

Action Planning for Packing Long Linear Elastic Objects Into Compact Boxes With Bimanual Robotic Manipulation

Wanyu Ma , *Graduate Student Member, IEEE*, Bin Zhang , Lijun Han , Shengzeng Huo , *Graduate Student Member, IEEE*, Hesheng Wang , *Senior Member, IEEE*, and David Navarro-Alarcon , *Senior Member, IEEE*

Abstract—In this article, we propose a new action planning approach to automatically pack long linear elastic objects into common-size boxes with a bimanual robotic system. For that, we developed a hybrid geometric model to handle large-scale occlusions combining an online vision-based method and an offline reference template. Then, a reference point generator is introduced to automatically plan the reference poses for the predesigned action primitives. Finally, an action planner integrates these components enabling the execution of high-level behaviors and the accomplishment of packing manipulation tasks. To validate the proposed approach, we conducted a detailed experimental study with multiple types and lengths of objects and packing boxes.

Index Terms—3-D point clouds, action planning, automatic packing, elastic objects, robotic manipulation.

e_{in}, e_{out} Shape differences of \mathbf{P} and \mathbf{P}^O to \mathbf{P}^* , respectively.
 e, e^* Total shape difference and its desired value.
 \mathbf{x} End-effector feedback pose.
 \mathbf{x}^* End-effector reference pose.
 \mathbf{u} End-effector target pose.
 $\{F\}$ Reference template frame.
 $\{FO\}$ Object body frame.
 $\Delta h, \Delta f$ Height offsets for hover and fixing action primitives.
 δ_i, δ_f Horizontal distances for the reference point generator.
 $\mathbf{p}^L, \mathbf{p}^G, \mathbf{p}^F$ Reference points for object placing, grasping, and fixing.

NOMENCLATURE

Symbol	Quantity
$\mathcal{O}(\eta, l_O, d_O)$	LEO of material η , length l_O , and rod diameter d_O .
$\mathcal{B}(l_B, w_B, h_B)$	Cuboid container of dimensions $l_B \times w_B \times h_B$.
\mathbf{P}^*	Offline reference template.
\mathbf{P}	Raw feedback point cloud inside the box.
\mathbf{P}^O	Ordered skeleton of the point cloud outside the box.
$\hat{\mathbf{p}}_i$	Corresponding point in \mathbf{P}^O to the i th point in \mathbf{P}^* .
\mathbf{p}_s^*	Split point.

I. INTRODUCTION

THE social distancing requirements imposed by the COVID-19 pandemic has forced many businesses to adopt online retail platforms. Recent reports [1] indicate that the pandemic has accelerated the shift away from physical to online stores by roughly five years. It is predicted that at the current pace, many aging societies will not have sufficient human workers within a decade to sort and pack products; human manual packing is simply unsustainable for these communities. A feasible strategy to deal with this record increasing demand for products and diverse commodities (whose heterogeneous properties may vary from compliant, to articulated, to deformable [2]) is to use dexterous robots that can automate the soft packing process. This approach can also help to optimize the current manual practice in the industry (which tends to use excessive packaging that wastes materials) and to improve social distancing (as no human workers are needed). Our goal in this article is precisely to develop efficient manipulation strategies that can automate the packing problem of a type of deformable objects with special properties resulting in high difficulty of its packing task.

To advance in the development of these valuable manipulation skills [3], [4], in this article, we focus on the challenging problem where a (long) linear elastic object (LEO) needs to be autonomously grasped, shaped/deformed, and placed within a compact box that optimizes its packing space, as depicted in

Manuscript received 21 April 2022; revised 24 September 2022; accepted 16 November 2022. Recommended by Technical Editor R. Carloni and Senior Editor N. G. Tsagarakis. This work was supported by the Research Grants Council (RGC) of Hong Kong under Grant 14203917. (Corresponding author: David Navarro-Alarcon.)

Wanyu Ma, Bin Zhang, Shengzeng Huo, and David Navarro-Alarcon are with the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mail: wanyu.ma@connect.polyu.hk; me-bin.zhang@connect.polyu.hk; kyle-sz.huo@connect.polyu.hk; david.navarro-alarcon@polyu.edu.hk).

Lijun Han and Hesheng Wang are with the Department of Automation, Shanghai Jiatong University, Shanghai 200240, China (e-mail: lijun_han@sjtu.edu.cn; wanghesheng@sjtu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMECH.2022.3223990>.

Digital Object Identifier 10.1109/TMECH.2022.3223990

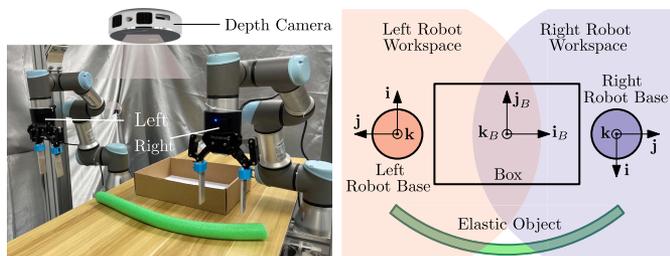


Fig. 1. Experimental setup with a box frame $\{F_B\} = \{i_B, j_B, k_B\}$, two robot arms, and the LEO, and a top-view camera.

Fig. 1. There are two main challenges that arise with the automation of this task: 1) due to the complexity of the shaping task (which is difficult to perform with a single continuous motion), several coordinated actions by collaborative arms are required to effectively deform and place the object within the box; 2) the typically occluded view from vision sensors during the task leads to partial observations of the manipulated object and the environment (this results in incomplete geometric information that complicates the real-time guidance of the robot's motion).

A. Related Work

1) *Robotic Packing in Logistics*: Although there has been a strong push toward robotizing the processing of product, e.g., with automated guided vehicles in distribution centers [5], packing remains a task entirely performed by human workers. Recently, many methods have been developed for the Amazon picking challenge to automatically recognize, collect, and transfer multiple types of products into boxes [6], [7], [8]. Note that the majority of these methods do not address (and underestimate) large-scale elastic deformations, e.g., those exhibited by LEOs; to optimize packing space, shape control is needed to transfer and arrange LEOs into compact boxes. The few works that do consider the arrangement of highly deformable materials [9] do not address shape control and are mostly confined to simple numerical simulations. As the booming e-commerce industry now extends to many nontraditional commodities (e.g., deformable groceries and household products [10]), it is essential to develop shape control methods that can deform highly elastic materials and, thus, save packing space. However, this challenging soft packing problem has not been sufficiently studied in the literature.

2) *Action Planning for Packing Tasks*: In contrast with traditional (low level) control methods for manipulating soft objects based on *continuous* trajectories [11] (i.e., with a single action), the robotic packing of a LEO requires to use *discrete* task planning with multiple (high level) actions. These types of methods decompose and plan the task in terms of a coordinated sequence of action primitives, each of which captures a specific motor behavior (this approach has been used in a wide range of applications, e.g., grasping [12], soccer [13], and assembly [14]). Action primitives methods have been proposed for packing and object arrangement problems, e.g., Schwar et al. [15] developed a controller for robotic picking and stowing tasks based on parameterized motion primitives, Zeng et al. [16] proposed a

method for manipulating objects into tightly packed configurations by learning pushing/grasping policies, and Capitanelli et al. [17] tackled the problem of reconfiguring articulated objects by using an ordered set of actions executed by a dual-arm robot. Yet, note that the action primitives adopted by these works cannot capture the complex behaviors that are needed to control the shape of a LEO during a packing task.

3) *Representation of Deformable Objects*: To visually guide the manipulation task, it is necessary for a controller to have a meaningful representation of the object. To this end, researchers have developed a variety of representation methods, e.g., physics-based approaches [18], [19], [20], [21], [22] (using mass-spring-damping models and finite element method), based on visual geometric features [23], [24], [25], [26] (using points, angles, curvatures, catenaries, contours moments, etc.), or data-driven representations [27], [28], [29], [30] (using Fourier series, FPFH, PCA, autoencoders, etc.). Typically, vision-based approaches are strongly affected by occlusions during the task (this is problematic for packing as a top observing camera will have incomplete observations during the task). To deal with this issue, many research works have addressed the estimation and tracking of the object's deformation in real time [31], [32], [33]. Yet, these works only consider with 2-D scenarios, hence, are not applicable to our 3-D LEO manipulation problem.

B. Our Contribution

In this article, we propose the following:

- 1) a new hybrid geometric model that combines online 3-D vision with an offline reference template to deal with camera occlusions,
- 2) a reference point planner that provides intermediate targets to guide the high-level packing actions, and
- 3) a cyclic action planner that coordinates multiple action primitives of a dual-arm robot to perform the complex LEO packing task.

The proposed methodology is original, and its demonstrated capabilities have not (to the best of our knowledge) been previously reported in the literature. To validate this new approach, we report a detailed experimental study with a dual-arm robot performing packing tasks with LEOs of various elastic properties.

The rest of this article is organized as follows. Section II presents the mathematical models. Section III describes the hybrid geometric model. Section IV presents the packing method, Section V reports the experiments. Finally, Section VI concludes this article.

II. MODELING

The key nomenclature used in this article is available in the Nomenclature.

A. Geometric Object Modeling

In our method, we use an RGB-D camera to capture point clouds of the scene in real time. During the task, the raw point cloud is split into two structures: P^O , which represents the

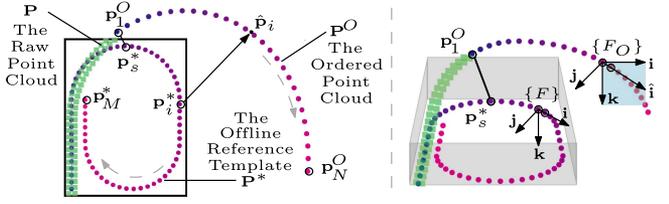


Fig. 2. Proposed hybrid geometric model of a LEO. Blue indicates the start curve and pink indicates its end. The axes \mathbf{i} , \mathbf{k} , and $\hat{\mathbf{i}}$ are all coplanar. Three sets of point clouds \mathbf{P} , \mathbf{P}^* , and \mathbf{P}^O , and two frames F and F_O are introduced.

object's part to be grasped by the robot, and \mathbf{P} , which represents the object's part already packed; spatially, these two structures correspond to the object's parts outside and inside the box, respectively. To provide an intuitive topology that facilitates the LEO's manipulation, the points in \mathbf{P}^O are ordered along the linear object's centerline. During initialization, the object's length l_O and width d_O are computed from the raw point cloud data. Note that, after initialization, l_O can update while d_O is fixed.

The offline reference template $\mathbf{P}^* = [\mathbf{p}_1^*, \dots, \mathbf{p}_M^*] \in \mathbb{R}^{3 \times M}$ is a predesigned geometric curve that represents the final configuration to be given to the object $\mathcal{O}(\eta, l_O, d_O)$ within the box $\mathcal{B}(l_B, w_B, h_B)$. Similarly to the raw feedback point cloud, the reference template \mathbf{P}^* is also separated into two parts at the split point \mathbf{p}_s^* , corresponding to \mathbf{P} and \mathbf{P}^O . Given the i th point \mathbf{p}_i^* on \mathbf{P}^* , we denote its corresponding point at \mathbf{P}^O as $\hat{\mathbf{p}}_i^*$. The length from $\hat{\mathbf{p}}_i^*$ to the end of the object \mathbf{p}_N^O equals the length from \mathbf{p}_i^* to \mathbf{p}_M^* . Note that the method to compute corresponding points depends. The split point \mathbf{p}_s^* is the s th point on the offline template \mathbf{P}^* where the length from \mathbf{p}_s^* to \mathbf{p}_M^* equals to the length of \mathbf{P}^O .

For the point clouds \mathbf{P}^* and \mathbf{P}^O , two important frames, the reference template frame, and the object body frame are defined. The reference template frame is denoted as $\{F\}$ at $\mathbf{p}_i^*, i = 1, \dots, M - 1$. Its z -axis is the unit vector vertically pointing to the table. The x -axis is the unit vector pointing to \mathbf{p}_{i+1}^* , and the y -axis is determined by the right-hand principle. The object body frame is denoted as $\{F_O\} = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ at a point in $[\mathbf{p}_i^O, \mathbf{p}_{i+1}^O], i = 1, \dots, N - 1$. Similarly, its z -axis is the unit vector vertically pointing to the table. Then, we introduce the unit tangent vector $\hat{\mathbf{i}}$ pointing from \mathbf{p}_i^O to \mathbf{p}_{i+1}^O . The y -axis \mathbf{j} is orthogonal to the plane $\mathbf{k} - \hat{\mathbf{i}}$. Note that, since the object is not in a plane parallel to the table, $\hat{\mathbf{i}}$ is not always orthogonal to \mathbf{k} . So the real x -axis of $\{F_O\}$ is determined by $\mathbf{i} = \mathbf{j} \times \mathbf{k}$. These definitions are depicted in Fig. 2.

To compute shape difference e between the feedback point cloud and the reference template, we must introduce the inside and outside the box errors between these two structures. For that, we define e_{in} as the average minimum Euclidean distance (i.e., the similarity [34]) between \mathbf{P} and the first s points of \mathbf{P}^* , and define as e_{out} the average Euclidean distance between the other $M - s$ points of \mathbf{P}^* and their corresponding points $\hat{\mathbf{p}}_i$. The total shape difference e is computed as a weighted combination of these two errors

$$e = we_{in} + (1 - w)e_{out} \quad (1)$$

for $w = \frac{s}{M}$ as normalization weight. This metric quantifies the accuracy of the automatic shaping process. Note that e_{in} and e_{out} are, respectively, averaged based on the number of points in \mathbf{P} and \mathbf{P}^O ; thus, they represent the distances between two pairs of points. Therefore, these errors are not significantly influenced if some feedback points are lost due to occlusion. Besides, e is weighted based on the number of points at the two sides of \mathbf{p}_s^* . This guarantees that the contribution of e_{in} and e_{out} is normalized. As the feedback point clouds of the LEO represent points over its surface whereas the reference template represents a centerline, therefore, the error e will ideally converge to a desired value $e^* = \frac{d_O}{2}$, i.e., half width of the object.

B. Action Planning

The robotic system considered in this study is composed of two end-effectors, one with an active grasp role and the other with an assistive fix role. It is assumed that the end-effectors are vertically pointing toward the box's plane. The configuration of the robotic arms is represented by a 4-degrees-of-freedom (DOF) pose vector $\mathbf{x} = [x, y, z, \theta]^T$ (comprised of position and orientation coordinates) and 1 DOF for the gripper's open/close configuration. The initial pose $\mathbf{x}(t_0)$ of the robot arms is assumed to be above the box and object.

The action planner that coordinates multiple action primitives performed by the robot arms is modeled with a classic state machine [35]. This state machine is represented by the tuple (S, T, A, G, R) , whose elements are defined as follows:

- 1) A : collection of action primitives for the end-effectors' 4-DOF pose.
- 2) G : collection of action primitives for the grippers' 1-DOF open/close configuration.
- 3) R : collection of the robot active and assistant roles.
- 4) S : collection of action planner states, each represented by a robot movement.
- 5) T : state transition function.

Aiming at recycling the modules of the state machine, the designed action primitives compose a periodic action planner loop that enables the robot to automatically perform the complex packing task.

C. Framework Overview

The overview of the proposed automatic packing approach is shown in Fig. 3. It is composed of a hybrid geometric model (green block), a reference point generator (orange block), and an action planner (blue block). The hybrid geometric model provides a robust representation of the object by combining an online part (\mathbf{P} and \mathbf{P}^O) and an offline part (\mathbf{P}^*). The reference point generator computes reference poses \mathbf{x}^* for the robot to perform grasping, placing, and fixing the object. The action planner commands the execution of the task based on a series of action primitives. The robotic platform (pink block) receives and executes the kinematic motion command (i.e., the target poses and gripper configurations) from the action planner, and returns an end flag to the control system after its completion. The state machine recycles a periodic action planner loop by alternating

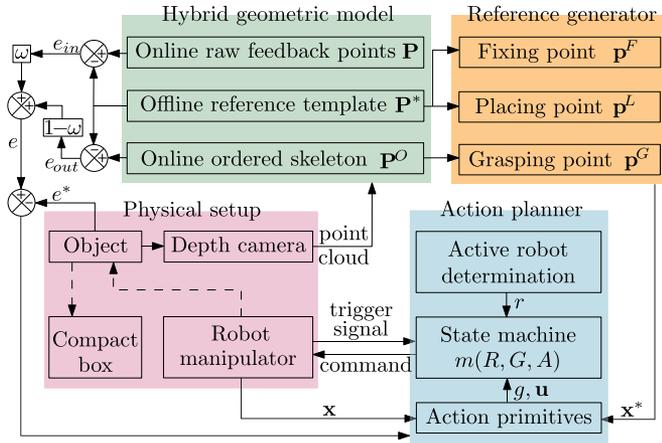


Fig. 3. Framework of the proposed approach for packing long LEOs into common-size boxes. Four components (a hybrid geometric model, a reference point generator, an action planner, and physical setup) are identified in different colors. Solid lines indicate data transmission and dashed lines indicate physical contact.

each robot arm between an active and an assistant role until the task is completed.

III. HYBRID GEOMETRIC MODEL

The proposed hybrid geometric model consists of \mathbf{P} (the raw feedback point cloud inside the box), \mathbf{P}^O (the ordered skeleton of the object's part outside the box), and \mathbf{P}^* (the generated offline reference template). It extracts the object's geometry in real time and generates the suitable target shape for packing LEOs, which are preconditions of reference point generation and packing progress measurement. On the one hand, the reference point generator replaces \mathbf{P} with the corresponding points (\mathbf{p}_1^* to \mathbf{p}_s^*) of \mathbf{P}^* and searches for the reference points in \mathbf{P}^O and \mathbf{P}^* , to deal with the typical occlusions that result from the grippers blocking the top-view camera. On the other hand, the shape difference e is computed as the combination of e_{in} (the distance from \mathbf{P} to the template points \mathbf{p}_i^* , $i = 1, \dots, s$) and e_{out} (the distance from \mathbf{P}^O to the template points \mathbf{p}_i^* , $i = s, \dots, M$), to monitor and quantify the object's packing.

A. Offline Reference Template

The offline reference template is needed to perform the packing task, as it provides the final target shape of the object and replaces the occluded parts with its offline 3-D points. To optimize packing space, the target shape for the long LEO is designed in the form of a modified spiral, which is composed of straight segments and concentric semicircles, as shown in Fig. 4. This target configuration is separated into periodic and aperiodic parts. The former consists of a semicircle followed by a straight segment. The latter only represents the beginning straight segment of the curve. Given a box-object pair, the maximum number of action planner loops (which equals to the number of grasps needed to complete the task) is one more than the total number of semicircles, i.e., $\lfloor \frac{w_B}{d_O} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the rounded down nearest integer operator.

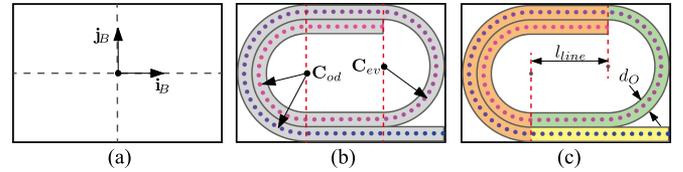


Fig. 4. Target shape Spiral of a LEO in a box. (a) shows the box's bottom. The points in gradients from pink to blue represents \mathbf{P} . (b) illustrates how Spiral is constructed with straight segments (between red dash lines) and two sets of concentric semicircles (the centers are C_{od} and C_{ev}). (c) illustrates the beginning segment (yellow) and periodic parts (orange and green) in Spiral.

Algorithm 1. The Description of the Shape Spiral.

Input: the box $\mathcal{B}(l_B, w_B, h_B)$, the object $\mathcal{O}(\eta, l_O, d_O)$
Output: the parameterized formula $\mathbf{P}^*(\lambda)$

- 1 $C_{od}(-\frac{l_B}{2} + \frac{w_B}{2}, 0, \frac{d_O}{2})$, $C_{ev}(\frac{l_B}{2} - \frac{w_B}{2} + \frac{d_O}{2}, \frac{d_O}{2}, \frac{d_O}{2})$;
- 2 $l_{line} = l_B - w_B + \frac{d_O}{2}$, $l_{count} = l_B - \frac{w_B}{2}$, $j = 0$, $\lambda = 0$;
- 3 **while** $\lambda < 1$ **do**
- 4 **if** $0 \leq \lambda l_O < l_B - \frac{w_B}{2}$ **then**
- 5 $\mathbf{P}(\lambda) = (\frac{l_B}{2}, -\frac{w_B + d_O}{2}, \frac{d_O}{2}) + (\lambda l_O, 0, 0)$;
- 6 **else**
- 7 $j = j + 1$ $\lambda = (l_B - \frac{w_B}{2})/l_O$;
- 8 $r_{sc} = \frac{w_B}{2} - \frac{d_O}{2}j$, $l_{semicircle} = \pi r_{sc}$;
- 9 **if** $\lambda l_O - l_{count} < l_{semicircle}$ **then**
- 10 $\phi = \frac{\lambda l_O - l_{count}}{r_{sc}}$;
- 11 $\mathbf{P}(\lambda) = C_{od} - r_{sc}(\sin \phi, \cos \phi, 0)$, j is odd;
- 12 $\mathbf{P}(\lambda) = C_{ev} + r_{sc}(\sin \phi, \cos \phi, 0)$, j is even;
- 13 $\lambda = \lambda + \frac{1}{M}$;
- 14 $l_{count} = l_{count} + l_{semicircle}$;
- 15 **if** $\lambda l_O - l_{count} \leq l_{line}$ **then**
- 16 $\mathbf{P}(\lambda) = C_{od} + (\lambda l_O - l_{count}, r_{sc})$, j is odd;
- 17 $\mathbf{P}(\lambda) = C_{ev} - (\lambda l_O - l_{count}, r_{sc})$, j is even;
- 18 $\lambda = \lambda + \frac{1}{M}$;
- 19 $l_{count} = l_{count} + l_{line}$;

We can compute the maximum object length that can be placed in the box (box capacity) with the spiral shape as follows:

$$l_O^{\max} = l_B - \frac{w_B}{2} + \sum_{j=1}^{\lfloor \frac{w_B}{d_O} \rfloor} \left(l_B - w_B + \frac{d_O}{2} + \pi \frac{w_B - d_O j}{2} \right). \quad (2)$$

We parameterize the centerline of the spiral shape [see Fig. 4(b)] with a normalized length $\lambda = i/M \in [0, 1]$, for $i = 1, \dots, M$. The parameterized centerline is denoted as $\mathbf{P}^*(\lambda)$, and the length of its curve is computed as follows:

$$l(\lambda) = \lambda l_O = \lambda \sum_{i=2}^M \|\mathbf{p}_i^* - \mathbf{p}_{i-1}^*\|_2. \quad (3)$$

The process for generating the target spiral shape is presented in Algorithm 1.

B. Online 3-D Vision

To compute the ordered skeleton \mathbf{P}^O , the point cloud processing algorithm extracts geometric information of the objects in real time. First, it smoothes the raw point clouds with a weighted filter [36] and downsamples it to optimize its computational cost. Next, it detects the boundaries of the object from

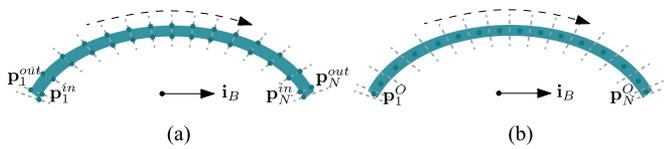


Fig. 5. Point cloud processing. (a) Boundary extraction. (b) Ordered skeleton.

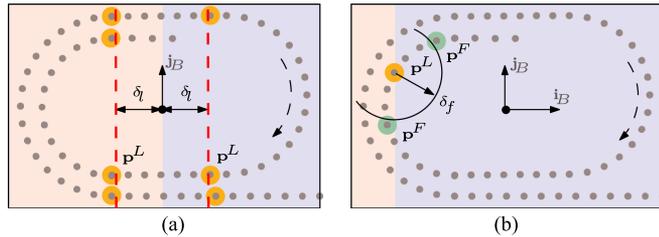


Fig. 6. Reference point generator. The offline reference template starts from a corner of the box and the point index clockwise increases. (a) Candidate placing points (orange). (b) Candidate fixing point (green) given a placing point (orange). Red and blue regions, respectively, indicate the workspace of the left and right-hand arms.

the point cloud. For that, we introduce a polar coordinate system with the origin at center of the box and its axis define along \mathbf{i}_B (as depicted in Fig. 1). Then, we segment the object into N sections by rotating (clockwise) a ray starting from \mathbf{i}_B around the center, and with a fixed angle interval (see Fig. 5). Along each ray, we search for nearest and farthest points in the raw point cloud feedback, which we denote as \mathbf{p}_i^{in} and $\mathbf{p}_i^{\text{out}}$, respectively; these points will serve as the components of the inner boundary and outer boundary of the object. Finally, the LEO's ordered skeleton \mathbf{P}^O is constructed by computing the mean of the raw feedback points between two adjacent rays. The length l_O and width d_O of the linear object are intuitively calculated as follows:

$$l_O = \sum_{i=2}^N \|\mathbf{p}_i^O - \mathbf{p}_{i-1}^O\|_2, \quad d_O = \frac{1}{N} \sum_{i=1}^N \|\mathbf{p}_i^{\text{out}} - \mathbf{p}_i^{\text{in}}\|_2. \quad (4)$$

IV. AUTOMATIC PACKING METHOD

A. Reference Points Generator

Our proposed manipulation method recycles a periodic action planner loop, which is composed of various high-level behaviors. To execute these behaviors, three types of points are planned, namely, the grasping reference \mathbf{p}^G , placing reference \mathbf{p}^L , and the fixing reference \mathbf{p}^F . The reference point generator (the orange block in Fig. 3) constructs the reference pose $\mathbf{x}^* = [x^*, y^*, z^*, \theta^*]^T$ for the robot based on the object body frame $\{F_O\}$ (which is computed from \mathbf{P}^O for \mathbf{p}^G) and the reference template frame $\{F\}$ (which is computed from \mathbf{P}^* for \mathbf{p}^L and \mathbf{p}^F). Fig. 2 conceptually depicts these frames.

The points \mathbf{p}^L represent the positions within the box where the LEO is to be placed by robot. These points are a subset of the offline reference template \mathbf{P}^* and are defined as $\mathbf{p}^L = \mathbf{p}_k^*$, for $k = 1, \dots, M$ as the index for the points in the template. The index k is chosen such that \mathbf{p}_k^* is approximately at a distance δ_l from the axis \mathbf{j}_B [i.e., along the red dash lines shown in Fig. 6(a)].

Depending on which robot arm plays the active packing role, \mathbf{p}_k^* is automatically chosen from the left or right to \mathbf{j}_B . The points \mathbf{p}^G indicate the positions on the object to be grasped by the robot. These points are the corresponding points to \mathbf{p}^L on the object's skeleton \mathbf{P}^O , and are computed as $\mathbf{p}^G = \hat{\mathbf{p}}_k$, for $k = 1, \dots, M$. The LEO's shaping behavior is achieved by driving \mathbf{p}^G into \mathbf{p}^L . The reference points generator contains some rules to online update the reference points (\mathbf{p}^G , \mathbf{p}^L , and \mathbf{p}^F) before every action planner loop starts. Note that the new \mathbf{p}^L can be the same as the ones in the past loops and not deleted from candidate list once it was used. Therefore, our method is robust to a type of pop-up situation when the object is still within the observable workspace.

In contrast with inelastic linear deformable objects, such as ropes or cords [31], LEOs have an intrinsic elastic energy that restores its shape to the original configuration. Thus, to steadily place the object in the box requires the assistant robot arm to fix the deformed object at a point \mathbf{p}^F while the active arm moves to a new grasping point \mathbf{p}^G . To compute \mathbf{p}^F , our method first obtains two candidate points along the object that are approximately at a distance δ_f from the placing point \mathbf{p}^L . The fixing point is selected at the same side of the assistant robot arm with respect to the active robot, see Fig. 6(b).

B. Action Primitives

As modeled in Section II-B, our method adopts two types of action primitives (for grippers and end-effectors) to compose high-level manipulation behaviors. The collection of action primitives for the 1-DOF grippers is as follows:

$$G = \{\text{Open}, \text{Close}\} = \{g_1, g_2\} \quad (5)$$

where the flags $g_1 = 1$ and $g_2 = 0$ define the opening and closing action of grippers, respectively.

The collection of five action primitives for the robotic end-effectors is as follows:

$$A = \{\text{Hover}, \text{Approach}, \text{Fix}, \text{Leave}, \text{Reset}\} \\ = \{a_1, a_2, a_3, a_4, a_5\}. \quad (6)$$

These end-effector action primitives are defined as follows.

- a_1 : *Hover*: The robot moves and stops above the reference point by an offset Δh . With this action, the robot is commanded with an end-effector target pose $\mathbf{u} = [x^*, y^*, z^* + \Delta h, \theta^*]^T$. This action is needed to avoid collisions with the object, and is done as a preparation step to perform fix and grasp actions.
- a_2 : *Approach*: The robot descends to z^* (viz. the height of the object's centerline). With this action, the robot is commanded with an end-effector target pose $\mathbf{u} = [x, y, z^*, \theta]^T$. This action, in combination with *Hover*, is needed to grasp and/or place the object by changing the gripper's configuration.
- a_3 : *Fix*: The robot descends to the object's surface, whose height is denoted by $z^* + \Delta f$. With this action, the robot is commanded as $\mathbf{u} = [x, y, z^* + \Delta f, \theta]^T$. This motion is needed to push the deformed elastic object and keep it inside the box, thus preventing it from returning to its original shape.

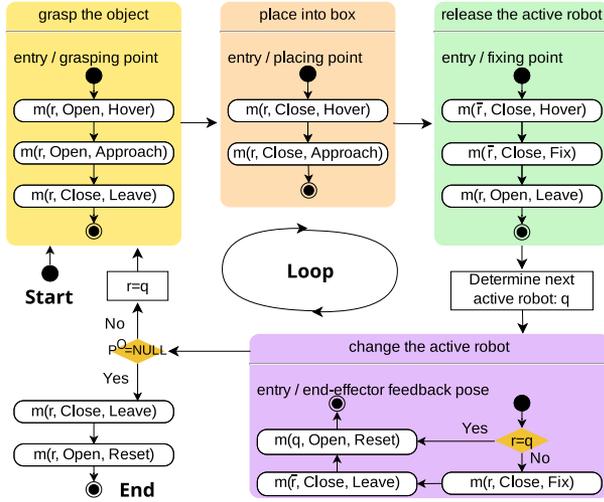


Fig. 7. Action planner for packing task. The arrow between states means $T(s) = 1$ while the arrow, representing $T(s) = 0$ pointing to the current state itself, is ignored to simplify the representation. High-level behaviors consist of robotic movements, e.g., grasping the object, placing it into the box, releasing the active robot after placing the object, and changing the identifier of the active robot. The results of reference point generator are inputs of behaviors as references. The specific action primitives determine the final target poses.

- a_4 : *Leave*: The robot returns to its initial height. With this action, the robot is commanded with an end-effector target pose $\mathbf{u} = [x, y, z(t_0), \theta]^T$. This action is needed to provide the robot with an obstacle-free region above the box's packing workspace.
- a_5 : *Reset*: The robot returns to its initial pose $\mathbf{u} = \mathbf{x}(t_0)$. This action is needed to visually observe the object with the top-view camera.

C. State Machine

The proposed state machine to automatically pack the long LEO has one periodic action planner loop (depicted in Fig. 7), which is iterated while monitoring the object's state until the task is completed. This sequence of actions is performed by collaborative robotic arms, identified as *Left* and *Right* (see Fig. 1) that can alternate between an active role and an assistant role. The former is in charge of grasping and placing the object into the box. The latter is in charge of immobilizing it while the arms change roles.

Our method uses a collection of robot roles $R = \{r, \bar{r}\}$, where r specifies which robot takes up the active packing role in a given cycle of the action planner loop. The identifier $r = \text{Left/Right}$ is automatically determined based on the proximity of \mathbf{P}^O to either the *Left* or *Right* robot. The assistant arm at the same cycle is denoted as \bar{r} , which for our dual-arm configuration, it simply represents the opposite arm, e.g., for $r = \text{Right}$, $\bar{r} = \text{Left}$.

The proposed state machine in Fig. 7 is composed of two layers. The first layer contains four high-level behaviors, namely, grasp the object, place it into the box, release the active robot, and change the active robot. The inputs of these high-level behaviors are the reference points \mathbf{p}^G , \mathbf{p}^L , and \mathbf{p}^F , and the reference poses \mathbf{x}^* . The second layer contains several low-level robot movements. These are modeled as elements in the collection

TABLE I
PROPERTIES OF THE OBJECTS IN EXPERIMENTS

Object	Length (%)	Width/Diameter (%)
$\mathcal{O}(\text{PEF}, 558, 38)$	98.08 ± 1.68	93.16 ± 6.32
$\mathcal{O}(\text{PEF}, 600, 38)$	97.73 ± 1.23	93.95 ± 4.74
$\mathcal{O}(\text{PEF}, 830, 38)$	98.41 ± 1.00	91.05 ± 4.47
$\mathcal{O}(\text{PEF}, 972, 38)$	98.80 ± 0.81	91.58 ± 5.26
$\mathcal{O}(\text{PUF}, 558, 30)$	97.83 ± 1.67	91.34 ± 7.33
$\mathcal{O}(\text{PUF}, 600, 30)$	97.77 ± 1.60	96.11 ± 6.07
$\mathcal{O}(\text{PUF}, 830, 30)$	98.46 ± 1.22	98.08 ± 5.02
$\mathcal{O}(\text{PUF}, 972, 30)$	98.91 ± 0.83	98.67 ± 6.10
$\mathcal{O}(\text{SCF}, 558, 34)$	97.80 ± 1.16	93.82 ± 5.59
$\mathcal{O}(\text{SCF}, 600, 34)$	98.18 ± 1.23	88.82 ± 7.35
$\mathcal{O}(\text{SCF}, 830, 34)$	98.83 ± 0.99	87.65 ± 7.65
$\mathcal{O}(\text{SCF}, 972, 34)$	98.89 ± 0.85	92.64 ± 6.47
$\mathcal{O}(\text{NL}, 600, 98)$	99.27 ± 1.90	96.22 ± 3.16

of states

$$S = \{s : s = m(R, G, A)\} \quad (7)$$

where the triple $m(R, G, A)$ defines the robot movements as a sequence of the following two commands: 1) first, the active/assistant robot $r/\bar{r} \in R$ performs the gripper action primitive $g_i \in G$, and 2) then, the robot performs the end-effector action primitive $a_j \in A$.

The result of the robot movement $m(R, G, A)$ corresponding to each state s is evaluated with the transition function

$$T(s) = \begin{cases} 1, & \text{once the robot completes the movement} \\ 0; & \text{otherwise.} \end{cases} \quad (8)$$

The proposed action planner stops when no object points are detected outside the box. The packing task succeeds when the shape difference e converges to e^* .

V. RESULTS

A. Experimental Setup

We conduct an experimental study to validate the proposed method. The developed experimental platform is composed of two 6-DOF robot manipulators (UR3) equipped with active grippers (Robotiq) that drive customized object grasping fixtures, and a top-view LiDAR camera (Intel RealSense L515) that captures real-time point clouds of the workspace. A table is placed between the two robot arms, with the packing box is rigidly attached to its surface. The robotic arms are controlled with a Linux-based PC (running Ubuntu 16.04), with ROS and RViz used for communication and visualization. Image processing is performed with the OpenCV libraries.

To test the robustness of our method for packing LEOs, we use 13 objects with different elastic properties, cross-sectional shapes, and object lengths. The density and Young's modulus of the object materials are listed in Table I. The cross sections, widths, and diameters of the objects are shown in Fig. 8. The 13 objects and its lengths are shown in Fig. 9, and 12 of these LEO are made of three materials: polyethylene foam (PEF), polyurethane foam (PUF), and silicone foam (SCF). These objects have four lengths: 558, 600, 830, and 972 mm. The thirteenth object is a pillow made of natural latex (NL) with a length of 600 mm. The objects in this study are all packed into boxes

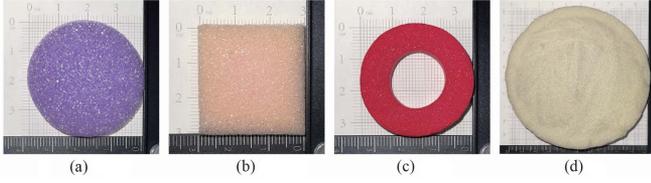


Fig. 8. Widths/diameters d_O and cross sections of objects made of different materials. The measurements of d_O are 38.0, 30.0, 34.0, and 98.0 mm. The cross sections are circle, square, ring, and circle.



Fig. 9. Packing 13 objects $\mathcal{O}(\eta, l_O, d_O)$ of different lengths into two boxes. PEF—polyethylene foam, PUF—polyurethane foam, SCF—silicone foam, and NL—natural latex.

of two different sizes, viz. $\mathcal{B}(270, 207, 80)$ and $\mathcal{B}(314, 232, 80)$ (given in mm units).

B. Vision-Based Computation of the Objects' Geometry

We validate the accuracy of the model (4) describing the objects' geometry (i.e., the length l_O and width d_O) by collecting ten measurements of their initial configuration over the table (similar to the one depicted in Fig. 1) and comparing the calculated dimensions with the ground truth, see Table II. The length l_O is computed from the ordered point cloud \mathbf{P}^O , whereas d_O is computed from the raw point cloud. The results in Table II show that the estimated object length l_O and width d_O are slightly smaller than the ground truth, which is caused

TABLE II
ACCURACY OF GEOMETRIC PROPERTY ESTIMATION

Object	Mean μ	Variance σ^2	3- σ Confidence Interval
$\mathcal{O}(\text{PEF}, 558, 38)$	17.5	0.299	99.79
$\mathcal{O}(\text{PEF}, 600, 38)$	18.0	0.411	99.84
$\mathcal{O}(\text{PEF}, 830, 38)$	18.1	1.087	99.87
$\mathcal{O}(\text{PEF}, 972, 38)$	18.5	0.874	99.42
$\mathcal{O}(\text{PUF}, 558, 30)$	12.7	0.540	97.48
$\mathcal{O}(\text{PUF}, 600, 30)$	15.1	0.078	98.61
$\mathcal{O}(\text{PUF}, 830, 30)$	16.0	0.300	97.85
$\mathcal{O}(\text{PUF}, 972, 30)$	15.1	0.360	98.03
$\mathcal{O}(\text{SCF}, 558, 34)$	19.2	2.049	99.64
$\mathcal{O}(\text{SCF}, 600, 34)$	19.6	0.130	99.73
$\mathcal{O}(\text{SCF}, 830, 34)$	17.0	1.120	98.89
$\mathcal{O}(\text{SCF}, 972, 34)$	20.3	1.435	99.85
$\mathcal{O}(\text{NL}, 600, 98)$	50.9	1.984	99.94

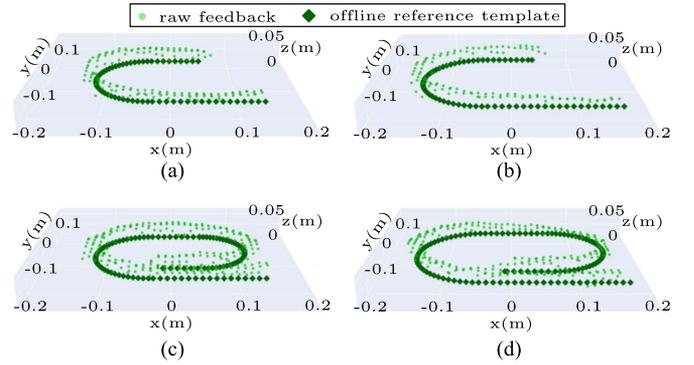


Fig. 10. Comparison of raw feedback and the offline reference template.

by the discretization of the continuous objects' arc-length and the partial view of their surface. This, however, does not affect the proposed manipulation strategy, as demonstrated in the experimental results that follow.

C. Similarity Analysis of the Reference Template

We set the distance of two adjacent points on the offline template \mathbf{P}^* as 5 mm, and the number of points in the offline reference template is $M = l_O/5$ mm. To verify if the designed Spiral shape is able to match the desired object in boxes, we compute the similarity between the point clouds of the reference template and the raw feedback of the object. To this end, we compute the set of minimum Euclidean distance between every point in \mathbf{P} to the points in \mathbf{P}^* as follows:

$$D = \{\min_j \|\mathbf{p}_i - \mathbf{p}_j^*\|_2 : \mathbf{p}_i \in \mathbf{P}, \mathbf{p}_j^* \in \mathbf{P}^*\}. \quad (9)$$

If the shape of the packed object matches Spiral well, D follows a Gaussian distribution $N(\mu, \sigma^2)$, with mean μ equal to the radius or half-width of the object $\mu \approx \frac{d_O}{2}$, and standard deviation $\sigma \approx 0$. The average value of the set D is equal to the error e_{in} . Fig. 10 presents the raw feedback point clouds \mathbf{P} and the offline reference template \mathbf{P}^* . A statistical analysis of the similarity is given in Table III, which demonstrates that the mean distances $\mu \approx \frac{d_O}{2}$ and the variances σ^2 are small.

TABLE III
PERFORMANCE OF THE METHOD IN PACKING TASKS

Algorithm 1. The Description of the Shape Spiral.

Input: the box $\mathcal{B}(l_B, w_B, h_B)$, the object $\mathcal{O}(\eta, l_O, d_O)$
Output: the parameterized formula $\mathbf{P}^*(\lambda)$

```

1  $\mathbf{C}_{od}(-\frac{l_B}{2} + \frac{w_B}{2}, 0, \frac{d_O}{2}), \mathbf{C}_{ev}(\frac{l_B}{2} - \frac{w_B}{2} + \frac{d_O}{2}, \frac{d_O}{2}, \frac{d_O}{2});$ 
2  $l_{line} = l_B - w_B + \frac{d_O}{2}, l_{count} = l_B - \frac{w_B}{2}, j = 0, \lambda = 0;$ 
3 while  $\lambda < 1$  do
4   if  $0 \leq \lambda l_O < l_B - \frac{w_B}{2}$  then
5      $\mathbf{P}(\lambda) = (\frac{l_B}{2}, -\frac{w_B}{2} + \frac{d_O}{2}, \frac{d_O}{2}) + (\lambda l_O, 0, 0);$ 
6   else
7      $j = j + 1 \quad \lambda = (l_B - \frac{w_B}{2})/l_O;$ 
8      $r_{sc} = \frac{w_B}{2} - \frac{d_O}{2}j, l_{semicircle} = \pi r_{sc};$ 
9     if  $\lambda l_O - l_{count} < l_{semicircle}$  then
10       $\phi = \frac{\lambda l_O - l_{count}}{r_{sc}};$ 
11       $\mathbf{P}(\lambda) = \mathbf{C}_{od} - r_{sc}(\sin \phi, \cos \phi, 0), j$  is odd;
12       $\mathbf{P}(\lambda) = \mathbf{C}_{ev} + r_{sc}(\sin \phi, \cos \phi, 0), j$  is even;
13       $\lambda = \lambda + \frac{1}{M};$ 
14       $l_{count} = l_{count} + l_{semicircle};$ 
15     if  $\lambda l_O - l_{count} \leq l_{line}$  then
16        $\mathbf{P}(\lambda) = \mathbf{C}_{od} + (\lambda l_O - l_{count}, r_{sc}), j$  is odd;
17        $\mathbf{P}(\lambda) = \mathbf{C}_{ev} - (\lambda l_O - l_{count}, r_{sc}), j$  is even;
18        $\lambda = \lambda + \frac{1}{M};$ 
19      $l_{count} = l_{count} + l_{line};$ 

```

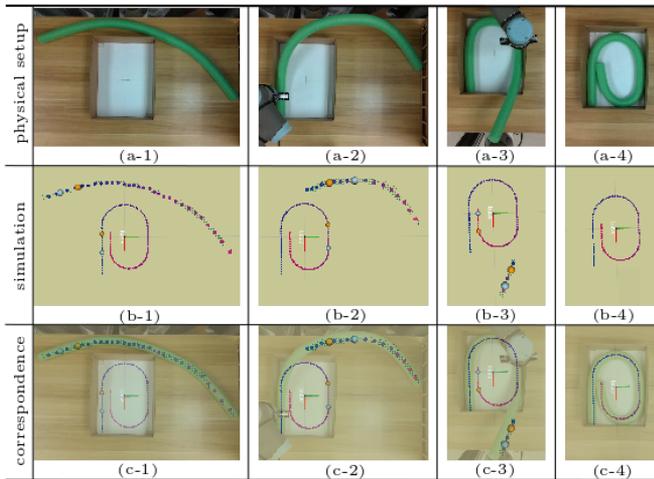


Fig. 11. Reference point generator of action planner. The light green points are raw feedback points \mathbf{P} . Blue indicates the start, red indicates the end, and the gradients of colors indicate the order of points. The grasping points and placing points are orange, and the fixing points are light blue. Blue indicates the start and red indicates the end, and the gradients of colors indicate the order of points.

D. Generation of Reference Points

In this section, we take $\mathcal{O}(\text{PEF}, 972, 38)$ as an example. The constant distance from \mathbf{p}^L to \mathbf{j}_B is set as $\delta_l = 50$ mm, and the distance from \mathbf{p}^F to \mathbf{p}^L as $\delta_f = 100$ mm. The selection of both parameters is flexible as long as they can protect robots from collision. In our research, all experiments use the same setting to generate the reference points. Based on these parameters, the generator automatically computes the placing points $\mathbf{p}^L = \mathbf{p}_k^*$ and grasping points $\mathbf{p}^G = \hat{\mathbf{p}}_k$, for and index $k = \{26, 64, 110\}$. The fixing points \mathbf{p}^F are determined within each cycle based on the location of the robot relative to each other. Fig. 11 depicts

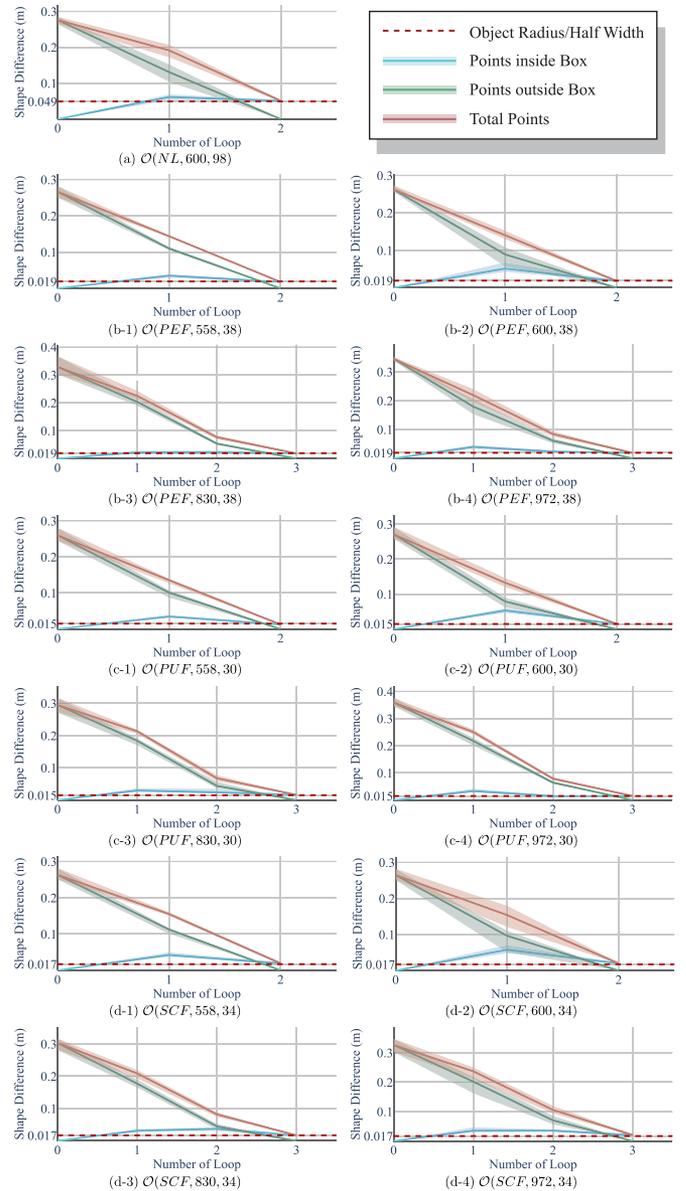


Fig. 12. Shape differences (inside box, outside box, and total) during manipulation.

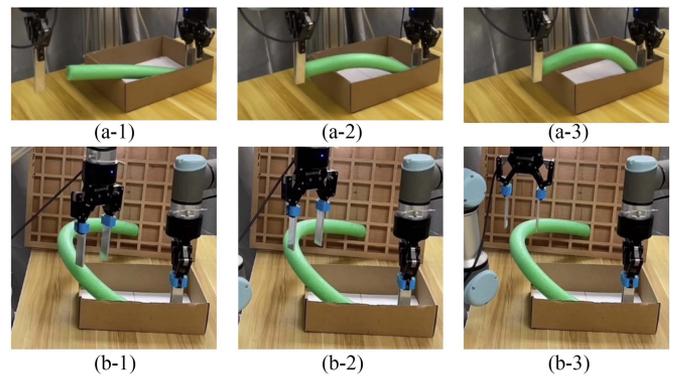


Fig. 13. Robotic movements with two modes of *Hover*. In (a-1)–(a-3), the robot moves at the constant height so that it touches the object causing the failure of grasping. In (b-1)–(b-3), the robot smoothly moves along the object with the following algorithm.

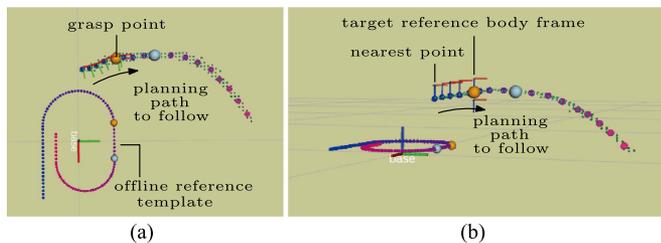


Fig. 14. Path planning for robots following the object without collision. (a) Top view. (b) Side view.

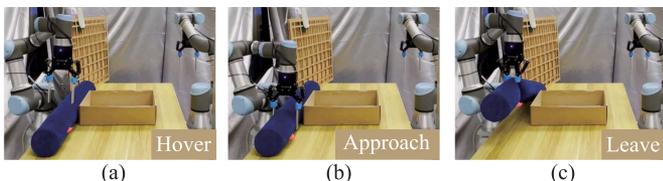


Fig. 15. Grasping manipulation (Loop 1) is composed of action primitives given a grasping point \mathbf{p}_1^G : (a) $m(\text{Left}, \text{Open}, \text{Hover})$, (b) $m(\text{Left}, \text{Open}, \text{Approach})$, (c) $m(\text{Left}, \text{Close}, \text{Leave})$.

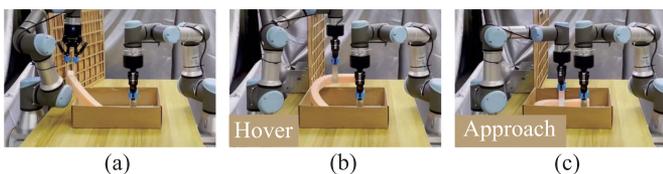


Fig. 16. Placing manipulation (Loop 2) is composed of action primitives given a grasping point \mathbf{p}_2^L : (a) initial state, (b) $m(\text{Left}, \text{Close}, \text{Hover})$, and (c) $m(\text{Left}, \text{Close}, \text{Approach})$.

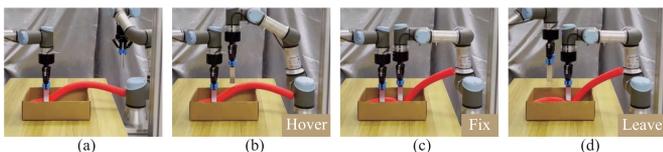


Fig. 17. Assistant robot help to release the active robot that is grasping the object after placing it in the box (Loop 2) given a fixing point \mathbf{p}_2^F : (a) initial state, (b) $m(\text{Right}, \text{Close}, \text{Hover})$, (c) $m(\text{Right}, \text{Close}, \text{Approach})$, and (d) $m(\text{Left}, \text{Open}, \text{Leave})$.

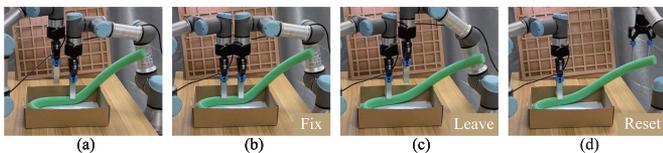


Fig. 18. Changing hand manipulation (Loop 2) is composed of action primitives conducted by both robots concerning their current positions: (a) initial state, (b) $m(\text{Left}, \text{Close}, \text{Fix})$, (c) $m(\text{Right}, \text{Close}, \text{Leave})$, and (d) $m(\text{Right}, \text{Open}, \text{Reset})$.

the reference template and the ordered skeleton, where \mathbf{p}^G and \mathbf{p}^L are represented by orange points, whereas \mathbf{p}^F by light-blue points. In this figure, we can see how the active robot grasps the object at \mathbf{p}^G (orange points on the object) and placed it at \mathbf{p}^L (corresponding orange points at the template). The figure also shows how the assistant robot fixes the object by pushing it at

\mathbf{p}^F (corresponding light-blue point at the reference template), which enables the active robot to be released and conduct the next action.

E. Shape Difference

This section validates the performance of the proposed automatic packing method with the 13 different objects shown in Fig. 9. Each experiment is conducted 10 times.¹ To quantify the progress and accuracy of the packing task, we compute the shape difference with visual feedback. This metric is only computed at the beginning of every loop, as there are occlusions and noisy points when the robots are moving. The blue, green, and red solid curves shown in Fig. 12, respectively, represent the errors e_{in} , e_{out} , and e that are obtained from ten automatic packing experiments. The red dashed line represents the errors' ideal value $e^* = \frac{d_o}{2}$. The blue curves start from zero when the object is completely outside of the box before the automatic manipulation, and converge to e^* after the object has been fully packed. The green curves start from large initial values when the objects lie on the table without deformation, and converge to zero when there are no points outside of the box after packing has been completed. The red curves, which represent the weighted average of the blue and green curves, start from the same initial values as the green curves, and monotonically decrease to e^* . These results quantitatively demonstrate that the proposed manipulation strategy can successfully deform and manipulate various types of LEOs into compact boxes.

F. High-Level Behaviors Constructed With Action Primitives

We set parameters for action primitives as $\Delta h = 100$ mm depending on the height at which the robot stop above the object, and $\Delta f = 0.7 \times \frac{d_o}{2}$ depending on the strength with which the robot pushes the object. Given appropriate parameters, our method is robust enough to get a 100% success rate. Generally, it is sufficient for the robots to move at a constant height to avoid collisions with the edge of the cuboid-shaped bin. However, as the ordered skeleton \mathbf{P}^O is higher than the box, this constant height may produce collisions with the object, as shown in Fig. 13 (a-1)–(a-3). To deal with this problem, we proposed the *Hover* action primitive, which guides the robot to move to the nearest point above the object, then, to move along the object's curvature until the gripper reaches the target grasping point \mathbf{p}^G , see Fig. 14. This valuable of action primitive enables to successfully grasp LEOs with complex bent geometries, while avoiding collisions with them, as demonstrated in Fig. 13 (b-1)–(b-3).

The four high-level behaviors (i.e., grasp the object, place it into the box, release the active robot, and change the active robot) are shown in Figs. 15–18. The experiment in Fig. 15 shows the how the robot autonomously grasps the object and reaches a safe height from the table. This figure shows that the *Left* robot performs *Hover* above \mathbf{p}^G , then *Approach* toward the grasping point \mathbf{p}^G with an open gripper, and finally, *Close*

¹[Online]. Available: <https://youtu.be/ZGJcRE2nqBc>

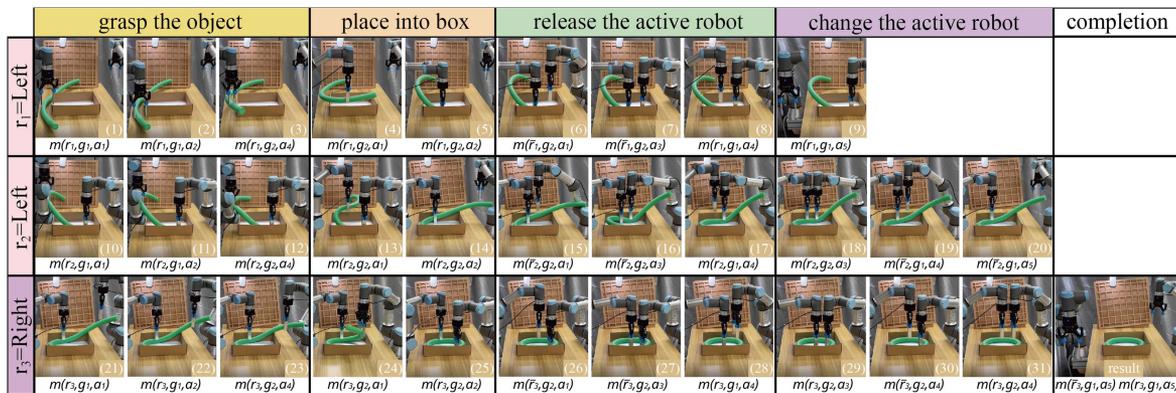


Fig. 19. Experiment process and action primitives of packing \mathcal{O} (PEF, 972, 38). The rows represent three action planner loops. The columns represent the behaviors of robots. The thumbnails demonstrate robotic movements. The first two loops are mainly executed by the left arm (grasping and placing) and assisted by the right arm (fixing). The third loop is mainly executed by the right arm (grasping and placing) and assisted by the left arm (fixing).

and *Leave* from \mathbf{p}^G toward the initial height $z(t_0)$. The second high-level behavior is depicted in Fig. 16. The purpose of these sequence of actions is to deform and place the grasped object at a specific position within the box. The figure shows the initial configuration where the object is grasped by the *Left* robot and the inside-the-box part is fixed by the *Right* robot; Then, the *Left* robot performs *Hover* and *Approach* toward \mathbf{p}^L , while holding the object.

The third high-level behavior (release the active robot) is depicted in Fig. 17. The purpose of these movements is to fix the object's shape while the active robot that is holding the object opens its gripper. The figure shows the initial configuration where the object (already inside the box) is grasped by the *Left* robot. Then, the *Right* robot performs *Hover* and *Approach* \mathbf{p}^F with a closed gripper. The *Left* robot *Open* its gripper, *Leave* the object, and returns to the initial height $z(t_0)$. The fourth high-level behavior (change the active robot) is depicted in Fig. 18. The purpose of these movements is to switch the active robot's identifier from *Left* to *Right*, this, in preparation for the *Right* robot to conduct the next grasp task. The figure shows the initial state where the *Left* robot is free, and the right robot is performing *Fix* onto the object. Then, the *Left* robot performs the *Fix* action while the *Right* robot performs *Leave* and then *Reset* to return to its initial position, which completes one cycle of the action planner loop.

We take \mathcal{O} (PEF, 972, 38) as a representative example to demonstrate the performance of the method. Fig. 19 depicts the complete process of the packing task, which consists of three cycles, corresponding to the three rows. Each thumbnail in the figure presents a movement $m(R, G, A)$ conducted by the robot arms. The periodic nature of the action planner is illustrated by the fact that the three cycles share the same first three high-level behaviors, viz. grasp the object, place into the box, and release the active robot. The first and second cycles only differ in the fourth high-level behavior, i.e., change the active robot, as the active robots in these two cycles are the same (thus, there is no need to change the active robot). The packing process ends in the third cycle, where the object has been completely packed into the box.

VI. CONCLUSION

In this article, we propose a complete method to pack long LEOs into compact boxes. First, we design a hybrid geometric model including an online 3D-vision method and an offline reference template to tackle occlusions during packing manipulations under a single-view camera. Online 3D-vision method extracts objects' geometric information in real time. Offline reference template is generated based on a designed shape *Spiral*. The effectiveness of *Spiral* is proved by the high similarity between the offline reference template and the shape of the packed object. Then, we propose a method to preliminary plan reference points for grasping, placing, and fixing. Next, we propose an action planner to compose defined action primitives as high-level behaviors and achieve packing tasks by repeating a periodic action planner loop. Finally, extensive experiments are conducted to verify the generality of our proposed method for various objects with different elastic materials, lengths, densities, and cross sections.

Although the method is designed for packing tasks, the defined action primitives and the reference point generator method can be used in other manipulation tasks (e.g., object sorting, multiple objects assemblies, etc). Also, the proposed perception method is able to work without markers and decrease computation time by extracting minimum geometric information of objects. A limitation of our method is that our perception method does not consider the situations where the object is outside the camera's view range. A possible solution is to employ multiview visual system to perceive the object.

For future work, we plan to explore the multiview vision and to extend the framework to other comprehensive tasks involving more types of objects (e.g., rigid, elastic, and articulated), as well as to optimize the packing to save space. Our team is currently working along this challenging direction.

REFERENCES

- [1] UN, "COVID-19 and e-commerce: A global review," New York, NY, USA, United Nations, White Paper, Mar. 2021.

- [2] S. L. Freichel, J. Wollenburg, and J. K. Wörtge, "The role of packaging in omni-channel fashion retail supply chains-how can packaging contribute to logistics efficiency?," *Logist. Res.*, vol. 13, no. 1, 2020.
- [3] J. Zhu et al., "Challenges and outlook in robotic manipulation of deformable objects," *IEEE Robot. Autom. Mag.*, vol. 29, no. 3, pp. 67–77, Sep. 2022.
- [4] X. Li, X. Su, and Y.-H. Liu, "Vision-based robotic manipulation of flexible PCBs," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 6, pp. 2739–2749, Dec. 2018.
- [5] V. Digani, L. Sabattini, C. Secchi, and C. Fantuzzi, "Ensemble coordination approach in multi-AGV systems applied to industrial warehouses," *IEEE Trans. Automat. Sci. Eng.*, vol. 12, no. 3, pp. 922–934, Jul. 2015.
- [6] M. Schwarz et al., "Fast object learning and dual-arm coordination for cluttered stowing, picking, and packing," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 3347–3354.
- [7] A. Yasuda, G. A. G. Ricardez, J. Takamatsu, and T. Ogasawara, "Packing planning and execution considering arrangement rules," in *Proc. IEEE Int. Conf. Robot. Comput.*, 2020, pp. 100–106.
- [8] A. Zeng et al., "Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2017, pp. 1386–1383.
- [9] D. Seita et al., "Learning to rearrange deformable cables, fabrics, and bags with goal-conditioned transporter networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 4568–4575.
- [10] C. E. Etumnu and N. O. Widmar, "Grocery shopping in the digital era," *Choices Mag.*, vol. 35, no. Quarter 2, pp. 1–8, 2020.
- [11] D. Navarro-Alarcon, Y. Liu, J. Romero, and P. Li, "On the visual deformation servoing of compliant objects: Uncalibrated control methods and experiments," *Int. J. Rob. Res.*, vol. 33, no. 11, pp. 1462–1480, 2014.
- [12] J. Felip and A. Morales, "Robust sensor-based grasp primitive for a three-finger robot hand," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2009, pp. 1811–1816.
- [13] G. Ficht et al., "NimbRo-OP2X: Adult-sized open-source 3D printed humanoid robot," in *Proc. IEEE-RAS 18th Int. Conf. Humanoid Robots*, 2018, pp. 1–9.
- [14] W. Wang, R. Li, Z. M. Diekel, and Y. Jia, "Robot action planning by online optimization in human-robot collaborative tasks," *Int. J. Intell. Robot. Appl.*, vol. 2, no. 2, pp. 161–179, 2018.
- [15] M. Schwarz et al., "NimbRo picking: Versatile part handling for warehouse automation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 3032–3039.
- [16] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2018, pp. 4238–4245.
- [17] A. Capitanelli, M. Maratea, F. Mastrogiovanni, and M. Vallati, "On the manipulation of articulated objects in human-robot cooperation scenarios," *Rob. Auton. Syst.*, vol. 109, pp. 139–155, 2018.
- [18] M. Kimura, Y. Sugiyama, S. Tomokuni, and S. Hirai, "Constructing rheologically deformable virtual objects," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2003, pp. 3737–3743.
- [19] N. Essahbi, B. C. Bouzgarrou, and G. Gogu, "Soft material modeling for robotic manipulation," in *Proc. Appl. Mech. Mater.*, 2012, vol. 162, pp. 184–193.
- [20] L. Han, H. Wang, Z. Liu, W. Chen, and X. Zhang, "Vision-based cutting control of deformable objects with surface tracking," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 4, pp. 2016–2026, Aug. 2021.
- [21] A. Petit, F. Ficuciello, G. A. Fontanelli, L. Villani, and B. Siciliano, "Using physical modeling and rgb-d registration for contact force sensing on deformable objects," in *Proc. Int. Conf. Informat. Control, Autom. Robot.*, 2017, pp. 4–33.
- [22] P. Kaufmann, S. Martin, M. Botsch, and M. Gross, "Flexible simulation of deformable models using discontinuous galerkin fem," *Graphical Models*, vol. 71, no. 4, pp. 153–167, 2009.
- [23] J. Qi et al., "Contour moments based manipulation of composite rigid-deformable objects with finite time model estimation and shape/position control," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 5, pp. 2985–2996, Oct. 2022.
- [24] Y. Wang, H. Wang, Z. Liu, and W. Chen, "Visual servo-collision avoidance hybrid task by considering detection and localization of contact for a soft manipulator," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 3, pp. 1310–1321, Jun. 2020.
- [25] F. Xu, H. Wang, J. Wang, K. W. S. Au, and W. Chen, "Underwater dynamic visual servoing for a soft robot arm with online distortion correction," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 3, pp. 979–989, Jun. 2019.
- [26] M. Laranjeira, C. Dune, and V. Hugel, "Catenary-based visual servoing for tether shape control between underwater vehicles," *Ocean Eng.*, vol. 200, 2020, Art. no. 107018.
- [27] D. Navarro-Alarcon and Y. Liu, "Fourier-based shape servoing: A new feedback method to actively deform soft objects into desired 2-D image contours," *IEEE Trans. Robot.*, vol. 34, no. 1, pp. 272–279, Feb. 2018.
- [28] Z. Hu, T. Han, P. Sun, J. Pan, and D. Manocha, "3-D deformable object manipulation using deep neural networks," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 4255–4261, Oct. 2019.
- [29] J. Zhu, D. Navarro-Alarcon, R. Passama, and A. Cherubini, "Vision-based manipulation of deformable and rigid objects using subspace projections of 2 d contours," *Rob. Auton. Syst.*, vol. 142, 2021, Art. no. 103798.
- [30] P. Zhou, J. Zhu, S. Huo, and D. Navarro-Alarcon, "LaSeSOM: A latent and semantic representation framework for soft object manipulation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 5381–5388, Jul. 2021.
- [31] T. Tang and M. Tomizuka, "Track deformable objects from point clouds with structure preserved registration," *Int. J. Robot. Res.*, vol. 41, 2018, Art. no. 0278364919841431.
- [32] C. W. M. T. S. Jin, W. Lian, and S. Schaal, "Robotic cable routing with spatial representation," *IEEE Robot. Autom.*, vol. 7, no. 2, pp. 5687–5694, Apr. 2022.
- [33] C. Chi and D. Berenson, "Occlusion-robust deformable object tracking without physics simulation," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2019, pp. 6443–6450.
- [34] C. F. R. C. D. Tian, H. Ochimizu, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3460–3464.
- [35] D. Hudson and C. D. Manning, "Learning by abstraction: The neural state machine," in *Proc. Adv. Neural Inf Process Syst.*, 2019.
- [36] T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions," *Technometrics*, vol. 37, no. 2, pp. 185–194, 1995.



Wanyu Ma (Graduate Student Member, IEEE) received the B.S. and M.A.Eng. degrees in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2016 and 2018, respectively. She is currently working toward the Ph.D. degree in mechanical engineering from The Hong Kong Polytechnic University, Hong Kong.

Her research interests include robotics, visual servoing, and deformable objects manipulation.



Bin Zhang received the B.S. degree in detection, guidance, and control technology from Beihang University, Beijing, China, in 2017, and the M.S. degree in control science and engineering from the China Academy of Space Technology, Beijing, China, in 2020. He is currently working toward the Ph.D. degree in mechanical engineering from The Hong Kong Polytechnic University, Hong Kong.

His research interests include multiagent systems and control theory.



Lijun Han received the B.S. degree in automation from Tianjin University, Tianjin, China, in 2019. She is currently working toward the Ph.D. degree in control technology and control engineering with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China.

Her research interests include visual servoing, robot control, and surgical robots.



Shengzeng Huo (Graduate Student Member, IEEE) received the B.S. degree in vehicle engineering from the South China University of Technology, Guangzhou, China, in 2019. He is currently working toward the Ph.D. degree in mechanical engineering from The Hong Kong Polytechnic University, Hong Kong.

His research interests include bimanual manipulation, deformable object manipulation, and robot learning.



David Navarro-Alarcon (Senior Member, IEEE) received the Ph.D. degree in mechanical and automation engineering from The Chinese University of Hong Kong, Hong Kong, in 2014.

From 2014 to 2017, he was a Postdoctoral Fellow and then a Research Assistant Professor with the CUHK T Stone Robotics Institute, Hong Kong. Since 2017, he has been with The Hong Kong Polytechnic University, Hong Kong, where he is currently an Assistant Professor with the Department of Mechanical Engineering, and

the Principal Investigator of the Robotics and Machine Intelligence Laboratory. His current research interests include perceptual robotics and control theory.



Hesheng Wang (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2002, and the M.Phil. and Ph.D. degrees in automation and computer-aided engineering from The Chinese University of Hong Kong, Hong Kong, in 2004 and 2007, respectively.

He is currently a Professor with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China. His current research

interests include visual servoing, service robot, computer vision, and autonomous driving.

Dr. Wang is an Associate Editor for *IEEE ROBOTICS AND AUTOMATION LETTERS* and *Assembly Automation and the International Journal of Humanoid Robotics*, and Technical Editor of the *IEEE/ASME TRANSACTIONS ON MECHATRONICS*. From 2015 to 2019, he was an Associate Editor for *IEEE TRANSACTIONS ON ROBOTICS*. He was the General Chair of the IEEE RCAR 2016 and the Program Chair of the IEEE ROBIO 2014 and IEEE/ASME AIM 2019.