JAIST Repository
https://dspace.jaist.ac.jp/

Title	On the Generality and Application of Mason's Voting Theorem to Center of Mass Estimation for Pure Translational Motion
Author(s)	Gao, Ziyan; Elibol, Armagan; Chong, Nak Young
Citation	IEEE Transactions on Robotics: 1-16
Issue Date	2024-04-22
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/18863
Rights	This is the author's version of the work. Copyright (C) 2024 IEEE. IEEE Transactions on Robotics, 2024, 1-16. DOI: https://doi.org/10.1109/TRO.2024.3392080. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

On the Generality and Application of Mason's Voting Theorem to Center of Mass Estimation for Pure Translational Motion

Ziyan Gao, Armagan Elibol, and Nak Young Chong, *Senior Member, IEEE*

***Abstract*—Object rearrangement is widely demanded in many of the manipulation tasks performed by industrial and service robots. Rearranging an object through planar pushing is deemed energy efficient and safer compared with the pick-and-place operation. However, due to the unknown physical properties of the object, re-arranging an object toward the target position is difficult to accomplish. Even though robots can benefit from multi-modal sensory data for estimating novel object dynamics, the exact estimation error bound is still unknown. In this work, firstly, we demonstrate a way to obtain an error bound on the center of mass (CoM) estimation for the novel object only using a position-controlled robot arm and a vision sensor. Specifically, we extend Mason's Voting Theorem (MVT) to object CoM estimation in the absence of accurate information on friction and object shape. The probable CoM locations are monotonously narrowed down to a convex region, and the Extended Voting Theorems (EVT's) guarantee that the convex region contains the CoM ground truth in the presence of contact normal estimation error and pushing execution error. For the object translation task, existing methods generally assume that the pusher-object system's physical properties and full-state feedback are available, or utilize iterative pushing executions, which limits the application of planar pushing to real-world settings. In this work, assuming a nominal friction coefficient between the pusher and object through contact normal error bound analysis, we leverage the estimated convex region and the Zero Moment Two Edge Pushing (ZMTEP) method [1] to select the contact configurations for object pure translation. It is ensured that the selected contact configurations are capable of tolerating the CoM estimation error. The experimental results show that the object can be accurately translated to the target position with only two controlled pushes at most.**

***Index Terms*—Object Physical Property Estimation, Planar Pushing, Non-Prehensile Manipulation.**

I. INTRODUCTION

PUSHING is one of the well-known non-prehensile manipulation skills that can be deployed in different types of robotic platforms. Compared with grasping, pushing is especially useful when the object is ungraspable because of its geometrical shape, dimension, or physical properties. Therefore, pushing has been used in various tasks such as object rearrangement [2], target acquisition [3], singulation [4], bin picking [5], in-hand re-grasping [6], and similar others. In this work, we are looking at ways to translate the novel object to the target position accurately and efficiently without using a state feedback controller. Here, the meaning of "novel" has two aspects; the physical properties are unknown, and the shape information is not perfectly known. Despite its simplicity, pushing introduces challenges around object rearrangement. Our method ensures that the pushed object undergoes pure translation to the target position with a minimum number of pushes.

The primary emphasis of this work is placed on the estimation of the CoM of novel objects as it plays an important role in pushing. To translate the object toward the target, the net force composited by contact forces must pass through the CoM and parallel to the target direction. In general, object inertial parameters identification leverages force and tactile sensing or multi-modal sensing as summarized in [7]. However, little research has addressed the estimation error bound, which is important to guarantee that the selected contact configuration is robust enough to tolerate the CoM estimation error. Furthermore, if the force measurement is unavailable and the object shape is partially unknown, the CoM estimation becomes more difficult. The proposed method, only collecting visual data of the pushed object, is easily applicable to the case where the object is neither attached at the end-effector nor graspable.

Inspired by the MVT, if the object is pushed along the contact normal and the object's sense of rotation is known, the CoM lies inside a convex region specified by the half-plane defined by the ray of the pusher motion. However, the MVT only determines the instant sense of rotation given a known contact normal. If the contact normal is imperfectly known and the pusher slides along the object, the object's sense of rotation may not follow the MVT, leading to the wrong CoM estimation of objects. In this work, the MVT is extended to facilitate the estimation of the CoM of novel objects in the presence of contact normal estimation error and continuous push in a straight line. We present the contact normal estimation error bound and a CoM region update rule, ensuring that the estimated CoM region encloses the CoM ground truth given a series of pushes. Then, to deal with the sliding between the pusher and the object, we give a sufficient condition for determining the sense of rotation of polygonal objects pushed continuously in a straight line.

For selecting a nominal friction coefficient between the pusher and object given a contact normal error bound, we

This work was supported in part by JSPS KAKENHI Grant Number JP23K03756, and in part by the Asian Office of Aerospace Research and Development under Grant/Cooperative Agreement Award No. FA2386-22-1-4042.

Z. Gao and N. Y. Chong are with the School of Information Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa 923-1292, Japan {ziyan-g, nakyoung}@jaist.ac.jp

A. Elibol is with IAS-8: Data Analytics and Machine Learning, Institute for Advanced Simulation, Forschungszentrum Jülich, 52428 Jülich, Germany a.elibol@fz-juelich.de

modify the ZMTEP [1] for object pure translation. The sampled contact configurations are capable of tolerating the CoM estimation error and guaranteed to resist any rotations during translation. When pushing the object using the sampled contact configurations, sticking contact between the pusher and the object is ensured. Therefore, no object state feedback is needed, as the object pose relative to the pusher remains the same. When the target position cannot be achieved by any of the contact configurations, we utilize two contact configurations whose pushing directions can positively span the target direction. In addition, our proposed method is insensitive to the pushing speed. We show through real experiments that the object can be translated to the target position with two pushes at most no matter how far the distance is from the target to the object, with a wide range of pushing speed which accounts for both the quasi-static and dynamic pushing.

In summary, this paper introduces the following theoretical contributions:

- Generalize and apply the MVT to CoM estimation of novel objects under unknown friction parameters and imperfect contact and shape information.
- A sufficient condition for pure translation of novel objects with the CoM estimation uncertainty.
- A sufficient condition for determining the sense of rotation of novel polygonal objects.

The assumptions are made as follows:

- The object is flat, and it does not tilt or flip during and after being pushed.
- The pusher, the object, and the support plane are rigid.
- Coulomb's law of friction applies.

The rest of the paper is organized as follows. In Section II, we review the related work in object inertial parameter estimation and planar pushing. In Section III, we present our CoM estimation method. In Section IV, based on the CoM estimation method, we explain the modification of ZMTEP for object pure translation. In Section V, we clarify the experimental setting for conducting object CoM estimation and pure translation. In Section. VI, we offer a comprehensive explanation of the implementation process for both the proposed method and the baselines. In Section. VII, we describe the experiments conducted to validate the modified ZMTEP method for purely translating objects. Finally, conclusions are drawn in Section VIII.

II. Related Work

A. Object Physical Property Estimation

A comprehensive survey on the estimation of objects' inertial parameters was performed in [7]. Our proposed method belongs to "exploratory methods" characterized by the interaction between the robot and the object. Here, we further categorize the exploratory methods as tactile, vision, and multi-modal sensor-based methods by the type of sensors used.

Among tactile sensor-based methods [8], [9], [10], [11], [12], [13], for ungraspable objects, Yu *et al.* [8] defined Gravity Equi-Effect Plane which contains the center of gravity. The object CoM was determined by the intersection of multiple Gravity Equi-Effect Planes found by multiple tip operations. The assumptions made were the rigid polyhedron-shaped object and sticking line contact between the object and the ground. For planar objects, Yu *et al.* [9] used two fingertip tactile sensors to measure the reactive forces during the interaction with the box, and formulated a least-squares problem based on the equation of motion and the measurement. However, this method assumed a known object shape and identical contact normals between two fingers. For graspable objects, the dynamics of the grasped object was estimated based on the measured forces from the fingertips [10]. Specifically, the estimation of friction between the fingertip and the object was done through a precise control scheme, measuring the forces when slippage occurs to estimate the boundary of the friction cone. The estimation of object inertial parameters was formulated as a factor graph taking the measurement uncertainty into account. Wang *et al.* [11] used a vision-based tactile sensor and encoded the object's inertial parameters and friction into a latent space using the deformation at the contact during object tilting and shaking. In [12], the CoM of a quadrilaterally-faced hexahedron object was identified through the cross-correlation between the force signals of two fingertips. Zhao *et al.* [13] detected the relative rotation angle due to slippage when the object was lifted and measured the gravitational force, and estimated the CoM using static equilibrium.

For vision sensor-based methods, one earlier work [14] uniformly sampled a set of hypothesized support points inside the convex region of the object projected onto the plane, and the friction parameters were estimated using a linear programming formulation. However, this method requires a known object shape and precise object pose tracking. More recently, in [15], the mass density was estimated using the impulse response to a striking motion. However, this method requires a high-rate camera and can not apply to fragile objects. Song *et al.* [16] proposed to learn the coupled mass-friction parameters using a differentiable simulator. However, the contact normal direction can not be well estimated as the object is approximated by rigidly connected 2D small grids. Li *et al.* [17] used recurrent neural networks to estimate the object CoM given a history of pushing interactions. Kumar *et al.* [18] employed a policy network to interact with an articulated object, and a predictor network to predict the mass distribution of an articulated object. Xu *et al.* [19] proposed a learning framework to encode the object's physical properties implicitly using high-speed pushing and colliding. In our previous work [1], [20], the object CoM was estimated based on the MVT, assisted by a prediction model of the pushed object motion [21].

Among multi-modal sensor-based methods, Tanaka *et al.* [22] estimated the visual center of the object and found pushing forces that can purely translate it. The estimation of the object mass was based on the measured contact forces and a known friction coefficient. In [23], the friction coefficients of the unknown objects were inferred from the visual color feature and partial haptic measurements. Veres *et al.* [24] proposed to learn the CoM implicitly and predict the grasp affordance in an end-to-end fashion. Specifically, several grasping trials were executed to collect the support set

for generating the grasp affordance.

In this work, we extend MVT to estimate the CoM of novel objects with inaccurate object shape and unknown friction. We show that the estimation of the CoM can be carried out accurately only using a vision sensor.

B. Object Rearrangement by Pushing

The previous studies [25], [26] focused on the use of open-loop pushing to reposition objects with known geometry and inertial properties using a flat pushing fence. However, this study aims to address the challenge of manipulating objects with inaccurate geometry and unknown inertial properties. Under the assumption of quasi-static pushing, an ellipsoid model was derived in [27] to approximate the limit surface [28], [29] which relates the applied force with object velocity. The ellipsoid model was then used for pushing controller design as in [30], [31], [32]. However, the ellipsoid approximation needs a known CoM and friction between the pusher and the slider and it under-fits the real object motion. Recently, in [33], instead of using an ellipsoid model, the Gaussian Process (GP) was used to model the dynamics of the object from experimental data. The authors leveraged GP and model predictive control to repose the object. The methods proposed in [30], [31], [32], [33] require accurate full-state feedback, which is expensive to obtain in the common setting.

The object rearrangement task can also be solved by multi-step pushing. In [17], [34], end-to-end learning models were trained to predict object motion. Then, a series of pushing actions were planned to achieve the goal pose. The authors in [35] showed the exact step bound for a given object rearrangement task and a set of known line pushes with known pushing effects. However, these methods lack efficiency, as the number of pushes is proportional to the difference between the initial and target object poses. In contrast, our proposed method for object translation only requires at most two pushes.

On the other hand, without estimating the object dynamics or prior knowledge of pushing effects, Lloyd and Lepora [36] showed that the novel object can be rearranged using tactile and vision sensor feedback. This method is time-consuming and computationally intensive as the contact normal needs to be inferred online through a convolutional neural network. We show that our method can translate the object efficiently and accurately with a variable speed range. In [37], the authors optimized the contact configurations to achieve minimum effort when pushing a known uniform-density object along the required trajectory using several mobile robots. In our work, the contact configuration is chosen to eliminate the uncertainty of the CoM. In [38], [39], physics engines were leveraged to plan the pushing action for uniformly distributed objects.

NOMENCLATURE

- $\mathbf{p}$ A sampled contact position on the object boundary in the image frame.
- $\mathbf{n}$ Contact normal direction represented by a unit two-dimensional vector.
- $\tilde{\mathbf{n}}, \tilde{\mathbf{n}}_l, \tilde{\mathbf{n}}_r$ Estimated contact normal direction (into the object) and its left and right limits.
- $\mathbf{R}_{fl}, \mathbf{R}_{fr}$ Rays delimiting the left and right boundaries of the friction cone.
- $\mathbf{R}_p$ Rays specifying the pusher motion.
- $\mathbf{R}_n, \mathbf{R}_{\tilde{n}}, \mathbf{R}_{\tilde{n}_l}, \mathbf{R}_{\tilde{n}_r}$ Rays starting from $\mathbf{p}$ and along $\mathbf{n}, \tilde{\mathbf{n}}, \tilde{\mathbf{n}}_l, \tilde{\mathbf{n}}_r$.
- $\mathbf{R}_{pl}, \mathbf{R}_{pr}$ Rays specifying the left or right bound of the difference between the commanded pushing and the real pushing.
- $\mathbf{C}_{\tilde{n}}$ Estimated contact normal error cone (bound), specified by two rays $\mathbf{R}_{\tilde{n}_l}$ and $\mathbf{R}_{\tilde{n}_r}$ delimiting the left and right boundaries of the bound.
- $\mathbf{C}_p$ Pushing execution error cone (bound), specified by two rays $\mathbf{R}_{pl}$ and $\mathbf{R}_{pr}$ delimiting the left and right boundaries for pushing execution error.
- ${}^{+}\mathbf{C}_{\tilde{n}}$ Enlarged $\mathbf{C}_{\tilde{n}}$ with $\mathbf{R}_{\tilde{n}_l}$ and $\mathbf{R}_{\tilde{n}_r}$ shifted to the left and right along the object boundary to minimally wrap $\mathbf{C}_p$. The shifted $\mathbf{R}_{\tilde{n}_l}$ and $\mathbf{R}_{\tilde{n}r}$ are represented by ${}^{+}\mathbf{R}_{\tilde{n}l}$ and ${}^{+}\mathbf{R}_{\tilde{n}r}$
- $\mathbf{C}_r$ Rotation uncertainty cone.
- $\mathbf{A}_{CoM}$ Possible CoM locations represented by a convex region inside the convex hull of the object mask.
- $\mathbf{A}_{\tilde{n}l}$ Area of intersection between the half plane (on which ${}^{+}\mathbf{R}_{\tilde{n}_l}$ has positive moment) and $\mathbf{A}_{CoM}$.
- $\mathbf{A}_{\tilde{n}r}$ Area of intersection between the half plane (on which ${}^{+}\mathbf{R}_{\tilde{n}_r}$ has negative moment) and $\mathbf{A}_{CoM}$.
- $\mathbf{A}_{pl}$ Area of intersection between the half plane (on which $\mathbf{R}_{\mathbf{p}_l}$ has positive moment) and $\mathbf{A}_{CoM}$.
- $\mathbf{A}_{pr}$ Area of intersection between the half plane (on which $\mathbf{R}_{\mathbf{p}_r}$ has negative moment) and $\mathbf{A}_{CoM}$.
- $\mathbf{p}_{CoM}$ CoM ground truth.
- $\tilde{\mathbf{p}}_{CoM}$ Estimated CoM, or the centroid of $\mathbf{A}_{CoM}$.
- α Half angle of the friction cone.
- δ Half angle of the contact normal error bound $\mathbf{C}_{\tilde{n}}$.
- θ_e Half angle of pushing execution error cone $\mathbf{C}_p$.
- d_e Max contact position error between the commanded and real contact position.

III. COM ESTIMATION

The objective of this study is to estimate the CoM of an object using vision-guided planar pushing in the presence of inaccurate shape information and unknown friction. This section begins with briefly examining the possibility of applying the MVT to the estimation of the CoM of an object. Under realistic conditions of inaccurate contact normal and friction cone between the pusher and the object, we introduce the concept of rotation indeterminacy cone with the estimation error bound on the contact normal. We then present the first EVT, which specifies the conditions for eliminating the error caused by inaccurate contact normal. Additionally, we introduce the second EVT, which specifies the conditions for eliminating the error introduced by the relative sliding between the pusher and the object. Next, we provide the update rule for the CoM region estimation and the selection method for the pusher motion. Finally, we summarize the proposed method in Alg. 1.

A. Applying MVT to Object CoM Estimation

In this work, the coordinate system is right-handed with the positive z-axis pointing out of the page. Given a pusher-

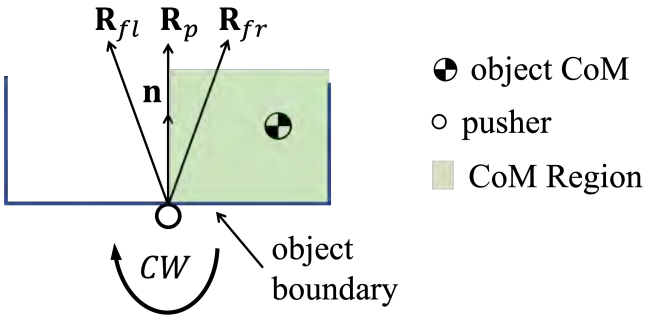

Fig. 1. Probable CoM region given $\mathbf{R}_p$, $\mathbf{R}_{fl}$, $\mathbf{R}_{fr}$, and the object sense of rotation. In this figure, $\mathbf{R}_p$ solely determines the rotation sense.

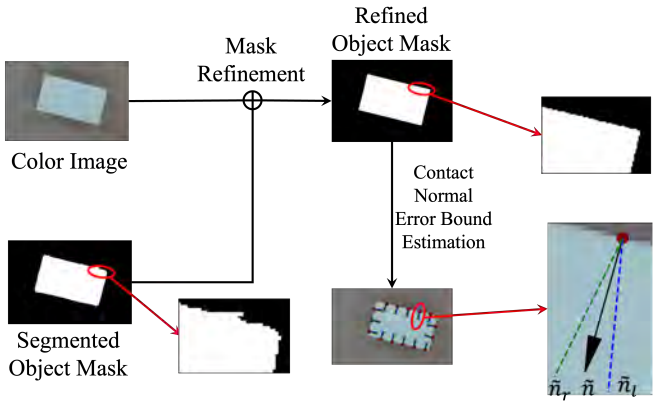

Fig. 2. The object mask is refined using RGB image and segmented object mask, and the contact normal error bound is estimated based on the refined mask.

object system, the CoM of the object is denoted as $\mathbf{p}_{CoM}$. The two rays delimiting the left and right boundaries of the friction cone are represented as $\mathbf{R}_{fl}$ and $\mathbf{R}_{fr}$, respectively. $\mathbf{R}_p$ specifies the ray of the pusher motion. MVT states that the spatial relationship among three rays $\mathbf{R}_{fl}$, $\mathbf{R}_{fr}$, $\mathbf{R}_p$, and the $\mathbf{p}_{CoM}$ vote for the slider object's sense of rotation. For instance, the object rotates clockwise if at least two rays have negative moments about $\mathbf{p}_{CoM}$ as illustrated in Fig. 1.

Projecting the object to the support plane, the $\mathbf{p}_{CoM}$ lies inside the convex hull of the object boundary. If $\mathbf{R}_p$, $\mathbf{R}_{fl}$, and $\mathbf{R}_{fr}$ are known and the object's instant sense of rotation is observed, according to the MVT, $\mathbf{p}_{CoM}$ is located in a convex region enclosed by the convex hull of the object and the middle of the three rays $\mathbf{R}_p$, $\mathbf{R}_{fl}$, and $\mathbf{R}_{fr}$. We refer to this convex region as the CoM region denoted by $\mathbf{A}_{CoM}$. By imparting a series of pusher motions to the object such that the middle ray of each of the motions passes $\mathbf{A}_{CoM}$, the region can be monotonously narrowed down. The smaller the region it gets, the more approximate its centroid is to the CoM.

If $\mathbf{R}_p$ is inside the friction cone as an *a priori* knowledge, it can determine the CoM region using the sense of rotation of the object, regardless of the directions of two other rays or friction cone limits. Under the condition that the friction between the pusher and the object is unknown but accurate contact normal is available, the $\mathbf{R}_p$ can be kept aligned with the contact normal. Since the contact normal evenly splits the friction cone, a minimum of two rays including $\mathbf{R}_p$ always vote for the same rotation sense. Therefore, $\mathbf{R}_p$ becomes the right or left border of the CoM region, depending on the rotation sense, as shown in Fig. 1. Furthermore, if the object is pushed along the contact normal passing through the centroid of the CoM region and rotated, one-half of the CoM region is reduced. If the object is purely translated, the CoM region degenerates to a line segment specified by the intersection between the CoM region and $\mathbf{R}_p$. If we use all pixels inside the CoM region to approximate the CoM region, ideally, the CoM ground truth can be bounded within one pixel by pushing the object $log_2(n)$ times, where n is the number of pixels occupying the CoM region.

In practice, both the accurate contact normal and friction cone at the contact point are unknown. To apply the MVT to narrow down the CoM region, at least the contact normal has to be identified. However, contact normal estimation is inaccurate due to sensor resolution or noise. $\mathbf{R}_p$ outside the friction cone might cause the CoM region update process to fail, and the real CoM would fall outside the CoM region. In addition, if real-time pose tracking of the pushed object is unavailable, sliding between the pusher and the object may lead to successive changes in the object's sense of rotation, *e.g.*, the object rotates counterclockwise and then clockwise so that the rotation sense determined by the MVT may be different from the real observation. All of these imperfections introduce challenges around the MVT. In order to estimate the CoM accurately in the presence of such challenges, we present extensions on the MVT by modeling and taking into account the contact normal estimation error bounds.

B. Contact Normal Estimation Error Bounds

To estimate the contact normal direction, we employ a vision sensor to capture images of the object to find the normal to the lateral surface using a set of pixels on the object boundary. To better extract the object boundaries and reduce the estimation error introduced by the sensor noise, we employ a high-resolution mask segmentation model called CascadePSP [40], and obtain the refined object mask using the raw RGB image and segmented object mask as shown in Fig. 2. Then, the surface normal can be estimated by analyzing the covariance matrix of a local image patch around a contact point. The eigenvector associated with the smallest eigenvalue of the covariance matrix is regarded as the surface normal [41]. The estimated contact normal, however, is inaccurate due to the sensor noise and errors introduced during conversion from the real world to the discretized image frame. To deal with this problem, we introduce an error bound on the contact normal estimation specified by the left and right limits $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$, respectively.

We formulate the contact normal error bound estimation problem as an optimization problem which will hereinafter be described. The superscript i denotes the index of a discretized contact point on the boundary of an object. We estimate the normal direction $\tilde{\mathbf{n}}^i$ at the contact point $\mathbf{p}^i$ using a set of k adjacent contact points $\{\mathbf{p}^{(i+j)}\}, j \in (-\frac{k-1}{2}, \frac{k-1}{2})$ on

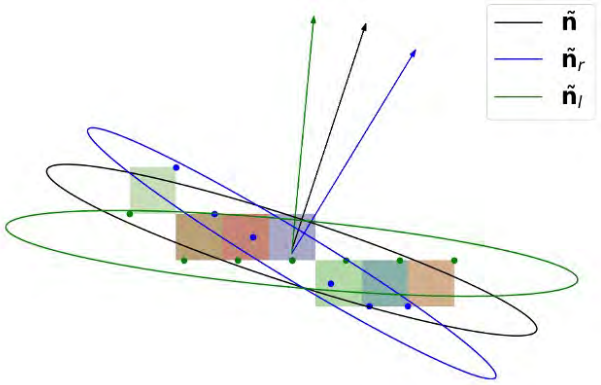

Fig. 3. Finding contact normal error bound by solving the optimization problem in Eq. 1, where $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$ are the left and right boundaries of $\mathbf{C}_{\tilde{n}}$. Local contour pixels are represented by cuboids. The circular dots represent the optimization results indicating the local curvatures. Ellipses are drawn using the dots, and the minor axes of the ellipses indicate the estimated contact normal error bounds.

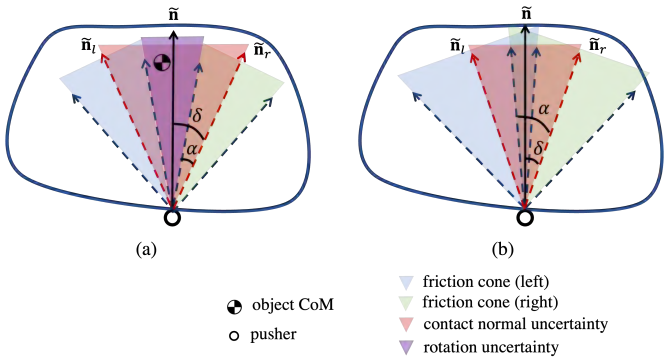

Fig. 4. Illustration of rotation indeterminacy cone $\mathbf{C}_r$. If object CoM lies inside $\mathbf{C}_r$ as shown in (a) and the object is pushed along $\tilde{\mathbf{n}}$, the rotation sense may be opposite to the case pushed along the contact normal ground truth. In (b), the presence of a larger friction cone eliminates rotation uncertainty.

a two-dimensional object boundary, where j is an integer and k is an odd number of neighborhood window size. To find the left limit of the contact normal error bound $\tilde{\mathbf{n}}_l$, we create k two-dimensional variables represented as $\{\mathbf{p}_v^{(i+j)}\}, j \in (-\frac{k-1}{2}, \frac{k-1}{2})$. The $\tilde{\mathbf{n}}_l$ can be found by minimizing the magnitude of the cross product of $\tilde{\mathbf{n}}_l^i$ and $\tilde{\mathbf{n}}^i$ given by Eq. 1.

$$\begin{aligned} \min_{\tilde{\mathbf{n}}_l^i} \quad & \tilde{\mathbf{n}}_l^i \times \tilde{\mathbf{n}}^i \\ \text{s.t.} \quad & \mathbf{Q}_v = \frac{1}{k}\Sigma_{j=i-\frac{k-1}{2}}^{i+\frac{k-1}{2}}(\mathbf{p}_v^j - \bar{\mathbf{p}}_v)(\mathbf{p}_v^j - \bar{\mathbf{p}}_v)^\top \\ & \mathbf{Q}_v \tilde{\mathbf{n}}_l^i = \lambda \tilde{\mathbf{n}}_l^i \\ & \mathbf{p}^j - \frac{1}{2} \le \mathbf{p}_v^j \le \mathbf{p}^j + \frac{1}{2} \\ & \mathbf{p}_v^j \le (\ge) \frac{1}{2}(\mathbf{p}_v^{j+1} + \mathbf{p}_v^{j-1}) \end{aligned} \tag{1}$$

$\mathbf{Q}_v$ is the covariance matrix calculated based on $\mathbf{p}_v^i$, and $\tilde{\mathbf{n}}_l^i$ is the eigenvector of the $\mathbf{Q}_v$. $\bar{\mathbf{p}}_v$ is the mean of k two-dimensional variables. $\mathbf{p}_v^i$ is bounded by the sub-pixel accuracy of $\mathbf{p}^i$. In addition, the local curvature is assumed either convex or concave as stated in the last constraint via $\le (\ge)$. In practice, we exclude the estimation of neighboring pixels that do not exhibit convex or concave behavior. Similarly, the right bound is found by minimizing the cross product of $\tilde{\mathbf{n}}$ and $\tilde{\mathbf{n}}_r$. Fig. 3 shows an example of finding $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$.

After obtaining $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$, the estimated contact normal $\tilde{\mathbf{n}}$ is recalculated to be aligned with the halfway between $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$. *i.e.*, $\tilde{\mathbf{n}} = \frac{\tilde{\mathbf{n}}_l + \tilde{\mathbf{n}}_r}{2}$. Provided that the contact normal estimation uncertainty can be calculated, in the following, we present how to utilize this uncertainty within our context of CoM estimation.

C. Instant Sense of Rotation Indeterminacy Analysis with Contact Normal Error Bounds

Using $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$ at a contact point, the contact normal error cone $\mathbf{C}_{\tilde{n}}$ can be constructed. We assume that the contact normal ground truth can be positively spanned by $\mathbf{C}_{\tilde{n}}$. If the robot pushes an object along the estimated contact normal, which lies outside the actual friction cone, and updates the CoM region based on the sense of rotation of the object, the CoM region and the CoM ground truth are mutually exclusive. In such a case, the ray of pushing direction does not cast a vote on the object's sense of rotation. Therefore, to ensure that the estimated CoM region encompasses the CoM ground truth, the contact normal estimation error bound must be considered.

In this study, we present the concept of the rotation indeterminacy cone $\mathbf{C}_r$ as a means to address the issue of exclusivity. This cone refers to the uncertainty bounds on the rotation sense, meaning that an update on the CoM region could go wrong if the CoM ground truth is inside this cone and the object is pushed along the estimated contact normal. If the $\mathbf{C}_r$ is accurately identified, the exclusivity case can be resolved by incorporating the region that is shared by $\mathbf{C}_r$ and the current CoM region, regardless of the sense of rotation of the object in updating the CoM region.

Assuming that the friction coefficient between the pusher and the object is known, the $\mathbf{C}_r$ can be determined via the following steps:

- Draw contact normal error bounds specified by $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$ at the contact point.
- Construct friction cones by regarding $\tilde{\mathbf{n}}_l$ and $\tilde{\mathbf{n}}_r$ as the contact normal direction.
- The left boundary of $\mathbf{C}_r$ is the right boundary of the friction cone which regards $\tilde{\mathbf{n}}_l$ as the contact normal.
- The right boundary of $\mathbf{C}_r$ is the left boundary of the friction cone which regards $\tilde{\mathbf{n}}_r$ as the contact normal.

$\mathbf{C}_r$ is illustrated in Fig. 4 (a). To enhance the expression of the idea, let us assume that the true contact normal is aligned with $\tilde{\mathbf{n}}_l$. In this case, $\tilde{\mathbf{n}}$ is outside the friction cone. If the CoM ground truth lies between the right boundary of the friction cone and $\tilde{\mathbf{n}}$, the update on the CoM region goes wrong. This is because $\tilde{\mathbf{n}}$ votes for counterclockwise rotation, however, the object will rotate clockwise according to the MVT.

The definition of the half angle of $\mathbf{C}_r$ is given by $\delta - \alpha$, where δ specifies the half-angle of the contact normal error bound, and α represents the friction angle. $\delta - \alpha$ is meaningful only if it is greater than zero. The instant object sense of rotation error only occurs when the CoM ground truth is inside

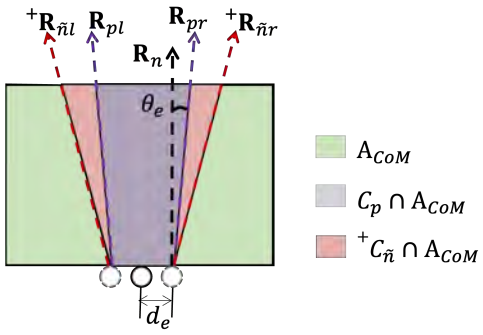

Fig. 5. Enlarged contact normal error bound $^{+}\mathbf{C}_{\tilde{\mathbf{n}}}$ which is specified by $^{+}\mathbf{R}_{\tilde{\mathbf{n}}l}$ and $^{+}\mathbf{R}_{\tilde{\mathbf{n}}r}$, and pushing execution error bound specified by $\mathbf{R}_{pl}$ and $\mathbf{R}_{pr}$.

the $\mathbf{C}_r$. To eliminate this error, the rotation indeterminacy region represented by $\mathbf{C}_r \cap \mathbf{A}_{CoM}$ must lie in the updated CoM region, no matter the object's sense of rotation. This guarantees that the inclusiveness between the CoM region and the CoM ground truth.

However, if the friction angle α is greater than the half-angle of contact normal error bound δ, the rotation indeterminacy cone becomes equal to $\emptyset$ as illustrated in Fig. 4 (b). In this case, the contact normal estimation error does not introduce any error in the rotation sense. Therefore, the estimation error due to inaccurate contact normal estimation can also be eliminated when the friction cone is larger than the estimated contact normal error bound.

Based on the aforementioned analysis, we introduce the first EVT stated in Theorem 1.

***Extended Voting Theorem (EVT)** 1:* If the friction angle is greater than the half-angle of contact normal error bound, and the object is pushed along the estimated contact normal, then the instant sense of rotation of the pushed object is solely determined by $\mathbf{R}_p$.

Proof: If the friction angle is greater than the half-angle of the contact normal error bound, the contact normal error bound is fully encompassed by the friction cone. Thus, any ray of pushing inside the contact normal error bound is also inside the friction cone. As a result, the estimated contact normal which evenly splits the contact normal error bound also is inside the friction cone. Based on the MVT, the ray of pushing $\mathbf{R}_p$ is the middle ray of $(\mathbf{R}_{fl}, \mathbf{R}_p, \mathbf{R}_{fr})$ so that $\mathbf{R}_p$ solely determines the object sense of rotation.

In practice, controlling the pusher motion precisely is difficult due to hand-eye calibration and object pose tracking errors. The execution error in pushing direction and contact point are specified by θ_e and d_e, where θ_e defines the maximal difference between the commanded and actual pushing directions, and d_e defines the maximal distance between the commanded and actual contact points. Now, the pushing execution error cone, denoted as $\mathbf{C}_p$, is specified by $\mathbf{R}_{pl}$ and $\mathbf{R}_{pr}$, which depend on θ_e, d_e as shown in Fig. 5. Any actual push action represented by $\mathbf{R}_p$ can be bounded by $\mathbf{C}_p$. Here the half angle of $\mathbf{C}_p$ is smaller than that of $\mathbf{C}_{\tilde{n}}$, which is considered reasonable if the camera is well calibrated. To avoid the estimation error introduced by the pushing execution error, the intersection region between the current CoM region and pushing execution error cone ($\mathbf{A}_{CoM} \cap \mathbf{C}_p$) must lie in the updated CoM region.

To incorporate both the contact normal estimation error and pushing execution error in the process of object CoM estimation, we enlarge the contact normal estimation error cone $\mathbf{C}_{\tilde{n}}$ by shifting $\mathbf{R}_{\tilde{n}l}$ and $\mathbf{R}_{\tilde{n}r}$ to minimally enclose the region $\mathbf{A}_{CoM} \cap \mathbf{C}_p$ as shown in Fig. 5, denoted as $^{+}\mathbf{C}_{\tilde{n}}$.

D. Sliding between Pusher and Object

The preceding sections examined the possibility of using the MVT to estimate the CoM of an object with the instant sense of rotation when the contact normal is inaccurate. However, in real-world conditions, the instant sense of rotation of an object may not always be available due to hardware limitations or lack of textures on the object for pose estimation. To address this issue, we approximate the instant sense of rotation by performing a short straight-line push. Based on the difference in object orientation before and after a push, the instant sense of rotation can be approximated. Nevertheless, the approximated sense of rotation of the object may be inaccurate because of the pusher sliding motion along the object boundary. The object could initially rotate clockwise and then later counterclockwise due to the alteration of the contact point. In this study, we present a condition that is adequate for determining the object's sense of rotation using the initial spatial relationship between $\mathbf{R}_p$ and $\mathbf{p}_{CoM}$ when the pusher slides along the object.

***Extended Voting Theorem (EVT)** 2:* If the $\mathbf{R}_p$ is initially inside the friction cone of a polygonal object and sliding occurs continuously between the pusher and the object, the object's sense of rotation is uniquely determined by the initial spatial relationship between $\mathbf{R}_p$ and CoM location.

In other words, given a polygonal convex object, if the ray of the pusher motion $\mathbf{R}_p$ lies inside the friction cone and the contact point stays on the same edge, then the object's sense of rotation remains unchanged regardless the sliding motion between the pusher and the object.

Proof: We prove EVT 2 by contradiction. Let us assume that $\mathbf{R}_p$ is within the friction cone and indicates a clockwise sense of rotation for the object. The hypothesis is that $\mathbf{R}_p$ eventually changes its indication due to the sliding motion. At first, the object rotates clockwise based on the MVT, while $\mathbf{R}_p$ rotates counterclockwise with respect to the object frame. Since the object has a polygonal shape, the directions of $\mathbf{R}_{fl}$ and $\mathbf{R}_{fr}$ with respect to the object frame will not change regardless of the direction of relative sliding motion between the pusher and the object. Consequently, $\mathbf{R}_p$ will move toward $\mathbf{R}_{fl}$. If the hypothesis holds, the rotation of $\mathbf{R}_p$ must rotate in a clockwise direction *w.r.t* the object frame, leading it to move closer to $\mathbf{R}_{fr}$ and eventually surpass the boundaries of the friction cone. However, this outcome is not feasible. Hence, this leads to a contradiction.

The proof of EVT 2 can also be done by enumerating the spatial relationships among $\mathbf{R}_p$, $\mathbf{R}_{fl}$, and $\mathbf{R}_{fr}$ during the sliding between the pusher and the object as illustrated in Fig. 6.

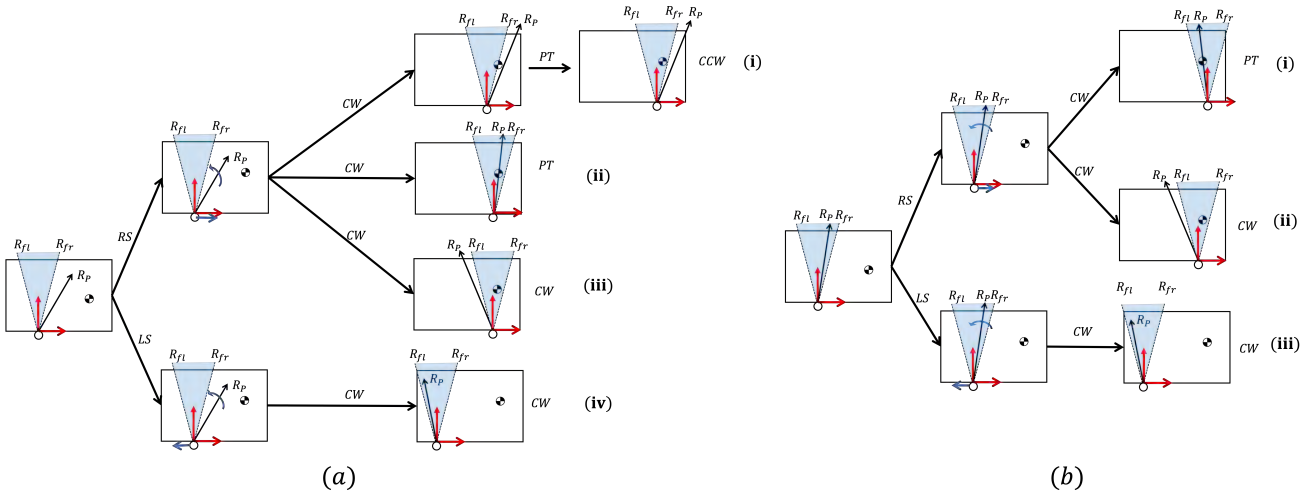

Fig. 6. Proof of EVT 2. LS or RS means that the pusher slides to the left or right *w.r.t.* the object fixed frame depicted in red color. (a) shows the case that the pushing direction $\mathbf{R}_p$ is initially outside the friction cone, while (b) shows the case that $\mathbf{R}_p$ is initially inside the friction cone.

We attach an instant frame to the object whose origin is at the contact point. One axis of the frame is along the contact normal and the other axis is along the object boundary. Based on the MVT, if the middle ray among three rays passes through the object CoM, the object translates along this ray. If the middle ray is $\mathbf{R}_p$, the object is purely translated and no sliding occurs between the pusher and the object. If the ray in the middle is the right boundary of the friction cone as illustrated in Fig. 6(a-i), the object is translated along $\mathbf{R}_{fr}$, meanwhile the pusher slides to the right with respect to the object-attached frame. In the case of continuous straight-line pushing, the object will then rotate counterclockwise. Therefore, successive changes in the object's sense of rotation might happen when the ray in the middle is the $\mathbf{R}_{fl}$ or $\mathbf{R}_{fr}$ as shown in Fig. 6(a-i).

Fig. 6 (a) illustrates the case that the $\mathbf{R}_p$ lies initially outside the friction cone. A unique sense of rotation is not guaranteed due to the unknown sliding direction between the pusher and the object. Fig. 6(b) shows the case where the $\mathbf{R}_p$ is initially inside the friction cone. In this case, regardless of the sliding direction, the successive changes in the object's sense of rotation will not occur. Therefore, if EVT 2 is satisfied, regardless of the pushing magnitude, the object's sense of rotation is initially determined and will not change.

E. CoM Region Update Rules

Here, we aim to answer the following question: Given a pusher motion and the resultant sense of rotation of an object, how can the object CoM region be maximally reduced in the presence of contact normal estimation error and pushing execution error?

If the friction cone is not guaranteed to be larger than the contact normal error bounds, based on the aforementioned instant object sense of rotation analysis, the intersection region between the rotation indeterminacy cone $\mathbf{C}_r$ and the CoM region must lie in the updated CoM region. However, the rotation indeterminacy cone cannot be found precisely if the friction is unknown. But the rotation indeterminacy cone is upper bounded by the contact normal error bound. Therefore, instead of finding an exact $\mathbf{C}_r$, the CoM region is updated in such a way that it contains the region formed by the contact normal error bound $\mathbf{C}_{\tilde{n}}$ and the current CoM region to ensure the inclusion of the CoM ground truth.

In the presence of both the pushing execution error and contact normal estimation error, $\mathbf{C}_{\tilde{n}}$ is replaced by ${}^{+}\mathbf{C}_{\tilde{n}}$ to avoid the potential estimation error caused by the pushing execution error. Therefore, if the object rotates counterclockwise, the region of $({}^{+}\mathbf{C}_{\tilde{n}} \cap \mathbf{A}_{CoM}) \cup \mathbf{A}_{\tilde{n}l}$ will be the updated CoM region. Similarly, if the object rotates clockwise, the region of $({}^{+}\mathbf{C}_{\tilde{n}} \cap \mathbf{A}_{CoM}) \cup \mathbf{A}_{\tilde{n}r}$ will form the updated CoM region. If the object is purely translated, meaning that $\mathbf{R}_p$ passes through the CoM, the region of $\mathbf{C}_p \cap \mathbf{A}_{CoM}$ is considered the updated CoM region.

On the other hand, if the friction cone is larger than the contact normal error bound, following EVT 1, the contact normal estimation error is eliminated. In this case, only the estimation error introduced by the pushing execution error is taken into account. Therefore, if the object rotates counterclockwise, the region $\mathbf{A}_{pl} \cup \mathbf{A}_{CoM}$ is selected as the updated CoM region. Similarly, the region $\mathbf{A}_{pr} \cup \mathbf{A}_{CoM}$ becomes the updated CoM region if the object rotates clockwise. The CoM region is $\mathbf{C}_p \cup \mathbf{A}_{CoM}$ if the object is purely translated. In summary, the updating rule is illustrated in Fig. 7.

F. Pusher Motion Selection

We now select the pusher motion, maximizing the reduction of the area of the CoM region using the CoM region update rule. Specifically, the scoring functions (s_n, s_p, s_d) are designed to quantify the area reduction. We aim to minimize the CoM region and avoid elongated shapes. In the case that the friction cone is not guaranteed to be larger than the contact normal estimation error cone, Eq. 2 is designed.

$$s_n(i) = \frac{2\mathbf{A}^i_{\tilde{n}l}\mathbf{A}^i_{\tilde{n}r}}{(\mathbf{A}^i_{\tilde{n}l})^2 + (\mathbf{A}^i_{\tilde{n}r})^2 + (\mathbf{A}_{CoM} \cap {}^{+}\mathbf{C}^i_{\tilde{n}})^2} \tag{2}$$

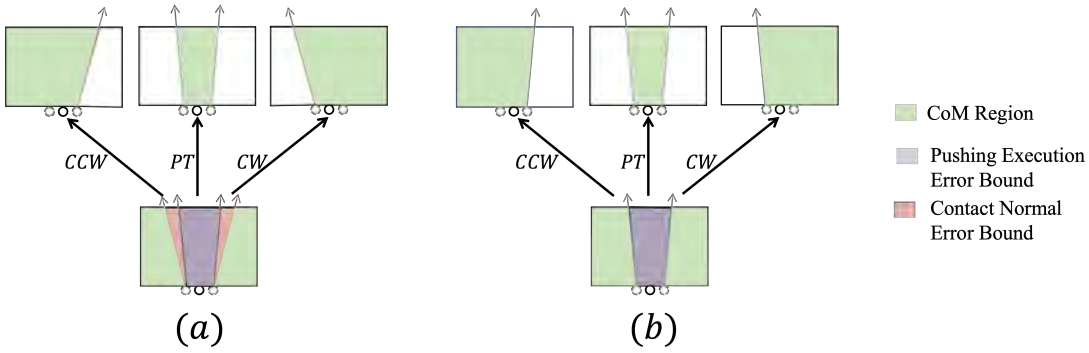

Fig. 7. CoM region updating rule: (a) when the friction cone is not guaranteed to be larger than the contact normal error bound, (b) when the contact normal estimation error is eliminated and only the pushing execution error is considered.

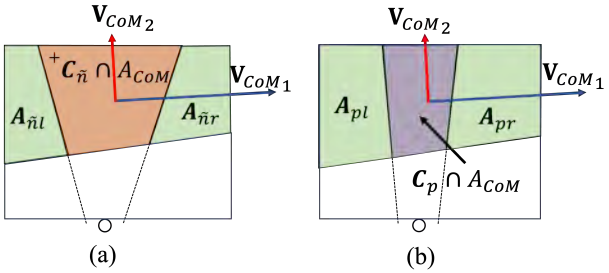

Fig. 8. Areas used for pusher motion selection. $\mathbf{V}_{CoM1}$ and $\mathbf{V}_{CoM2}$ are the principal axes calculated based on the current CoM region $\mathbf{A}_{CoM}$. The areas are used to select a pusher motion (a) when both contact normal estimation error and pushing execution error are present, and (b) when the friction cone at the contact is larger than the estimated contact normal error bound.

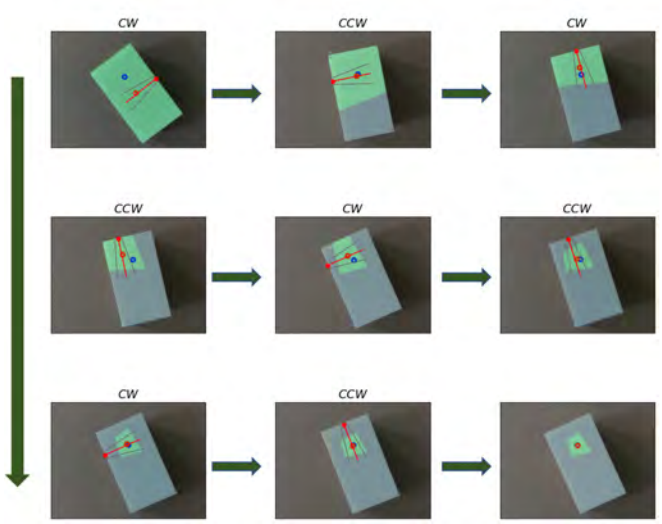

Fig. 9. CoM estimation process. In each figure, the green area is the CoM region, the red arrow shows the pushing direction, and dashed lines are the left and right limits of $^{+}\mathbf{C}_{\tilde{n}}$. The CoM ground truth is depicted as a green dot, and the estimated CoM is depicted as a red dot.

where the subscript n infers the presence of contact normal error bound. i represents the index of the sample within the set of sampled pusher motions. Specifically, the boundaries $^{+}\mathbf{R}_{\tilde{n}l}$ and $^{+}\mathbf{R}_{\tilde{n}r}$ of enlarged contact normal error bound $^{+}\mathbf{C}_{\tilde{n}}$ are used to split the current CoM region into at most three regions: $\mathbf{A}_{\tilde{n}l}$, $\mathbf{A}_{\tilde{n}r}$ and $^{+}\mathbf{C}_{\tilde{n}} \cap \mathbf{A}_{CoM}$, as shown in Fig. 8. It aims to maximize the product of the size of $\mathbf{A}_{\tilde{n}l}$ and $\mathbf{A}_{\tilde{n}r}$ and minimize $\mathbf{A}_{CoM} \cap {}^{+}\mathbf{C}^i_{\tilde{n}}$, since $\mathbf{A}_{CoM} \cap {}^{+}\mathbf{C}^i_{\tilde{n}}$ is always included in the updated CoM region regardless of the object sense of rotation. Consequently, the pusher motion of the highest score is determined by $\mathbf{A}^i_{\tilde{n}l}$ and $\mathbf{A}^i_{\tilde{n}r}$ of the same size and the minimum intersection between $\mathbf{A}_{CoM}$ and $^{+}\mathbf{C}^i_{\tilde{n}}$.

In the case that the friction cone is larger than the estimated contact normal error cone, Eq. 3 is used where the subscript p infers the presence of pushing execution error bound, as the contact normal estimation error does not cause a CoM estimation error.

$$s_p(i) = \frac{2\mathbf{A}^i_{pl}\mathbf{A}^i_{pr}}{(\mathbf{A}^i_{pl})^2 + (\mathbf{A}^i_{pr})^2 + (\mathbf{A}_{CoM} \cap \mathbf{C}^i_p)^2} \tag{3}$$

Apart from the area of the CoM region, the shape of the CoM region, which depicts the uncertainty in estimation, should also be optimized to reduce the estimation error. In the case where the CoM region is elongated, there is a possibility of significant estimation errors even if the ground truth CoM is guaranteed to be within the CoM region. This is particularly true when the ground truth CoM is far away from the centroid of the CoM region. By controlling the shape of the CoM region, it is possible to decrease the estimation error in the worst-case scenario. Therefore, Eq. 4 evaluates the geometric isotropy of the CoM region based on the direction of pushing motion indicated by the subscript d. Specifically, $\mathbf{V}_{CoM_2}$ is the second principal eigenvector of the current CoM region. This cost function evaluates how well the selected pushing action reduces the uncertainty along the first principal direction of $\mathbf{A}_{CoM}$. We therefore combine Eq. 2 or Eq. 3 with Eq. 4 in a weighted-sum fashion.

$$s_d(\tilde{\mathbf{n}}^i, \mathbf{V}_{CoM_2}) = \|\tilde{\mathbf{n}}^{i\top}\mathbf{V}_{CoM_2}\| \tag{4}$$

G. Pipeline of CoM Estimation

Based on the EVT 1 and EVT 2, we propose the CoM region update rule and pusher motion selection strategy. By repeatedly applying the pusher motion and CoM region update rule, the CoM region can be narrowed down until no further push can reduce the CoM region or a maximal number of

pushes is achieved. The pipeline of the CoM estimation is illustrated in Alg. 1, also supported by an example in Fig. 9.

Algorithm 1: CoM Region Narrowing Down Process

Input: $agent, \{\mathbf{p}\}^i, \{\tilde{\mathbf{n}}\}^i, \{\tilde{\mathbf{n}}_l\}^i, \{\tilde{\mathbf{n}}_r\}^i, \theta_e, \theta_{PT}, d_e, \mathbf{w}, T$

/* θ_e, d_e are used to construct $\mathbf{C}_p$ and calculate $\mathbf{A}_{pl}$ and $\mathbf{A}_{pr}$. T is the maximal time steps for interaction. $\mathbf{w}$ is a two-dimensional weight vector. θ_{PT} is an angle threshold determining if the object is purely translated. */

Output: $\mathbf{A}_{CoM}$

Initialize $\mathbf{A}_{CoM}$ /* The region inside the convex hull of the object. */

for $t = 1$ *to* T **do**

  /* Pusher motion selection */

  /* opencv library */

  $\mathbf{V}_{CoM_2} \leftarrow PCA(\mathbf{A}_{CoM})$

  if *include uncertainty* **then**

    Construct $\{^{+}\mathbf{C}_{\tilde{n}}\}^i, \{\mathbf{C}_p\}^i, \{\mathbf{A}_{\tilde{n}l}\}^i, \{\mathbf{A}_{\tilde{n}r}\}^i$ based on $\{\mathbf{p}\}^i, \{\tilde{\mathbf{n}}\}^i, \{\tilde{\mathbf{n}}_l\}^i, \{\tilde{\mathbf{n}}_r\}^i, \theta_e, d_e$

    i=argmin $\mathbf{w}^\top \begin{bmatrix} s_n(i) \\ s_d(\tilde{\mathbf{n}}^i, \mathbf{V}_{CoM_2}) \end{bmatrix}$

    /* Pusher-object contact interaction */

    $\Delta\theta \leftarrow agent.execute(\mathbf{p}^i, \tilde{\mathbf{n}}^i)$

    /* CoM region update */

    if $\|\Delta\theta\| > \theta_{PT}$ **then**

      if $\Delta\theta > 0$ **then**

        $\mathbf{A}_{CoM} \leftarrow (^{+}\mathbf{C}^i_{\tilde{n}} \cap \mathbf{A}_{CoM}) \cup \mathbf{A}^i_{\tilde{n}r}$

      else

        $\mathbf{A}_{CoM} \leftarrow (^{+}\mathbf{C}^i_{\tilde{n}} \cap \mathbf{A}_{CoM}) \cup \mathbf{A}^i_{\tilde{n}l}$

    else

      $\mathbf{A}_{CoM} \leftarrow (\mathbf{C}^i_p \cap \mathbf{A}_{CoM})$

  else

    Construct $\{\mathbf{C}_p\}^i, \{\mathbf{A}_{pl}\}^i, \{\mathbf{A}_{pr}\}^i$ based on $\{\mathbf{p}\}^i, \{\tilde{\mathbf{n}}\}^i, \theta_e, d_e$

    i=argmin $\mathbf{w}^\top \begin{bmatrix} s_p(i) \\ s_d(\tilde{\mathbf{n}}^i, \mathbf{V}_{CoM_2}) \end{bmatrix}$

    $\Delta\theta \leftarrow agent.execute(\mathbf{p}^i, \tilde{\mathbf{n}}^i)$

    if $\|\Delta\theta\| > \theta_{PT}$ **then**

      if $\Delta\theta > 0$ **then**

        $\mathbf{A}_{CoM} \leftarrow (\mathbf{C}^i_p \cap \mathbf{A}_{CoM}) \cup \mathbf{A}^i_{pr}$

      else

        $\mathbf{A}_{CoM} \leftarrow (\mathbf{C}^i_p \cap \mathbf{A}_{CoM}) \cup \mathbf{A}^i_{pl}$

    else

      $\mathbf{A}_{CoM} \leftarrow (\mathbf{C}^i_p \cap \mathbf{A}_{CoM})$

IV. Object Translation based on CoM Region

Our recent work [1] has shown that the ZMTEP can purely translate novel objects, if two contact normal directions positively span the pushing direction, and the lines specified by two contact normals and the pushing direction that passes through the CoM ground truth intersect at a single point. Specifically, the pusher needs to touch the object at two different edges, and the ZMTEP selects a two-edge-contact configuration with the highest tolerance against the touch position and CoM estimation errors. The larger the distance between two contact points, the larger the tolerance on CoM estimation error.

If the exact contact normal and friction cone are known at the contact point, for a specific two-edge-contact configuration, two contact forces are bounded by $(\mathbf{R}^i_{fl}, \mathbf{R}^i_{fr})$ and $(\mathbf{R}^j_{fl}, \mathbf{R}^j_{fr})$ at each contact point $\mathbf{p}^i$ and $\mathbf{p}^j$, respectively. Given a target pushing direction and two-edge-contact configuration, the CoM tolerance range can be found as follows:

- draw line that is parallel to the target pushing direction and is passing through the intersection of two left boundaries of friction cones at two contact positions represented by $\mathbf{R}^i_{fl} \cap \mathbf{R}^j_{fl}$.
- draw the other line that is parallel to the target pushing direction and is passing through $\mathbf{R}^i_{fr} \cap \mathbf{R}^j_{fr}$
- find the region encompassed by the intersection of two lines as well as the convex hull of the object.

It should be noted that the CoM tolerance range is only valid when the target pushing directions can be positively spanned by both $(\mathbf{R}^i_{fl}, \mathbf{R}^j_{fl})$ and $(\mathbf{R}^i_{fr}, \mathbf{R}^j_{fr})$. Using the CoM region, without the CoM ground truth, the following conditions should be satisfied for pure translation of an object.

- The CoM tolerance range of the two-edge-contact configuration completely encompasses the CoM region.
- The target pushing direction can be positively spanned by contact forces at two contact positions.

These two conditions are sufficient conditions for object pure translation with the estimated CoM region.

In this study, we modify the ZMTEP method to accommodate the situations where the contact normal, friction cone, and object CoM are coarsely known. Given the target direction, estimated object CoM region, and estimated contact normal error bounds, our goal is to efficiently find the two-edge-contact configurations that ensure pure object translation toward the target position.

Theoretically, if the friction angle is two times greater than the half-angle of the contact normal error bound given by Eq. 5, it is ensured that the friction cone includes the estimated contact normal error bound. In other words, any contact force inside the contact normal error bound can be positively spanned by the left and right boundaries of the friction cone. We refer to Eq. 5 as the friction condition for pure translation when the object geometry is not perfectly known. To make Eq. 5 fulfilled, we wrap the pusher with high-frictional materials. After that, the approximated CoM tolerance range can be found by utilizing the left and right limits of contact normal error bounds $\mathbf{R}_{\tilde{n}l}$ and $\mathbf{R}_{\tilde{n}r}$.

$$\alpha \geq 2\delta \tag{5}$$

The computational complexity for finding two-edge-contact configurations is $\mathbf{O}(n^2)$, where n represents the number of distinct points on the object contour, as we iterate through all

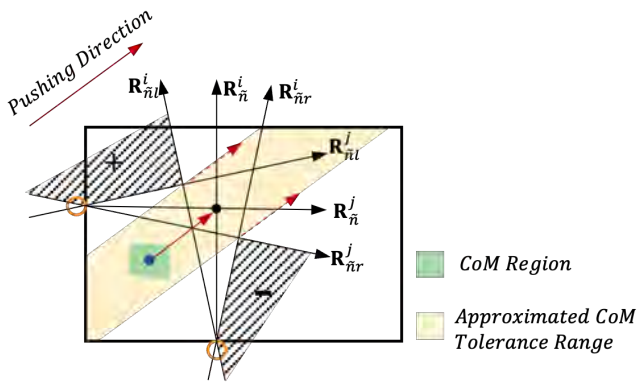

Fig. 10. The approximated CoM tolerance range is depicted by the light yellow colored region. The pushing direction is specified based on the centroid of the CoM region and the intersection point between $\mathbf{R}^i_{\tilde{n}}$ and $\mathbf{R}^j_{\tilde{n}}$. The plus and minus signs in the hatched regions are the moment labels of the two-edge contact configuration.

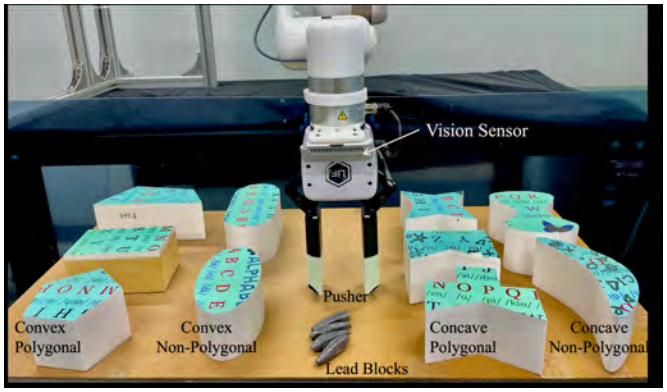

Fig. 11. Experimental setup for CoM estimation and pure translation of objects with different shapes.

possible combinations of two contact points. To efficiently determine the configuration for any target directions, we sample uniformly a set of two-edge-contact configurations. We then delineate the pushing direction for each sampled configuration using the centroid of CoM region and the intersection point $\mathbf{R}^i_{\tilde{n}} \cap \mathbf{R}^j_{\tilde{n}}$ as shown in Fig. 10. We finally check if the sampled configuration and its corresponding pushing direction satisfy the sufficient conditions for pure translation or not. The steps involved in the above pipeline are described below:

- Pair each contour point with another contour point on different edges, represented by a set of $\{\mathbf{p}^i, \mathbf{p}^j, \tilde{\mathbf{n}}^i, \tilde{\mathbf{n}}^j, \mathbf{C}^i_{\tilde{n}}, \mathbf{C}^j_{\tilde{n}}, \mathbf{n}_d\}^m$ with m examples. Here, $\mathbf{n}_d$ denotes the target pushing direction specified by $\tilde{\mathbf{p}}_{CoM}$ and $\mathbf{R}^i_{\tilde{n}} \cap \mathbf{R}^j_{\tilde{n}}$.
- Remove examples whose $(\mathbf{R}^i_{\tilde{n}l}, \mathbf{R}^j_{\tilde{n}l})$ and $(\mathbf{R}^i_{\tilde{n}r}, \mathbf{R}^j_{\tilde{n}r})$ can not positively span $\mathbf{n}_d$.
- Remove examples whose tolerance range cannot fully cover the CoM region.
- Remove examples whose distance between two contact points is larger than the maximum stroke of the gripper.

What remains of the initial m examples are all two-edge-contact configurations and their corresponding stable translation directions. As the remained contact configurations satisfy the sufficient condition of pure translation, the tolerance range formed by two parallel arrows always encloses the CoM region and there must be no intersection between the two parallel arrows and the CoM region.

It is well established that the real vector space R^n can be spanned positively by a minimum of $n+1$ vectors. Hence, if there exist three stable pushing directions that can positively span R^2, any target direction can be achieved. An object can be translated through a single push if the target direction is the same as one of the three stable pushing directions. Alternatively, an object can be translated through a linear combination of two stable pushes. To determine whether to use one or two pushes to translate an object, we set a small threshold and find the set of two-edge-contact configurations whose cosine distance between the corresponding pushing directions and the target direction is below the specified threshold. After that, we remove the configurations which do not satisfy the sufficient conditions for pure translation from the set. Finally, if the set is still non-empty, we simply choose the one whose stroke distance is the largest, because it has the largest CoM tolerance range.

However, if the set after removal is empty, the target direction cannot be achieved using only one push. In this case, we use two configurations in a sequence whose corresponding pushing directions (referred to as $\mathbf{n}^i_d, \mathbf{n}^j_d$) can positively span the target direction. An optimal combination of two pushing directions can be found by minimizing the sum of travel distances. Let d_{target} denote the pushing distance and $\mathbf{n}_{target}$ denote the pushing direction. The pushing distance of each configuration can be found by solving Eq. 6.

$$\begin{aligned} \min_{\mathbf{n}^i_d, \mathbf{n}^j_d} \quad & d^i + d^j \\ \text{s.t.} \quad & [d^i, d^j]^\top = d_{target} [\mathbf{n}^i_d, \mathbf{n}^j_d]^{-1} \mathbf{n}_{target} \\ & d^i \geq 0 \\ & d^j \geq 0 \end{aligned} \tag{6}$$

where d^i and d^j are the corresponding pushing distances to achieve the target position. To minimize the travel distance to the target position, we choose two configurations with which the sum of pushing distances is minimal.

V. Experiment Setting

The experimental setup is shown in Fig. 11, where an RGB-D vision sensor is attached to the end link of the robot arm, capturing the overhead image of the object. We conducted two experiments; one was to validate the proposed method for CoM estimation, and the other one was to purely translate the object to the target position, leveraging the estimated CoM region and the modified ZMTEP method.

The experimental setup includes both high-friction and low-friction pushing. In high friction pushing, we use the pusher with a rubber wrap, while in low friction pushing, we simply remove the wrap. We 3D printed ten hollow objects in different shapes categorized into four groups based on the convexity and edge curvature as shown in Fig. 11. The hollow space inside each object is divided into multiple uniform grids where we

insert four lead blocks and change the CoM ground truth. Fig. 14 illustrates an instance of lead block configurations for each object. In the high friction setting, the frictional coefficient between the pusher and the objects is higher than 0.70. In the low friction setting, the frictional coefficient between the pusher and the objects is lower than 0.18.

We attach a textured surface paper on top of each object to monitor the change in pose of the pushed object with the Scale Invariant Feature Transform (SIFT) [42]. Prior to push initiation, the robotic arm adjusts its joint angles to ensure that a complete object shape is observable within the camera's field of view. Then, we remove the background from the image and register it that contains solely an object. The camera captures the pushing execution at the rate of 16.7 Hz. For each image frame obtained, we extract the keypoints and their descriptors. The projective transformation matrix is computed based on the corresponding keypoints between the registered image and each image frame using a brute force matcher and Lowe's test ratio. The difference in object pose from the pose before the object is pushed can be acquired by decomposing the projective transformation matrix. We compute the pose estimation error for object image, capturing multiple images of an object (whose pose remains unchanged with respect to the robot base) from different robot poses (or camera angles). We found that the position estimation error is negligible and the orientation estimation error is less than one degree.

VI. CoM Estimation Experiment

In this section, we assess the accuracy of our proposed CoM estimation method detailed in Alg. 1 and compare it with two baseline methods proposed by Lynch [14] and Li [17]. In [14], the author used a linear programming formulation to estimate the normalized magnitude of the friction forces between the object and the ground.

We discovered that a quadratic objective function yields more accurate results, and reformulated the problem as a quadratic programming problem. Therefore, we refer to the method in [14] as quadratic programming (QP). Additionally, we combine Alg. 1 and QP to improve the accuracy and effectiveness of an estimation.

A. Proposed CoM Estimation Experiment

Alg. 1 starts with finding the contact normal and pushing error bounds. This section explains the methods employed to compute the contact normal and pushing error bound. Furthermore, it outlines the process of determining the object sense of rotation based on the pose change of the pushed object, and steps for conducting experiments.

1) Contact Normal and Push Execution Error Bound: Firstly, we use the RGB-D image and deep CascadePSP model to obtain the object mask and contour points. We then remove the sharp points and their 10 adjacent pixels to reduce the estimation error. After that, we leverage 7 adjacent pixels for each contour point to estimate the contact normal error bound.

To construct the pushing execution error bound, we move the robot to each corner of the object and measure the position error. After multiple measurements, we found that the robot can move to each corner within position error $2mm$, and the difference between the command pushing direction and the real pushing direction is less than 3°. Therefore, we set d_e to $2mm$ and θ_e to 3° in Alg. 1.

2) Object Sense of Rotation: In the proposed CoM estimation pipeline, we set θ_{PT} to 1°, *i.e.*, if the accumulated change in orientation of the pushed object, denoted as $\Delta\theta$, is less than 1°, the motion is considered pure translation. If $\|\Delta\theta\| > \theta_{PT}$, we check the sign of $\Delta\theta$ to identify the object's sense of rotation. If we have a high frictional pusher and a convex polygonal object, the sense of rotation is determined according to the EVT 2 without the need to continuously track the object's pose.

3) Experiment: Striking a balance between outputs of Eq. 2 and Eq. 4 (or similarly Eq. 3 and Eq. 4), we set the weight vector $\mathbf{w}$ to $[0.5, 0.5]^{\top}$ in Alg. 1. Therefore, the CoM region does not tend to elongate. To evaluate the CoM estimation accuracy, we assign each object three different lead block configurations. For each configuration, we conducted three experiments in both high-friction and low-friction settings. In each run of Alg. 1, the object was pushed a total of eight times, with each push being executed $5cm$ along the estimated contact normal direction. After each push, the CoM region is updated, and the centroid of the updated CoM region is regarded as the estimated CoM. We refer to this method as **EVT-based CoM Estimation (EVT-CoM)**. The estimation procedure may be terminated early if no push contributes to narrowing the CoM region. We conducted 180 CoM estimation experiments in total.

B. Quadratic Programming

In [14], a set of hypothesis contact points between the object and the table should be determined. In our study, we uniformly selected contact points within the object mask. To ensure an accurate estimation of frictional distribution, we set the number of hypothesis contact points to a value greater than 20 based on the findings in [14]. Each hypothesis contact point has a corresponding non-negative weight, which is analogous to the normalized magnitude of the frictional forces. If the weights can be accurately estimated, the CoM of the object can be determined by summing the weighted positions of all hypothetical contact points.

Moreover, the contact points between the pusher and the object are uniformly selected in [14]. Following this idea, we uniformly sample contact points around the object's perimeter. The robot pushes the object $5cm$ along the estimated contact normal direction at each contact point. Based on the change in the pose of the object, the motion of each hypothesis contact point can be computed so that the direction of the frictional force at the hypothesis contact point can be estimated.

After synchronizing the pusher and object motion, a quadratic programming method is employed to minimize the magnitude of the moment at the contact point between the pusher and the object. This is based on the principle that the sum of the moments of frictional forces at the contact point between the pusher and the object should be zero under the assumption of quasi-static interaction. Valid solutions on

weights of hypothesis contact points should be non-negative, and the sum of weights is required to be one to avoid an all-zero solution. As the exact contact normal and friction cone at the contact point between the pusher and the object are unknown, there is no constraint on the direction of the resultant of all frictional forces applied to the object from the ground.

To evaluate the influence of the number of pushes on estimation accuracy, the robot pushes the object around its perimeter 25 times. We then uniformly sample the pushing instances, which is referred to as **QP Uniform Sampling (QP-uniform)**.

C. PushNet

PushNet [17] is a recurrent neural network trained in such a way to capture the object dynamics based on the history of pushing instances. Specifically, the input is the current object mask, start and end pusher positions in the image frame, and the output is two folds; the first one is the object CoM with respect to the image frame and the other one is an encoded feature vector for pushing action selection purpose. As we only compare the CoM estimation accuracy, we use the first output of the network. To make the pushing sequence for PushNet, we re-utilize the pushing instances collected in Section VI-B.

D. Combining the CoM region with Quadratic Programming

This study investigates the potential benefits of incorporating our proposed framework into quadratic programming to enhance estimation accuracy and efficiency. Instead of uniform sampling of the pushing instances, we select the pushing action in a manner that narrows down the CoM region. Additionally, we integrate the CoM region into the quadratic programming formulation as linear constraints to find the optimal solution. Specifically, the weighted sum of hypothesis positions must fall within the CoM region. To ensure a fair comparison of method performance, we utilize the previously collected data from Section VI-A. As each pushing motion is selected based on the current CoM region in each step, we refer to this method as **QP EVT-based Sampling (QP-EVT)**.

E. CoM Estimation Result

The estimation error is illustrated in Fig. 12. In general, it can be observed that the estimation accuracy of PushNet does not improve as the number of pushes increases. One potential explanation for this is that PushNet is trained on a simulation dataset, which introduces a sim-to-reality gap that hinders its generalization to our experimental settings. Conversely, the other methods demonstrate a reduction in estimation error with an increasing number of pushes. **EVT-CoM** consistently achieves a lower estimation error after each push, when compared to the **QP-uniform**, and ultimately attains a lower estimation error. However, it is worth noting that **QP-EVT** achieves the lowest estimation error or equivalent performance to the **EVT-CoM**. Furthermore, our findings suggest that **QP-EVT** can achieve comparable or superior results with only two pushes.

Fig. 13 illustrates the final estimation errors of all methods, excluding PushNet, for different categorical objects in both high and low friction scenarios. The mean and standard deviation of PushNet are found to be 63.1mm and 26.5mm, respectively, which is incomparably worse than the others. According to Fig. 13, we found that **EVT-CoM** has similar performance in different frictional settings. **QP-EVT** achieves a lower estimation error in the high friction scenario. In addition, comparing two quadratic programming-based methods, **QP-EVT** benefits significantly from our push action selection strategy in reducing estimation errors, regardless of the frictional setting.

Fig. 14 illustrates the results of the CoM region identification for all objects with one location configuration of lead blocks both in high and low friction settings. Each sub-figure shows an image overlay of three runs of Algorithm 1. Due to the pushing execution errors, the three CoM regions are not exactly identical. However, our proposed algorithm takes the contact normal and pushing execution errors into account to ensure that the CoM ground truth is inside the CoM region. As a result, the CoM ground truth is also contained inside the intersection of three CoM regions. Notably, it is observed that the CoM regions estimated in high frictional settings tend to be smaller than those in low frictional settings. This is because there is no need to deal with the uncertainty in contact normals in high frictional settings.

The CoM regions for convex polygonal objects with the other two lead block location configurations in high frictional settings are depicted in Fig. 15. This is to confirm the validity of the EVT 2. It is noticeable that the CoM ground truth remains within the intersection of three CoM regions even though the estimation is solely based on the difference between the initial and final poses of the object.

VII. Object Translation Experiment

This study conducts object translation experiments with high frictional pushing to verify the validity of Eq. 5. Following the completion of the CoM estimation experiments, three CoM regions are obtained for each object in all three lead block configurations. The intersected CoM region and estimated contact normals are then utilized to select two-edge-contact pushing configurations. The stable pushing directions can positively span R^2 space visualized in Fig. 16, where the two-edge-contact configurations and corresponding pushing directions are depicted in the same color. In order to assess the effectiveness of object pure translation, the robot is tasked to push each object in nine different directions. The angles between the pushing direction and the horizontal direction in the image frame are evenly distributed within the range of $(-180^{\circ}, 180^{\circ})$. To determine the factors that influence the accuracy of the modified ZMTEP, the object is pushed with varying distances (100mm and 180mm) and speeds (30mm/s and 150mm/s). Throughout the pushing process, the rotations and translations relative to the pusher are recorded. Overall, the translation experiment is repeated a total of 1080 times.

The mean and standard deviations of translation and rotation errors for pure translation experiments are presented in Table I. It can be observed that the pose error is generally small for all object types, indicating that our ZMTEP method translates

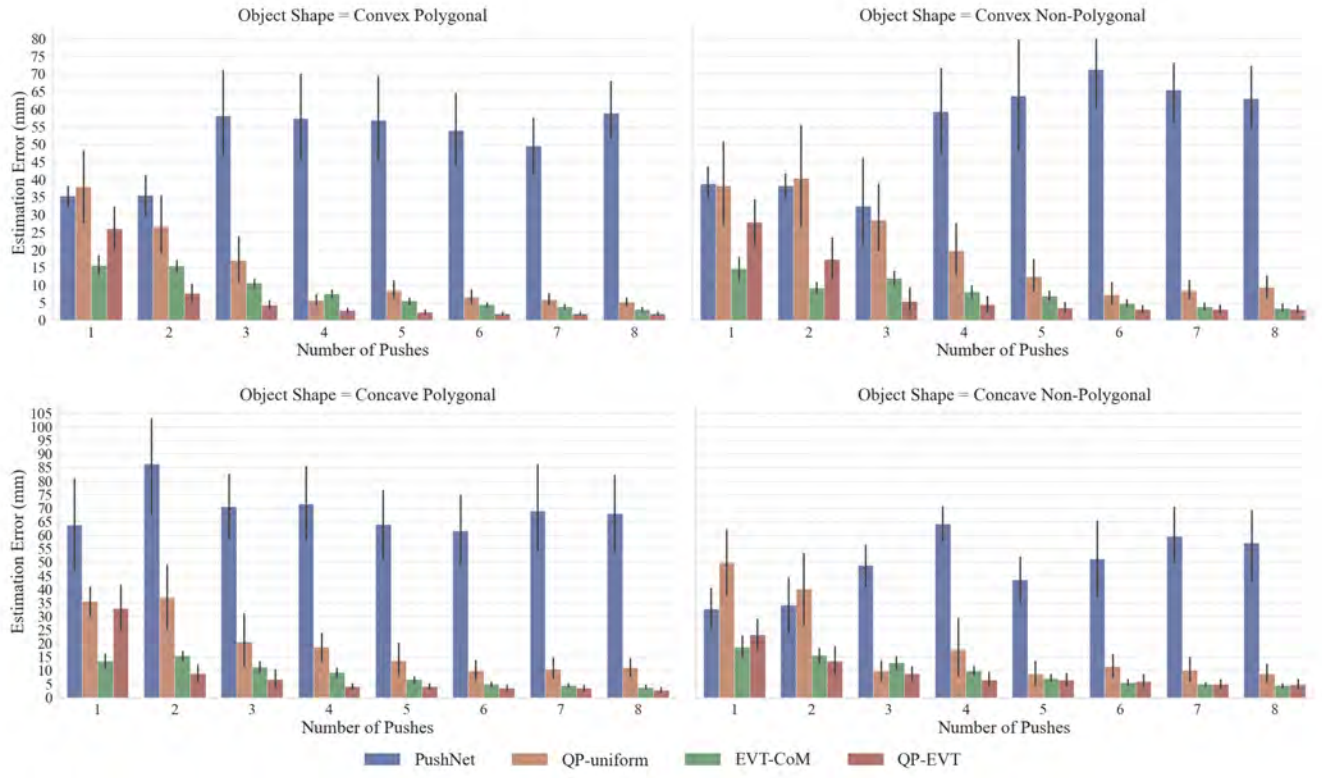

Fig. 12. Error in estimating the CoM of objects after each push. Each sub-figure displays the estimation error for objects within the same category. The mean estimation error is visually represented by the height of each bar, while the solid black line at the middle of each bar represents the standard deviation.

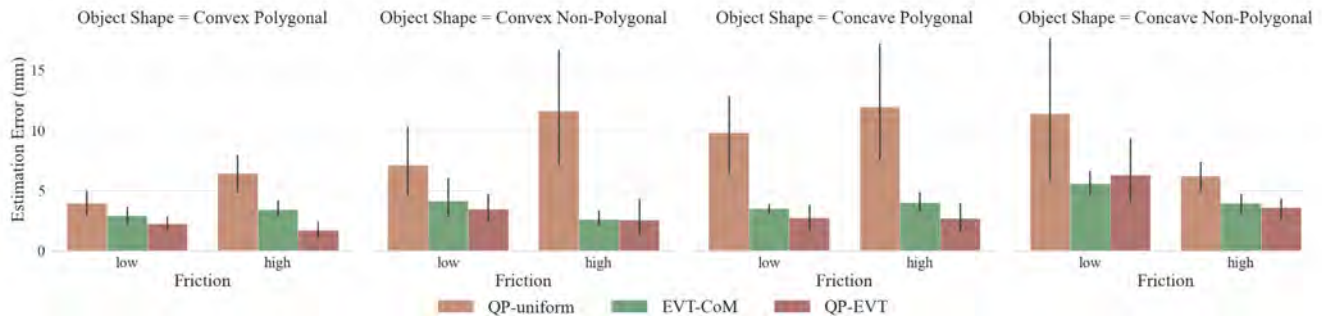

Fig. 13. Estimation error after the last push under both high and low frictional settings.

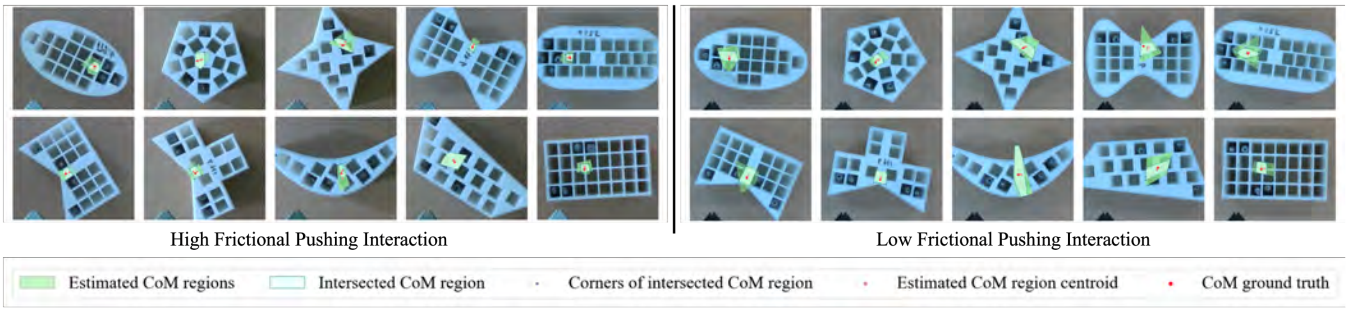

Fig. 14. Illustrations of CoM regions for both high and low frictional pushing interactions. Each image shows one lead block configuration and three overlapping CoM regions.

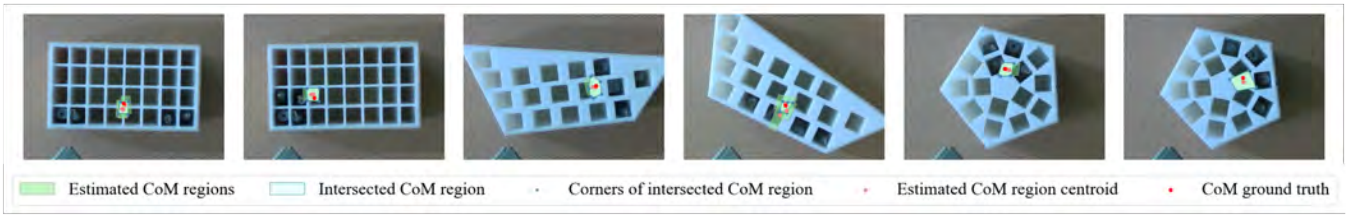

Fig. 15. Illustrations of CoM regions of the convex polygonal objects with the other two lead-block configurations in the high frictional setting.

Fig. 16. Pushing directions that are available for each test object based on the intersected CoM region. Two-edge-contact pushing configurations are illustrated on the boundary of the object. The corresponding directions are illustrated on the centroid of the CoM region, using the same color.

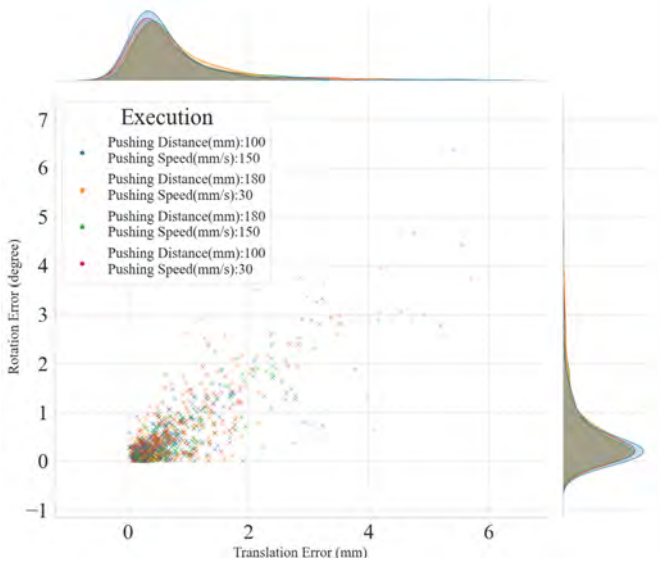

Fig. 17. Joint plot of translation and rotation error.

TABLE I
TRANSLATION AND ROTATION ERRORS

Object Category	Translation Error(mm)		Rotation Error(degree)	
	mean	std	mean	std
Convex Polygonal	0.72	0.84	0.52	0.77
Convex Non-Polygonal	0.72	0.79	0.48	0.57
Concave Polygonal	0.58	0.41	0.48	0.62
Concave Non-Polygonal	1.11	1.02	0.76	0.85

novel objects without changing their orientations to the target position using the CoM region and contact normal estimates. However, for concave non-polygonal objects (in particular, slender objects), the object pose error is relatively larger compared with other object types. This can be attributed to the potential slippage of the pusher with a small gripping force when the target direction significantly deviates from the pusher's contact normal directions.

Fig. 17 shows the joint distributions of translation and rotation errors of the pushed object. It can be seen from the figure that the performance of the proposed method does not depend on translation distance and/or pushing speed to a noticeable degree. To be more specific, the change in pushing distance and speed has a limited effect on the final pose of the pushed object. In contrast to almost all previous work relying on quasi-static pushing, our method enabled to translate arbitrarily shaped objects with variable speeds up to $150mm/s$ under our experimental setup. However, from a practical point of view, we found difficulties in controlling two fingers of the parallel-jaw gripper to simultaneously touch the object's different edges. Therefore, a slight change in the object's pose occurs during the initial touching. We conjecture that this asynchrony between two fingers of initial touching could be the main factor that leads to the error.

VIII. CONCLUSION

In this work, extending Mason's Voting Theorem, we proposed a pusher motion selection strategy and CoM region update rule to efficiently find a convex region containing the

CoM ground truth given noisy object shape information. With the nominal friction cone defined for novel objects whose actual value is not known, we kept the pusher motion at the contact point aligned with the contact normal. As the exact contact normal was difficult to extract from the vision sensor, we found a contact normal error bound acting as an upper bound of the rotation uncertainty cone and eliminated the CoM estimation error due to the inaccurate contact normal.

We presented two Extended Voting Theorems widely applicable to real-world settings. The first theorem states that the uncertainty in the direction of contact normal will not introduce any error in the CoM estimation procedure if the friction cone at the contact point is larger than the contact normal error bound. The second theorem states that continuous pushing in a straight line will not introduce the CoM estimation error if the object is a polygon and the pushing direction lies initially inside the friction cone at the contact point. We conducted a large number of the CoM estimation experiments using a parallel-jaw gripper as the pusher and ten CoM-controllable objects in real settings. We compared our CoM estimation method with two baseline methods. The experimental results indicated that our method outperformed the baselines in terms of estimation accuracy. Furthermore, our estimation pipeline combined with quadratic programming achieved the lowest estimation error, significantly reducing the required number of pushes. In addition, we have proven the validity of the EVT 2 by showcasing the estimated CoM region for polygonal convex objects in a high frictional setting. It is worth noting that this estimation does not entail any implementing measures for object tracking or considering relative sliding between the pusher and the object. Despite these simplifications, the estimated CoM region encompassed the true CoM position.

Based on the estimated CoM region and the modified ZMTEP method, we sampled a set of two-edge contact configurations that can tolerate the CoM estimation error. In contrast to existing methods which usually make use of full state feedback or multiple short-length pushes, we show that if the stable pushing directions positively span R^2, our proposed method can efficiently translate the object using at most two pushes regardless of the pushing direction and pushing magnitude.

Currently, the second EVT does not guarantee the unique sense of rotation for non-polygonal objects. In the future, toward extending our method to deal with the CoM estimation of non-polygonal objects, we will explore different ways of pushing or integrating the pose estimation of the pushed object. This work has focused on rearranging the position of the object. Optimizing the contact configuration for rearranging both the position and orientation of the object is also a potential direction.

REFERENCES

[1] Z. Gao, A. Elibol, and N. Y. Chong, "Zero moment two edge pushing of novel objects with center of mass estimation," *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 3, pp. 1487–1499, 2023.

[2] C. Song and A. Boularias, "Object rearrangement with nested nonprehensile manipulation actions," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 6578–6585.

[3] B. Huang, S. D. Han, A. Boularias, and J. Yu, "Dipn: Deep interaction prediction network with application to clutter removal," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4694–4701.

[4] A. Eitel, N. Hauff, and W. Burgard, "Learning to singulate objects using a push proposal network," in *Robotics Research: The 18th International Symposium ISRR*. Springer, 2020, pp. 405–419.

[5] M. Danielczuk, J. Mahler, C. Correa, and K. Goldberg, "Linear push policies to increase grasp access for robot bin picking," in *IEEE International Conference on Automation Science and Engineering*, 2018, pp. 1249–1256.

[6] N. Chavan-Dafle, R. Holladay, and A. Rodriguez, "Planar in-hand manipulation via motion cones," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 163–182, 2020. [Online]. Available: https://doi.org/10.1177/0278364919880257

[7] N. Mavrakis and R. Stolkin, "Estimation and exploitation of objects' inertial parameters in robotic grasping and manipulation: A survey," *Robotics and Autonomous Systems*, vol. 124, no. 103374, 2020.

[8] Y. Yu, K. Fukuda, and S. Tsujio, "Estimation of mass and center of mass of graspless and shape-unknown object," in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, vol. 4, 1999, pp. 2893–2898 vol.4.

[9] Y. Yu, T. Arima, and S. Tsujio, "Estimation of object inertia parameters on robot pushing operation," in *IEEE International Conference on Robotics and Automation*, 2005, pp. 1657–1662.

[10] B. Sundaralingam and T. Hermans, "In-hand object-dynamics inference using tactile fingertips," *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1115–1126, 2021.

[11] C. Wang, S. Wang, B. Romero, F. Veiga, and E. Adelson, "Swingbot: Learning physical features from in-hand tactile exploration for dynamic swing-up manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5633–5640.

[12] K. Yao, M. Kaboli, and G. Cheng, "Tactile-based object center of mass exploration and discrimination," in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, 2017, pp. 876–881.

[13] Z. Zhao, X. Li, C. Lu, and Y. Wang, "Center of mass and friction coefficient exploration of unknown object for a robotic grasping manipulation," in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2018, pp. 2352–2357.

[14] K. M. Lynch, "Estimating the friction parameters of pushed objects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1993, pp. 186–193.

[15] M. Artashes and D. Burschka, "Visual estimation of object density distribution through observation of its impulse response," in *VISAPP*, 2013.

[16] C. Song and A. Boularias, "A probabilistic model for planar sliding of objects with unknown material properties: Identification and robust planning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 5311–5318.

[17] J. K. Li, W. S. Lee, and D. Hsu, "Push-net: Deep planar pushing for objects with unknown physical properties," in *Robotics: Science and Systems*, 2018.

[18] K. N. Kumar, I. Essa, S. Ha, and C. K. Liu, "Estimating mass distribution of articulated objects using non-prehensile manipulation," *arXiv preprint arXiv:1907.03964*, 2019.

[19] Z. Xu, J. Wu, A. Zeng, J. B. Tenenbaum, and S. Song, "Densephysnet: Learning dense physical object representations via multi-step dynamic interactions," in *Robotics: Science and Systems*, 2019. [Online]. Available: http://www.zhenjiaxu.com/DensePhysNet/

[20] Z. Gao, A. Elibol, and N. Y. Chong, "Estimating the center of mass of an unknown object for nonprehensile manipulation," in *IEEE International Conference on Mechatronics and Automation*, 2022, pp. 1755–1760.

[21] ——, "Planar pushing of unknown objects using a large-scale simulation dataset and few-shot learning," in *IEEE International Conference on Automation Science and Engineering*, 2021, pp. 341–347.

[22] S. Tanaka, T. Tanigawa, Y. Abe, M. Uejo, and H. Tanaka, "Active mass estimation with haptic vision," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, 2004, pp. 256–261 Vol.3.

[23] T. N. Le, F. Verdoja, F. J. Abu-Dakka, and V. Kyrki, "Probabilistic surface friction estimation based on visual and haptic measurements," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2838–2845, 2021.

[24] M. Veres, I. Cabral, and M. Moussa, "Incorporating object intrinsic features within deep grasp affordance prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6009–6016, 2020.

[25] S. Akella and M. T. Mason, "Posing polygonal objects in the plane by pushing," *The International Journal of Robotics Research*, vol. 17, no. 1, pp. 70–88, 1998. [Online]. Available: https://doi.org/10.1177/027836499801700107
[26] Q. Li and S. Payandeh, "Manipulation of convex objects via two-agent point-contact push," *The International Journal of Robotics Research*, vol. 26, no. 4, pp. 377–403, 2007. [Online]. Available: https://doi.org/10.1177/0278364907076819
[27] K. Lynch, H. Maekawa, and K. Tanie, "Manipulation and active sensing by pushing using tactile feedback," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1992, pp. 416–421.
[28] S. Goyal, A. Ruina, and J. Papadopoulos, "Planar sliding with dry friction part 1. limit surface and moment function," *Wear*, vol. 143, no. 2, pp. 307–330, 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0043164891901043
[29] ——, "Planar sliding with dry friction part 2. dynamics of motion," *Wear*, vol. 143, no. 2, pp. 331–352, 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0043164891901054
[30] F. Bertoncelli, F. Ruggiero, and L. Sabattini, "Linear time-varying mpc for nonprehensile object manipulation with a nonholonomic mobile robot," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 11 032–11 038.
[31] J. Zhou, Y. Hou, and M. T. Mason, "Pushing revisited: Differential flatness, trajectory planning, and stabilization," *International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1477–1489, 2019. [Online]. Available: https://doi.org/10.1177/0278364919872532
[32] F. R. Hogan, E. R. Grau, and A. Rodriguez, "Reactive planar manipulation with convex hybrid mpc," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 247–253.
[33] M. Bauza, F. R. Hogan, and A. Rodriguez, "A data-efficient approach to precise and controlled pushing," in *Conference on Robot Learning*. PMLR, 2018, pp. 336–345.
[34] Z. Gao, A. Elibol, and N. Y. Chong, "A few-shot learning framework for planar pushing of unknown objects," *Intelligent Service Robotics*, vol. 15, no. 3, pp. 335–350, 2022.
[35] C.-Y. Chai, W.-H. Peng, and S.-L. Tsao, "Object rearrangement through planar pushing: A theoretical analysis and validation," *IEEE Transactions on Robotics*, vol. 38, no. 5, pp. 2703–2719, 2022.
[36] J. Lloyd and N. F. Lepora, "Goal-driven robotic pushing using tactile and proprioceptive feedback," *IEEE Transactions on Robotics*, pp. 1–12, 2021.
[37] F. Bertoncelli, M. Selvaggio, F. Ruggiero, and L. Sabattini, "Task-oriented contact optimization for pushing manipulation with mobile robots," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 1639–1646.
[38] A. Pasricha, Y.-S. Tung, B. Hayes, and A. Roncone, "Pokerrt: Poking as a skill and failure recovery tactic for planar non-prehensile manipulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4480–4487, 2022.
[39] Z. Dong, S. Krishnan, S. Dolasia, A. Balakrishna, M. Danielczuk, and K. Goldberg, "Automating planar object singulation by linear pushing with single-point and multi-point contacts," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2019, pp. 1429–1436.
[40] H. K. Cheng, J. Chung, Y.-W. Tai, and C.-K. Tang, "CascadePSP: Toward class-agnostic and very high-resolution segmentation via global and local refinement," in *CVPR*, 2020.
[41] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," in *Proceedings of the 19th annual conference on computer graphics and interactive techniques*, 1992, pp. 71–78.
[42] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2. Ieee, 1999, pp. 1150–1157.

Ziyan Gao received the B.S. in mechanical engineering from the Qingdao University of Science and Technology, Qingdao, China, in 2016. He received the M.S. and Ph.D. degrees in information science from the Japan Advanced Institute of Science and Technology, Ishikawa, Japan, in 2019 and 2022, respectively. He is currently a post-doctoral researcher with the Japan Advanced Institute of Science and Technology. His research interests include object physical parameter estimation, non-prehensile manipulation, robotic grasping, and deep neural networks.

Armagan Elibol is a senior researcher at the Institute for Advanced Simulation, IAS-8, Forschungszentrum Juelich in Germany. He obtained a Ph.D. in Computer Science (2011) from the Computer Vision and Robotics Group (VICOROB) of the University of Girona, Spain, and both a B.Sc. and M.Sc. in Mathematical Engineering from Yildiz Technical University in Turkiye, respectively 2002 and 2004. He has worked in different universities in Japan, South Korea, and Turkiye. His research interests include the wide area of Computer Vision and Robotics.

Nak Young Chong received the B.S., M.S., and Ph.D. degrees from Hanyang University, Seoul, Korea, in 1987, 1989, and 1994, respectively. From 1994 to 2003, he was with Daewoo Heavy Industries, Korea Institute of Science and Technology, Mechanical Engineering Laboratory-AIST/MITI, and Intelligent Systems Institute-AIST. In 2003, he joined the faculty of JAIST, where he currently is a professor of information science and served as a councilor, director of the Center for Intelligent Robotics, and chair professor of Intelligent Robotics Group. He was a visiting scholar at Northwestern University, Georgia Institute of Technology, University of Genoa, and Carnegie Mellon University, and served as an associate faculty at the University of Nevada Las Vegas and Kyung Hee University. He serves/served as editor of the IEEE Robotics and Automation Letters, Intelligent Service Robotics, and International Journal of Advanced Robotic Systems, and associate editor of the IEEE Transactions on Robotics. He served as program (co)-chair for JCK Robotics 2009, ICAM 2010, IEEE Ro-Man 2011/2013/2022, IEEE CASE 2012, URAI 2013/2014, DARS 2014, ICCAS 2016, IEEE ARM 2019. He was a general (co)-chair of URAI 2017 and UR 2020. He also served as co-chair for IEEE RAS Networked Robots TC from 2004 to 2006, and Fujitsu Scientific System WG from 2004 to 2008.