 2, no. 4, pp. 2111–2118, Oct. 2017.
- [11] T. Lozano-Perez, M. T. Mason, and R. H. Taylor, “Automatic synthesis of fine-motion strategies for robots,” *Int. J. Robot. Res.*, vol. 3, no. 1, pp. 3–24, 1984.
- [12] G. Kahn, A. Villafior, V. Pong, P. Abbeel, and S. Levine, “Uncertainty-aware reinforcement learning for collision avoidance,” 2017, *arXiv:1702.01182*.
- [13] C.-Y. Kuo, A. Schaarschmidt, Y. Cui, T. Asfour, and T. Matsubara, “Uncertainty-aware contact-safe model-based reinforcement learning,” *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 3918–3925, Apr. 2021.
- [14] M. A. Lee et al., “Guided uncertainty-aware policy optimization: Combining learning and model-based strategies for sample-efficient policy learning,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 7505–7512.
- [15] J. Bohg et al., “Interactive perception: Leveraging action in perception and perception in action,” *IEEE Trans. Robot.*, vol. 33, no. 6, pp. 1273–1291, Dec. 2017.
- [16] S. Dragiev, M. Toussaint, and M. Gienger, “Uncertainty aware grasping and tactile exploration,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2013, pp. 113–119.
- [17] M. Björkman, Y. Bekiroglu, V. Högman, and D. Kragic, “Enhancing visual perception of shape through tactile glances,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 3180–3186.
- [18] T. Matsubara and K. Shibata, “Active tactile exploration with uncertainty and travel cost for fast shape estimation of unknown objects,” *Robot. Auton. Syst.*, vol. 91, pp. 314–326, 2017.
- [19] M. C. Koval, N. S. Pollard, and S. S. Srinivasa, “Pose estimation for planar contact manipulation with manifold particle filters,” *Int. J. Robot. Res.*, vol. 34, no. 7, pp. 922–945, 2015.
- [20] F. Wirmshofer, P. S. Schmitt, P. Meister, G. v. Wichert, and W. Burgard, “State estimation in contact-rich manipulation,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 3790–3796.
- [21] R. Paolini and M. T. Mason, “Data-driven statistical modeling of a cube regrasp,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 2554–2560.
- [22] N. Chavan-Dafle et al., “Extrinsic dexterity: In-hand manipulation with external forces,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2014, pp. 1578–1585.
- [23] N. Chavan-Dafle and A. Rodriguez, “Prehensile pushing: In-hand manipulation with push-primitives,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 6215–6222.
- [24] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework,” *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.
- [25] Y. Konishi, Y. Hanzawa, M. Kawade, and M. Hashimoto, “Fast 6D pose estimation from a monocular image using hierarchical pose trees,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 398–413.
- [26] U. Klank, D. Pangercic, R. B. Rusu, and M. Beetz, “Real-time cad model matching for mobile manipulation and grasping,” in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2009, pp. 290–296.
- [27] G. Du, K. Wang, S. Lian, and K. Zhao, “Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: A review,” *Artif. Intell. Rev.*, vol. 54, pp. 1677–1734, 2020.
- [28] R. Li et al., “Localization and manipulation of small parts using gelsight tactile sensing,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 3988–3993.
- [29] J. Bimbo, L. D. Seneviratne, K. Althoefer, and H. Liu, “Combining touch and vision for the estimation of an object’s pose during manipulation,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 4021–4026.
- [30] Y. Ding, J. Bonse, R. Andre, and U. Thomas, “In-hand grasping pose estimation using particle filters in combination with haptic rendering models,” *Int. J. Humanoid Robot.*, vol. 15, no. 01, 2018, Art. no. 1850002.
- [31] T. D. Barfoot and P. T. Furgale, “Associating uncertainty with three-dimensional poses for use in estimation problems,” *IEEE Trans. Robot.*, vol. 30, no. 3, pp. 679–693, Jun. 2014.