2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 1-5, 2023. Detroit, USA

Vision-Based In-Hand Manipulation of Variously Shaped Objects via Contact Point Prediction

Yuzuka Isobe[1], Sunhwi Kang[2], Takeshi Shimamoto[2], Yoshinari Matsuyama[2], Sarthak Pathak[3] and Kazunori Umeda[3]

Abstract— **In-hand manipulation (IHM) is an important ability for robotic hands. This ability refers to changing the position and orientation of a grasped object without dropping it from the hand workspace. One major challenge of IHM is to achieve a large range of manipulation (especially rotation), regardless of the shape, size, and the orientation during manipulation of the grasped object. There are two main challenges - the manipulation range (due to the range of motion of the hand) and keeping the object grasped under all shapes and orientations. Specifically, even when the contact points between the hand and the object switch and the positions of these points change due to its shape and changing orientation, constant grasp of the object is required.**

This paper presents an IHM method for a robotic hand with belts, based on the prediction of the contact-point changes via image information. The focus is on a robotic hand that has a two-fingered parallel gripper with conveyor belts which can continuously manipulate an object through a large range. A stereo camera is attached to the hand. First, the contour of the grasped object is acquired from the camera. From the contour, the switching of the contact points between the surfaces of the belts and the object is predicted. Then, the positions of the contact points in the next frame are estimated by rotating the contour. The velocities of the belts are calculated based on the prediction of the switching. The fingers are controlled to follow the estimated positions of the contact points, via a feed-forward control. The effectiveness of the proposed method is verified through in-hand manipulation experiments for 22 objects of various shapes and sizes.

I. INTRODUCTION

In-hand manipulation (IHM) is an important ability for robotic hands. It refers to changing both the position and the orientation of a grasped object without dropping it [1], [2], [3]. The main application of this is a pick-and-place motion, which is a basic task for robotic hands used in various industries, such as manufacturing, logistics, and retail.

One major challenge is how to achieve both grasping and manipulation (translation and rotation) simultaneously [4]. The grasping capability for IHM represents the robustness to the object characteristics (size, shape, orientation, and so on) or the disturbances during manipulation. The manipulation capability shows how far objects can be translated or rotated, i.e., the range of displacement and the rotational angle. Focusing especially on the grasping capability, unique hands with units that are deformable to fit the shape of the object have been proposed [5], [6]. Since these hands can extend the contact region between the surface of the hands and the grasped object, a stable grasp can be achieved. Additionally, by using underactuated grippers, the control method to achieve stable grasp during manipulation has been proposed by a self-supervised learning [7]. However, due to the low degrees of freedom (DOF) of these hands, the controllable ranges of translation and rotation are limited. Therefore, so as to achieve high ranges of manipulation, it is necessary to once place the object on the extrinsic environment ([8], [9]) or *regrasp* repeatedly. Here, *regrasping* is known as a motion of releasing a finger away from part of the object surface and then contacting it at another part [1], [2]. These motions increase the time required to conduct the manipulation, which can be a challenge in actual tasks. Attempting to solve this problem, robotic hands with high DOF have been developed. Zhou *et al.* [10] designed a 13-DOF robotic hand composed of five fingers and a palm. It can grasp objects with various outlines and both translate and rotate them in large ranges. In addition, as the number of rigid fingers increases, the contact regions can be expanded. Andrychowicz *et al.* [11] designed a rigid five-fingered hand to achieve IHM with high capabilities without the placing motion. Although both the grasping and manipulation capabilities are implemented simultaneously using these approaches, complex control systems are required. Additionally, the regrasping behavior is required even with these configurations.

In contrast to multi-fingered hands, as an approach to improve the manipulation capability, robotic hands with rotatable elements have been proposed in many studies. Ichikura *et al.* [12] discussed a unique hand with a belt passed between parallel grippers and a novel motion inspired by diabolo juggling to accomplish IHM with variously shaped objects. This motion was required to invert the hand and vibrate the object. As a configuration with rotatable elements installed on each gripper, Tahara *et al.* [13] constructed a two-fingered robotic hand with rotatable fingertips. Since each fingertip rolls individually, the grasped object can be translated and rotated by rolling each of them in respective same and opposite directions. The roller-based hand shown by Yuan *et al.* [14] has also been equipped with rotatable fingertips. Further, a gripper with a conveyor-belt surface on each finger has been demonstrated [15]. This configuration

[1]Yuzuka Isobe is with the Graduate School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan isobe@sensor.mech.chuo-u.ac.jp

[2]Sunhwi Kang, Takeshi Shimamoto, and Yoshinari Matsuyama are with R&D Revision, Panasonic Connect Co., Ltd., 3-1-1 Yagumo-naka-machi, Moriguchi City, Osaka 570-8501, Japan {kang.sunhwi, shimamoto.takeshi, matsuyama.yoshinari}@jp.panasonic.com

[3]Sarthak Pathak and Kazunori Umeda are with the Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan {pathak, umeda}@mech.chuo-u.ac.jp

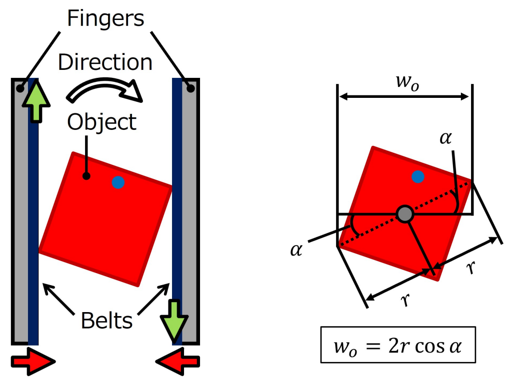
2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-9190-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IROS55552.2023.10341968

allows for large ranges of both the translation of a pinched object inside the hand and the rotation of it. Similar methods using the configuration of a gripper with belts have also been researched [16], [17], [18], [19]. By using such rotatable configurations, the rotation range of the grasped object can be increased according to the DOF of the belts. That is, if the belt can loop endlessly, the object can be rotated one or more arbitrary revolutions without explicit regrasping unless it is dropped. In order not to drop the object, these hands have also been configured with underactuated or soft fingers. Although the fingers attempt to passively adjust to the shape of the object, the object is rotated excessively for the fingers to follow the object. Specifically, when the contact points between the surfaces of the rotatable configurations and object switch during rotation or the positions of the points change largely, the fingers may not keep the contact. However, how much the object should be rotated to prevent dropping depends on the object's shape and orientation (that relate to the above changes in the contact points), and also the dynamic characteristics of the fingers. Therefore, adjusting the rotational angle based on the changes in the contact points and enabling the fingers to follow the object remain as challenges.

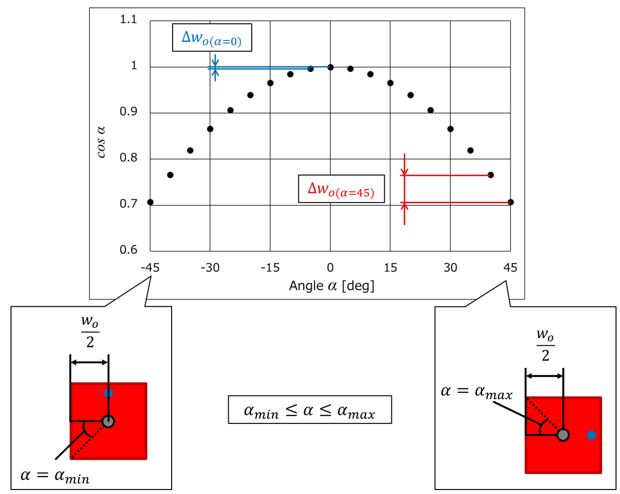
This paper presents an in-hand manipulation method based on the prediction of the changes in the contact points between a robotic hand and a grasped object. For the purpose of improving the rotatable range of the grasped object, a two-fingered parallel gripper with conveyor belts are used as a robotic hand. To increase the variety of manipulatable objects, the control methods for both the fingers and the belts are developed using an image from a stereo camera attached to the hand. From a camera image, the contour of the grasped object is acquired. Based on this contour, the switching of the contact point between each belt and the object is predicted. According to the prediction, the rotational angle of the object in the next frame so as to not drop it, is determined. To rotate the object to the determined angle, the velocities of the belts are calculated, supposing that no slippage occurs between the belts and the object. Then, since using only a feed-back control can cause dropping of the object due to primary delay, a feed-forward control of the fingers is adopted. Based on the contour of the object, it is estimated where the contact points are moving to in the next frame. The fingers are forward controlled to the estimated positions. The validity of this approach is verified through in-hand manipulation experiments for 22 samples with 11 shapes and 2 sizes.

The novelties and contribution of the proposed method are as follows:

- Improving both grasping and manipulation capabilities; extend both the rotational range and the variety of manipulatable objects, based on the use of the fingers with belts and the detection of the object's shape and orientation from a camera.
- Keeping the grasp during manipulation by both predicting the switching of the contact points between the belt and the object and estimating the positions of the contact points in the next frame.
- Continuous control and real-time performance without explicitly regrasping the object.



(a) The calculation of the width w_o when an object with a square cross section is rotated by fingers.



(b) The differences in the changes Δw_o in the width between angles depend on the orientation.

Fig. 1. An example of in-hand manipulation of an object with a square cross section and the changes in the required width of the fingers.

II. Challenging Points

In this section, one situation in which the fingers may drop the object is detailed as a main challenge on which this paper is focused. Fig. 1(a) shows a two-fingered parallel gripper with belts grasping an object with a square cross section and rotating it in the direction indicated in the figure. Here, the width w_o is the distance between the leftmost and rightmost points of the object and thus indicates the positions to which the fingers should move. As illustrated in the figure, the radius of rotation is indicated as r, and the angle between the horizontal axis and the line connecting each contact point and the rotational center is α . Additionally, w_o can be calculated by using r and $\cos\alpha$. It follows that, when the radius r or $\cos\alpha$ changes, the width w_o changes correspondingly. During manipulation of a nondeformable (rigid) object, the changes in radius are caused by the switching of the contact points.

Besides, $\cos\alpha$ varies between the lower and upper limits (α_{min} and α_{max}) from time to time, as long as the object is rotated. α_{min} and α_{max} are defined by the shape of the object.

Even though the angular velocity of the object is constant, the changes in $\cos\alpha$ per unit of time may not be constant. Fig. 1(b) shows the cosine value at each angle α. In this example, the angle α is between $\alpha_{min} = -45$ and $\alpha_{max} = 45$ [deg]. As the object rotates from -45 to 0 [deg], $\cos\alpha$ increases, and w_o becomes larger. The smallest changes Δw_o are indicated in blue in the graph in Fig. 1(b). Then, rotating from 0 to 45 [deg], $\cos\alpha$ decreases, and the width w_o becomes shorter. In contrast, the changes in $\cos\alpha$ between each angle increase, and Δw_o becomes larger in a negative direction. Then, when the angle is $\alpha_{max} = 45$ [deg], each contact point switches, and the angle α calculated by new contact point is $\alpha_{min} = -45$ [deg]. Consequently, Δw_o has the maximum value in a negative direction, as written in red in the figure, just before the angle becomes $\alpha_{max} = 45$ [deg]. That is, the object is most likely to be dropped just before any contact point switches.

The rotational angle of the object until the contact point switches (i.e. how much angle is required for the switching) varies depending on the shape and orientation of the object. Conversely, detecting the shape and orientation leads to predict the switching of the contact point. Therefore, we consider that the method to predict the switching and control a robotic hand based on it is expected to improve the grasp capability of IHM. In this research, we propose the use of a camera to detect the required information.

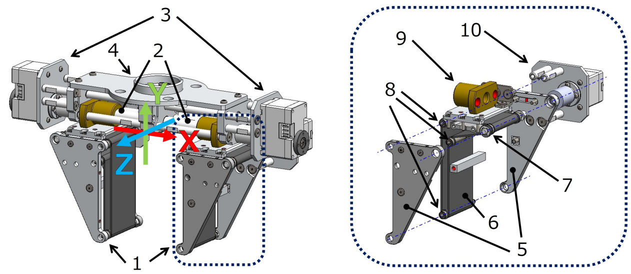
III. PROPOSED SYSTEM

A. Hardware components

In this work, we consider a two-fingered parallel gripper with conveyor belts as a robotic hand, as depicted in Fig. 2. This structure allows the grasped object to be rotated through one or more revolutions, while maintaining the grasp. Each finger unit (1) is individually controlled via a lead screw (2) by the respective motor (3). On the surface of the finger unit, a circular belt (6) is fabricated. The belt is regulated by frames (5) and three idler shafts (8) both to prevent slacking and to enable the grasped object to be conveyed along the Y(vertical) direction. The belt also wraps around a drive shaft (7) connected to a motor (10). A nut (9) is attached to the lead screw (2) on the upper side of the unit (1). These enable both the units and the grasped object to be translated along the X(horizontal) direction. As a supplement, the hand can be attached to a robotic arm via a base part (4).

With this mechanical structure, the grasped object can be translated along the X direction by actuation of the fingers. Controlling each belt to move in the same direction also enables the object to be translated along the Y direction. When the belts move in the different directions, the object can rotate around the Z axis.

The camera is mounted on the robotic hand with an attachment and captures the hand workspace continuously. In this paper, a camera is arranged directly facing the workspace. Note that no force sensor is installed due to the difficulty of design and fabrication.



1. Finger unit
2. Lead screw to move the finger 1 along the X direction
3. Finger motor connected to the lead screw 2
4. Base part connectable to a robotic arm (not shown)
5. Frame
6. Circular belt
7. Drive shaft to circulate the belt 6
8. Idler shaft regulating the belt 6 to move along the Y direction
9. Nut for lead screw 2
10. Belt motor connected to the drive shaft 7

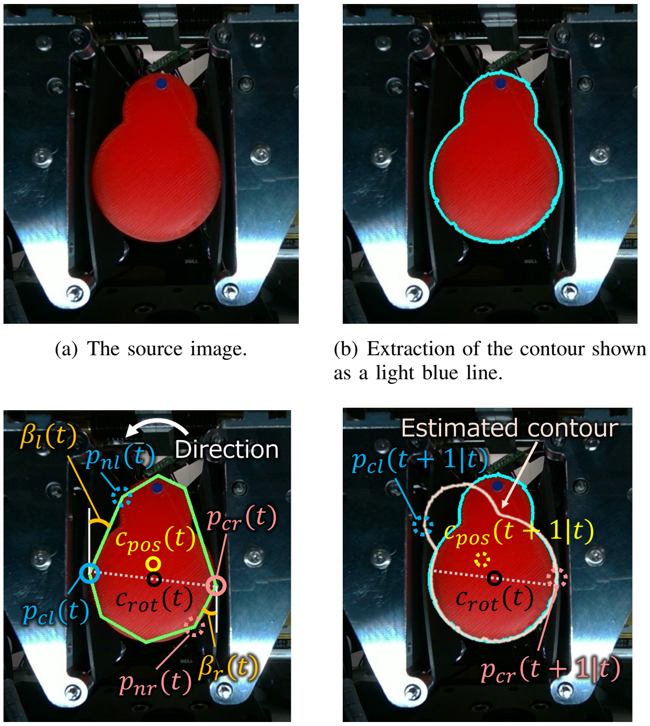
Fig. 2. The hardware component used in the proposed system and the coordinate system for control.

B. Predicting the switching of the contact points

This section introduces a method to acquire information from an image and predict the switching of the contact points. Utilizing the prediction for the belt control enables the fingers to move appropriately to grasp an object, according to the shape, size, and orientation of the object, even while manipulating it.

As mentioned in Section II, when the object is rotated while reducing its width w_o, the closer the switches of the contact points are, the higher the velocity $\Delta\dot{w}_o$ of the changes Δw_o in w_o may be, even if the angular velocity is constant. In our method, by slowing down the angular velocity of the object, $\Delta\dot{w}_o$ is adjusted so as not to drop it. For this adjustment, the rotational angle β that is the angle by which the object needs to be rotated for each contact point to switch is predicted. By adjusting the angular velocity to be directly proportional to β, Δw_o is regulated to prevent dropping. Here, β is calculated based on the contour of the object obtained from a camera image. The process to obtain such information is detailed below.

First, a contour of the grasped object is extracted from a color image, as in Fig. 3(a); the contour is depicted as a light blue line in Fig. 3(b). Next, the contour is approximated as a polygon using the Douglas–Peucker algorithm with the tolerance ε. A convex hull of the polygon is produced, as indicated by the green outline in Fig. 3(c). Note that the vertices of the hull can be regarded as candidates for the contact points because a parallel gripper can contact only the convex surface. The leftmost and rightmost vertices are defined as the respective left and right contact points $p_{cl}(t)$, $p_{cr}(t)$, indicated as blue and red circles, respectively, in the figure. Further, from neighboring vertices of these contact points and considering the direction of rotation, it can be



(a) The source image.

(b) Extraction of the contour shown as a light blue line.

(c) Calculation of the contact information when rotating in a counter-clockwise direction.

(d) Estimation of the next contour and contact points.

Fig. 3. The process of calculating the required information. The angle β between a vertical line and the line connecting the current and next contact points, p_c and p_n, is calculated to be used to control the belt. Rotating the contour, both the centroid and contact point in the next frame are estimated as $c_{pos}(t+1|t)$ and $p_c(t+1|t)$, respectively.

estimated which vertex will make contact next, namely the next contact points, $p_{nl}(t)$ and $p_{nr}(t)$. Then each angle, $\beta_l(t)$ and $\beta_r(t)$, between a vertical line and the line connecting the present and next contact points is calculated. In Fig. 3(c), the arced orange lines show the calculated angles for the left and right sides when rotating in a counterclockwise direction. $\beta_l(t)$ and $\beta_r(t)$ are the desired angles (β for the left and right side, respectively).

During this process, it has not been necessary to track any feature point between frames. This means that the prediction of the contact-point switching, which relates to the possibility of dropping the object, does not require any past information. Whatever orientation the object had in previous frames, the changes in width in the next frame are determined by both the current orientation and the rotational angle between frames. In addition, unless the object is dropped, the contact points remain on the leftmost and rightmost vertices, i.e., on the surface of the belts. Because these principles do not change regardless of the object ' s position, orientation, and characteristics, the proposed method does not need a tracking process.

C. Estimating the position of each contact point

In this section, the approach to finger control is described. The purpose of the approach is to improve the grasping capability by estimating the contact points in the next frame. It has been mentioned that the required width of the fingers during object rotation can vary, and grasping might fail because fingers cannot follow the changes in width. Though it is important to control the fingers to follow the changing width, a primary delay occurs with a feedback system. Since the amount of changes in width varies between frames depending on the shape, size, and orientation of the object, the fingers cannot follow along, due to the delay, when the amount per unit of time are large. Additionally, the delay may cause an increased load to both the hand and the grasped object, not just dropping of the object. Therefore, based on the object information, estimating where the fingers should be moved in the next frame contributes to maintaining the grasp. The procedures for estimating the requires position of each finger one frame after are detailed.

Fig. 3(c) shows a color image describing the contour of an object and the calculated contact points. The yellow circle indicates the centroid of the object and the criteria of its position, namely the positional center $c_{pos}(t)$. The center of rotation (called the rotational center $c_{rot}(t)$) is defined to be at half the distance between the points, indicated by a black circle in the figure. This definition is based on the assumption that the object is homogeneous and no slippage occurs between the surfaces of the belt and the object. Then, by rotating each point $p_i(t) = (u_i(t), v_i(t))^\top$ on the object contour around $c_{rot} = (u_{rot}(t), v_{rot}(t))^\top$ to the angle $\Delta\widehat{\theta}(t)$, the moved point $p_i(t+1|t)$ in the next frame is estimated as follows:

$$
\begin{aligned}
l_i &= \|p_i(t) - c_{rot}(t)\|_2, \\
\alpha_i(t+1|t) &= \tan^{-1}\left(\frac{v_i(t) - v_{rot}(t)}{u_i(t) - u_{rot}(t)}\right) + \Delta\widehat{\theta}(t), \\
p_i(t+1|t) &= l_i \begin{pmatrix} \cos\alpha_i(t+1|t) \\ \sin\alpha_i(t+1|t) \end{pmatrix} + c_{rot}(t),
\end{aligned}
\tag{1}
$$

where l_i is the length of the line connecting each point $p_i(t)$ to the rotational center $c_{rot}(t)$ and $\alpha_i(t+1|t)$ is the estimated angle between the horizontal axis and the line in the next frame. The estimated contour is depicted in Fig. 3(d) as a pale pink line. As a result of the rotation, the positional center $c_{pos}(t)$ and the contact points $p_{cl}(t)$, $p_{cr}(t)$ at time t move to $c_p(t+1|t)$, $p_{cl}(t+1|t)$ $p_{cr}(t+1|t)$ in the next, respectively. These values are calculated from the next contour and obtained as a result of Eq. 1. These are regarded as the estimated information and used to calculate the control command detailed in the next section.

D. Computation of the control command

In the proposed system, the fingers and belts are controlled to manipulate the object to both arbitrary position and orientation, i.e., in three degrees of freedom. They are given as the goal position X_{goal} and Y_{goal} and orientation θ_{goal}, defined as each position on the X and Y axes and rotation angle around the Z axis. Note that, the computation of the control assumes that no slippage occurs between the hand and the object.

In Sections III-B and III-C, for the simplicity of explanation, all parameters have been defined as two-dimensional values. For the control, they are converted into three-dimensional ones using stereo images, and hereafter, the converted ones are indicated by corresponding uppercase letters, except the angles such as β. Additionally, each superscript denotes the axis to which the value refers, e.g., C_{pos}^{X} represents the X value of the point C_{pos}.

Based on the information acquired from a camera, four control values $\widehat{\boldsymbol{q}}(t)$ for the fingers and belts are calculated.

$$\widehat{\boldsymbol{q}}(t)=\begin{pmatrix}\widehat{\dot{f}}_l(t)\\ \widehat{\dot{f}}_r(t)\\ \widehat{\dot{b}}_l(t)\\ \widehat{\dot{b}}_r(t)\end{pmatrix} \tag{2}$$

Here, $\widehat{\dot{f}}_l(t)$ and $\widehat{\dot{f}}_r(t)$ are the command values for the velocities of the left and right fingers along the X axis, and $\widehat{\dot{b}}_l(t)$ and $\widehat{\dot{b}}_r(t)$ are those for each belt along the Y axis.

At first, based on the smallest angle of β_l, β_r and the required angle to the goal θ_{goal}, the desired angle $\Delta\widehat{\theta}(t)$ to rotate the object by the next frame is determined as follows:

$$\Delta\widehat{\theta}(t)=K_\theta\cdot\min(\beta_\mathrm{l}(\mathrm{t}),\beta_\mathrm{r}(\mathrm{t}),\theta_\mathrm{goal}-\theta(\mathrm{t})). \tag{3}$$

where K_θ is a rotation gain and $\theta(t)$ is the observed orientation of the object in the current image.

Next, the required displacement of the left and right fingers ($\Delta\widehat{f}_l(t)$ and $\Delta\widehat{f}_r(t)$) to follow the movement of the contact points in the next frame is acquired.

$$\begin{aligned}\begin{pmatrix}\Delta\widehat{f}_l(t)\\ \Delta\widehat{f}_r(t)\end{pmatrix}&=K_{f,pos}(X_{goal}-C_{pos}^{X}(t+1|t))\begin{pmatrix}1\\1\end{pmatrix}\\&+\begin{pmatrix}P_{cl}^{X}(t+1|t)-f_l(t)\\ P_{cr}^{X}(t+1|t)-f_r(t)\end{pmatrix}-\begin{pmatrix}-X_m\\ X_m\end{pmatrix},\end{aligned} \tag{4}$$

where $K_{f,pos}$ indicates the gain values of the fingers related to translation; $C_{pos}^{X}(t+1|t)$, $P_{cl}^{X}(t+1|t)$, and $P_{cr}^{X}(t+1|t)$ are X components of the estimated centroid $c_{pos}(t+1|t)$ and contact points $p_{cl}(t+1|t)$ and $p_{cr}(t+1|t)$, $f_l(t)$ and $f_r(t)$ are the present positions of the fingers given by the encoders of each motor, and X_m is a constant that shows the deflection of the belts related to the grasp force.

Then, according to the desired angle $\Delta\widehat{\theta}(t)$ and the estimated Y component $C_{pos}^{Y}(t+1|t)$ corresponding to the centroid $c_{pos}(t+1|t)$, the desired displacement of belts $\Delta\widehat{b}_l(t)$ and $\Delta\widehat{b}_r(t)$ between frames is evaluated as follows:

$$\begin{aligned}\begin{pmatrix}\Delta\widehat{b}_l(t)\\ \Delta\widehat{b}_r(t)\end{pmatrix}&=K_{b,pos}(Y_{goal}-C_{pos}^{Y}(t+1|t))\begin{pmatrix}1\\1\end{pmatrix}\\&+\begin{pmatrix}R(t)\Delta\widehat{\theta}(t)\\ -R(t)\Delta\widehat{\theta}(t)\end{pmatrix},\end{aligned} \tag{5}$$

where $K_{b,pos}$ is the translation gain, and $R(t)$ illustrates the rotational radius which is calculated as the half lengths of the lines connecting the current rotational center $C_{rot}(t)$ (corresponding to $c_{rot}(t)$) to either contact point.

TABLE I

THE SPECIFICATIONS OF THE HAND, THE DEVICES, AND THE PARTS.

Maximum width of fingers	60 mm
Material of the belt	Chloroprene rubber(Shore A45)
Thickness of the belt	1 mm
Length of the belt in direction Y	57 mm
Width of the belt in direction Z	30 mm
Motor	Dynamixel XL330-M288-T
Camera	RealSense D405 (Intel)
Resolution	848×480 pixel

TABLE II

THE SOFTWARE SETUPS OF THE PROCESS.

Tolerance ε	25 pixels
K_θ / $K_{f,pos}$ / $K_{b,pos}$	0.3 / 0.5 / 0.1
$X_{goal}/Y_{goal}/\Theta_{goal}$	0.0 mm / −41.0 mm / 360.0 deg
X_m	1.25 mm

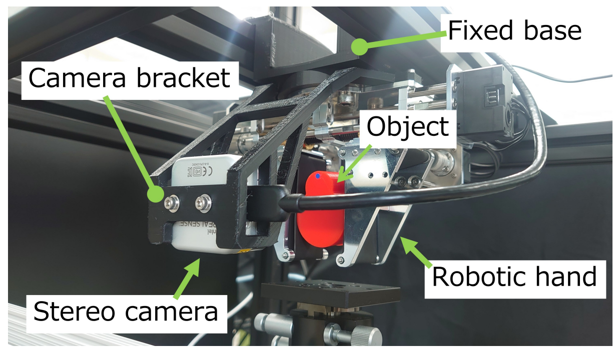


Fig. 4. The experimental setup. The robotic hand was mounted to the fixed base. A stereo camera was attached to the hand via a camera bracket.

The control value $\widehat{\boldsymbol{q}}(t)$ is calculated by dividing the given displacements of both fingers and belts by the unit of time.

IV. EXPERIMENTS

The validity of the proposed method has been verified through the experiments to manipulate grasped objects of various shapes and sizes.

A. Experimental conditions

1) Instruments and parameters: In the experiments, a robotic hand described in Fig. 2 was used. Table I shows the details of the specifications of the hand, the devices, and the parts used. Fig. 4 shows the experimental setups. The hand mounted to the fixed base. A stereo camera was attached to the hand via a bracket, so as to capture directly in front of the hand workspace. The software setups of the process are shown in Table II.

2) Tested samples: Some of the tested samples were selected by referring to the objects in [7], [12]. In addition, we prepared samples based on the number of vertices, the aspect ratio, symmetry, and convexity of the section shape. The 11 sample shapes tested are listed in Fig. 5. In total, 22 samples were tested; 11 samples are depicted in the figure; the others had homothetic shapes whose dimensions, indicated by the blue arrows, were 22.5 mm. All samples had the same thickness, 20 mm in direction Z. All samples

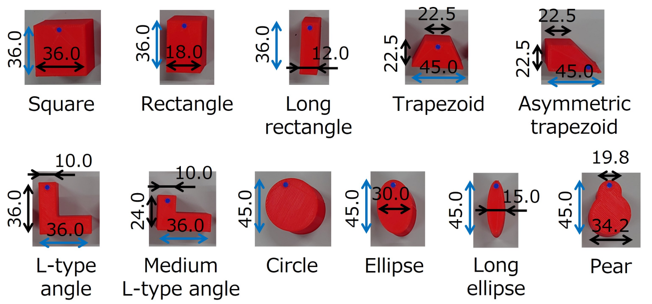


Fig. 5. The appearance of tested samples with their dimensions [mm]. In total, 22 samples, consisting of 11 types of shapes and 2 sizes (large/small samples of similar shapes), were tested. These are the large ones; the others had homothetic shapes whose dimensions, indicated by the blue arrows, were 22.5 mm. Thickness of all samples was 20 mm in direction Z.

TABLE III

DETAILS OF THE TESTING CONDITIONS. EACH ROW INDICATES WHETHER EACH CONTROL WAS USED.

Condition	Method			
	Proposed	A	B	C
Belt Control (Section III-B)	True	True	False	False
Gripper Control (Section III-C)	True	False	True	False

were produced by a 3D printer with PLA plastic. A blue marker was placed on the observable surface to calculate the orientation of each object. The angle between the X axis and the line connecting both the object centroid and the marker was defined as the orientation at the frame.

B. Evaluation and comparison

For the evaluation metrics, both the maximum angle and the average angular velocity for each sample through 10 trials were calculated. In addition to the proposed method, the experiments were also implemented under three other conditions, shown in Table III. The purpose of the comparison was to evaluate the contributions of the belt and finger control. The methods were conditioned based on whether each control was applied. In the table, the condition is "True" if the method used each control and "False" if it did not. The average metrics and their standard deviations over 10 trials for each sample were compared.

C. Results

Please refer to the submitted supplemental video of summarized behavior throughout the manipulation experiments. The results of the experiments are illustrated in Table IV and Table V. In the former table, the cases in which the goal was achieved are in bold red type. With respect to the cases in which none of the methods could achieve the given goal, each maximum angle in all methods is in non-bold red type. As to the velocities detailed in the latter table, each highest value of all methods is also shown in bold red type. In total, using the proposed method for both large and small objects, the average values for the rotatable angle and angular velocity were 205.64 [deg] and 8.58 [deg/s], and 248.13 [deg] and 15.71[deg/s], respectively. The angles of both objects and the velocity of the large ones were the highest values of all methods. The velocity of the small objects was the second highest and 0.25 [deg/s] lower than that in method A. Additionally, the proposed method had the highest number of results shown in red among all methods. That is, the proposed method has achieved high performance independent of the shape and size of the object. The frame rate was 30.8 [fps] throughout the experiments.

D. Discussion

In the experiments, 10 samples could be rotated by one revolution using the proposed method through 10 trials. With respect to the others, the failure situation was classified into two types: unrotated (unable to rotate) and dropped. Four samples (both sizes of Rectangle and both sizes of Trapezoid) have not been dropped, but could not achieve the goal neither. Eight samples (both sizes of Long rectangle, both sizes of Asymmetric trapezoid, both sizes of L-type angle, and both sizes of Long ellipse) were dropped from the hand. Regarding the failure situation, the former ones (unrotated) were related to the manipulation capability, whereas the latter ones (dropped) were due to the grasping one. The experimental results are classified into these two capabilities and discussed below. Additionally, we refer to the effectiveness of the proposed method through the comparison to the other conditions with these capabilities.

1) The manipulation capability: The results of method C shows the manipulation capability of the mechanical structure of the hand, i.e., a two-fingered parallel gripper with conveyor belts. Thanks especially to the circular belts, even with just a feedback control, which lets the fingers follow the contact points obtained from images, nine samples were successfully rotated by one revolution. Comparing the results of method A with those of method C, it is demonstrated that the proposed belt control could improve the capability in terms of both rotatable angle and angular velocity. For the large and small samples, the rotated angle using method A was 2.67 [deg] and 15.41 [deg] greater, respectively, than that using method C. The velocity results were improved by 1.21 [deg/s] and 4.37 [deg/s], respectively. The differences in improvement between the large and small samples might be due to the differences in the ease of rotating each sample. The ease depends on how large a moment is necessary to rotate the object, determined by the contact dimension and how the belt velocity contributes to the angular moment. The more points at which the object and belt come into contact, the harder it is to rotate the object. All of the samples classified as unrotated objects could not be rotated further after it was observed in the images that multiple points had been contacted. It is considered that the rotational moments from multiple points have canceled each other out, so that the required moment could not occur. Another factor is the relation between the direction of the belt movement and the angular moment. The moment from the belts moving vertically varies according to the shape and orientation of the object. For example, comparing the sample Square with Rectangle, the moment from the belt velocity for the former

TABLE IV

THE RESULTS FOR ROTATIONAL ANGLE [DEG]. THE CASES IN BOLD RED TYPE ARE GOALS THAT WERE ACHIEVED. THOSE IN NON-BOLD RED TYPE INDICATE THE MAXIMUM ANGLES OF THOSE FOR WHICH THE GOAL WAS NOT ACHIEVED.

Shape	Size	Proposed method		Method A		Method B		Method C	
		ave.	std. dev.	ave.	std. dev.	ave.	std. dev.	ave.	std. dev.
Square	large	**359.48**	0.43	**359.54**	0.54	**359.15**	0.16	**359.48**	0.75
	small	**359.15**	0.10	**359.12**	0.11	**359.13**	0.11	**359.13**	0.11
Rectangle	large	91.51	0.48	91.47	0.53	77.72	0.59	78.06	1.37
	small	90.40	1.24	90.29	1.10	90.29	0.92	90.16	1.44
Long rectangle	large	63.83	1.23	64.08	1.28	61.65	1.79	63.03	1.86
	small	59.49	1.51	58.20	1.28	57.51	1.19	58.13	0.84
Trapezoid	large	84.17	6.02	82.82	6.58	78.32	1.68	77.03	1.31
	small	88.76	1.63	87.73	1.18	87.50	2.49	87.52	2.83
Asymmetric trapezoid	large	76.00	1.36	77.65	1.22	75.68	1.96	74.21	1.51
	small	293.78	110.77	227.79	126.62	78.57	1.32	77.32	2.16
L-type angle	large	**359.43**	0.30	**359.72**	0.51	**359.66**	2.70	357.38	3.02
	small	**359.73**	0.89	**359.96**	1.51	**359.21**	0.20	**359.55**	0.79
Low L-type angle	large	95.37	54.63	76.12	0.82	76.12	1.09	76.63	1.54
	small	348.66	33.96	**359.87**	0.88	341.43	55.32	341.15	56.57
Circle	large	**359.86**	0.75	**360.39**	0.94	**359.95**	0.95	**359.74**	0.87
	small	**360.63**	0.99	**361.22**	0.36	**360.61**	0.84	**361.18**	0.13
Ellipse	large	**359.34**	0.26	**359.46**	0.61	**359.58**	0.81	**359.30**	0.60
	small	**359.51**	0.26	**359.62**	1.93	**359.82**	0.94	**359.88**	0.87
Long ellipse	large	53.86	1.52	54.46	2.02	52.10	1.75	51.21	1.69
	small	49.34	2.94	48.54	2.82	47.82	2.36	47.83	2.54
Pear	large	**359.16**	1.57	358.84	1.15	**359.30**	0.58	**359.11**	0.06
	small	**359.99**	0.95	**359.86**	0.62	**360.27**	1.04	**360.86**	0.81
Total	large	205.64	6.23	204.05	1.47	201.75	1.28	201.38	1.33
	small	248.13	14.12	242.93	12.58	227.47	6.07	227.52	6.28

TABLE V

THE RESULTS FOR ROTATIONAL VELOCITY [DEG/S]. THE CASES IN BOLD RED TYPE INDICATE THE HIGHEST VELOCITY AMONG ALL METHODS.

Shape	Size	Proposed method		Method A		Method B		Method C	
		ave.	std. dev.	ave.	std. dev.	ave.	std. dev.	ave.	std. dev.
Square	large	4.29	0.38	4.09	0.38	6.60	0.03	**6.63**	0.07
	small	9.15	0.14	9.12	0.07	10.38	0.41	**10.52**	0.07
Rectangle	large	3.73	0.98	3.37	0.46	7.35	0.03	**7.38**	0.02
	small	9.00	1.84	8.84	1.01	9.20	0.83	**9.57**	0.62
Long rectangle	large	**8.39**	0.32	8.18	0.21	7.83	0.10	7.78	0.09
	small	**12.62**	0.34	12.26	0.24	10.87	0.29	10.76	0.10
Trapezoid	large	6.70	3.00	6.64	3.69	**7.36**	0.12	7.31	0.05
	small	11.37	1.72	**14.88**	0.97	10.81	0.68	10.94	0.68
Asymmetric trapezoid	large	**7.20**	0.29	6.83	0.37	6.50	0.09	6.45	0.07
	small	**20.70**	2.01	19.76	2.11	12.34	0.25	12.31	0.25
L-type angle	large	**7.36**	0.16	7.36	0.18	7.20	0.04	7.20	0.04
	small	**12.35**	0.24	12.11	0.22	11.55	0.03	11.56	0.11
Low L-type angle	large	**8.23**	1.43	8.08	0.28	7.01	0.08	6.99	0.09
	small	9.90	1.90	**15.37**	0.74	12.46	1.31	12.58	1.09
Circle	large	**21.49**	1.06	19.60	1.74	6.65	0.02	6.65	0.05
	small	**32.00**	0.51	31.83	0.29	13.03	0.05	13.01	0.03
Ellipse	large	8.62	0.14	**8.70**	0.13	7.86	0.06	7.87	0.12
	small	18.70	0.29	**19.22**	0.41	15.17	0.05	15.16	0.07
Long ellipse	large	5.73	0.17	5.79	0.21	**7.07**	0.18	6.85	0.17
	small	11.30	0.65	10.94	0.62	**11.34**	0.25	11.29	0.22
Pear	large	12.65	1.14	**13.31**	0.35	7.59	0.03	7.57	0.02
	small	25.70	3.91	**26.21**	0.58	14.68	0.16	14.70	0.08
Total	large	**8.58**	0.82	8.36	0.73	7.18	0.07	7.15	0.07
	small	15.71	1.23	**16.41**	0.66	11.99	0.39	12.04	0.30

was greater than that for the latter. However, even if the belt velocity is raised to expand the moment, the challenge of the contact dimension remains. Attempting to resolve this problem, the shape of the belt surface will be designed to occur the appropriate moment by belts, such as concave-convex shape.

In terms of the finger control, the metric of the rotatable angle for the proposed method has been promoted in comparison with that for method A. However, the metrics of angular velocity were lower, especially for Trapezoid and Low L-type angle samples. In the experiments for these samples with the proposed method, it took longer time than that with the method A, just after it has been observed that the multiple points contacted. Although this was when the contact points were supposed to switch, the switch could not be implemented due to the lack of angular moment. However, the fingers have been opened by the feed-forward control, since it was unfortunately estimated that the switch did. Then, a short-term slippage occur, and the fingers closed so as not to drop the object. In the experiments with the proposed method, this process was sometimes repeated. These failures were caused by the cancellation of moment between multiple contact points, as well as the above challenge for unrotated objects. Therefore, the mentioned enforcement of the belt design is considered to settle it too. By such consideration of the design, it is supposed that the manipulation capability can be improved.

2) The grasping capability: In the experiments for some samples, each sample was dropped when rotation was attempted. Depending on the shape and orientation of such samples, the rotation was caused by the fingers, not just the belts. That is, the grasping force also enabled generation of the angular moment. Due to the integration of the angular moments from both the fingers and belts, the larger moment possibly occurred and the required width of the fingers could decrease than predicted. Dealing with it, the approach to rotate the object just by the fingers is considered. In addition, the belts will be controlled in the opposite direction to a goal angle so that it does not rotate more than predicted. Based on the combination of the fingers and belts, we will attempt to improve grasping capability in the future.

V. CONCLUSIONS

In this paper, we proposed an in-hand manipulation method for a robotic hand with belts via predicting the changes in the contact points from camera images. To improve the rotatable range of the grasped object, a two-fingered parallel gripper equipped with conveyor belts were used as a robotic hand. The control methods for both the fingers and the belts have been developed using an image from a stereo camera attached to the hand, so as to increase the variety of manipulatable objects. At first from time-series images, the contour of the grasped object was acquired. Based on the contour, the switching of the contact points between the surfaces of the belts and the object was predicted. According to the prediction of the switching, the velocities of the belts were controlled to adjust the rotational angle of the object so as to allow the fingers to not drop it. Then, also from the acquired contour, the contact points in the next frame are estimated. The fingers were moved according to the estimated points. The improvement of both grasping and manipulation capabilities has been evaluated through experiments to manipulate 22 samples (11 shapes and 2 sizes) to one revolution. Additionally, the proposed method was compared with the other three methods. Throughout the experiments, the average rotational angle for the proposed method was the largest of all; 205.64 [deg] for large 11 samples and 248.13 [deg] for small ones. Though the number of the samples that could reach the goal was 10 samples for the proposed method and was also the highest, 12 samples were dropped or not rotated anymore.

In future works, two approaches will be applied separately to promote the manipulation and grasping capabilities individually. One approach is mechanical, involving the design of the belt geometry. This approach is aimed ensuring that the moment to rotate the grasped object is applied appropriately. The other approach is a control method to maintain the grasp. Combining the control of fingers and belts, the desired grasp and rotation are achieved simultaneously.

ACKNOWLEDGMENT

This paper is based on results obtained from a project, JPNP20016, subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] R. R. Ma and A. M. Dollar, "On Dexterity and Dexterous Manipulation," in Proc. of Int. Conf. on Advanced Robotics, pp. 1–7, 2011.

[2] I. M. Bullock *et al.*,"A Hand Centric Classification of Human and Robot Dexterous Manipulation," IEEE Trans. on Haptics, Vol. 6, No. 2, pp. 129–144, 2013.

[3] A. Bicchi, "Hands for Dexterous Manipulation and Robust Grasping: A Difficult Road Toward Simplicity," IEEE Trans. on Robotics and Automation, Vol. 16, No. 6, pp. 652–662, 2000.

[4] A. M. Okamura *et al.*,"An Overview of Dexterous Manipulation," in Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 255–262, 2000.

[5] N. Chavan-Dafle *et al.*,"Pneumatic Shape-Shifting Fingers to Reorient and Grasp," in Proc. of Int. Conf. on Automation Science and Engineering, pp. 988–993, 2018.

[6] J. Amend and H. Lipson, "The JamHand: Dexterous Manipulation with Minimal Actuation," Journal of Soft Robotics, Vol. 4, No. 1, pp. 70–80, 2017.

[7] A. S. Morgan *et al.*, "Towards Generalized Manipulation Learning through Grasp Mechanics-Based Features and Self-Supervision," IEEE Trans. on Robotics, Vol. 37, No. 5, pp. 1553–1569, 2021.

[8] Y. Hou *et al.*,"Fast Planning for 3D Any-Pose-Reorienting Using Pivoting," IEEE Int. Conf. on Robotics and Automation, pp. 1631–1638, 2018.

[9] N. Chavan-Dafle and A. Rodriguez, "Stable Prehensile Pushing: In-Hand Manipulation with Alternating Sticking Contacts," IEEE Int. Conf. on Robotics and Automation, pp. 254–261, 2018.

[10] J. Zhou *et al.*,"BCL-13: A 13-DOF Soft Robotic Hand for Dexterous Grasping and In-Hand Manipulation," IEEE Robotics and Automation Letters, Vol. 3, No. 4, pp. 3379–3386, 2018.

[11] M. Andrychowicz *et al.*, "Learning Dexterous In-Hand Manipulation," Int. Journal of Robotics Research, Vol. 39, No. 1, pp. 3–20, 2020.

[12] H. Ichikura and M. Higashimori, "In-Hand Manipulation Inspired by Diabolo Juggling," IEEE Trans. on Robotics and Automation, Vol. 7, No. 4, pp. 12227–12234, 2022.

[13] K. Tahara *et al.*,"External Sensorless Dynamic Object Manipulation by a Dual Soft-Fingered Robotic Hand with Torsional Fingertip Motion," in Proc. of IEEE Int. Conf. on Robotics and Automation, pp.4309–4314, 2010.

[14] S. Yuan *et al.*,"Design of a Roller-Based Dexterous Hand for Object Grasping and Within-Hand Manipulation," in Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 8870–8876, 2020.

[15] V. Tincani *et al.*,"Implementation and Control of the Velvet Fingers: a Dexterous Gripper with Active Surfaces," in Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 2744–2750, 2013.

[16] Y. Cai *et al.*, "In-Hand Manipulation in Power Grasp: Design of an Adaptive Robot Hand with Active Surfaces," in Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 10296–10302 , 2023.

[17] K. Morino *et al.*,"Sheet-Based Gripper Featuring Passive Pull-In Functionality for Bin Picking and for Picking Up Thin Flexible Objects," IEEE Robotics and Automation Letters, Vol. 5, No. 2, pp. 2007–2014, 2020.

[18] A. Kakogawa *et al.*,"Sensor-Less and Control-Less Underactuated Grippers with Pull-In Mechanisms for Grasping Various Objects," Frontiers in Robotics and AI, Vol. 8, pp. 1–11, 2021.

[19] R. R. Ma and A. M. Dollar, "In-Hand Manipulation Primitives for a Minimal, Underactuated Gripper with Active Surfaces," in Proc. of ASME Int. Design Engineering Technical Conf. and Computers and Information in Engineering Conf., pp. 1–7, 2016.