Soft Human-Robot Handover Using a Vision-Based Pipeline

Chiara Castellani, Enrico Turco, *Member, IEEE*, Valerio Bo, *Member, IEEE*, Monica Malvezzi, *Member, IEEE*, Domenico Prattichizzo, *Fellow, IEEE*, Gabriele Costante, *Member, IEEE*, and Maria Pozzi, *Member, IEEE*

***Abstract*—Handing over objects is an essential task in human-robot collaborative scenarios. Previous studies have predominantly employed rigid grippers to perform the handover, focusing on generating grasps that avoid physical contact with people. In this paper, we present a vision-based open-palm handover solution where a soft robotic hand exploits contact with the human hand for improved grasp success and robustness. The human-robot physical interaction allows the robotic hand to slide over the human palm and firmly cage the object. The identification of the human hand plane and object pose is achieved through a versatile perception pipeline that exploits a single RGB-D camera. Through experimental trials, we show that the system achieves successful grasps over multiple objects with different geometries and textures. A comparative analysis assesses the robustness of the proposed *soft* handover method against a baseline approach. A study with 30 participants evaluates users' perception of human-robot interaction during the handover, highlighting the effectiveness and preference for the proposed pipeline.**

***Index Terms*—Grasping, physical human-robot interaction. soft robot applications.**

I. Introduction

ENABLING a safe, robust, and deliberate physical interaction between humans and robots is crucial for the deployment of collaborative robots in real-world scenarios [1]. In this context, the handover task is particularly challenging as it involves managing the physical exchange of the object, as well as the cognitive process of deciding what needs to be passed and when and where the transfer should happen [2]. Handover requires the human and the robot to operate in close proximity and with high coordination in time and space. Thus, a robust perception pipeline is essential. Previous works usually rely on vision or proximity sensing in the *pre-handover* phase and sometimes on force and tactile sensing in the *physical handover* phase [2].

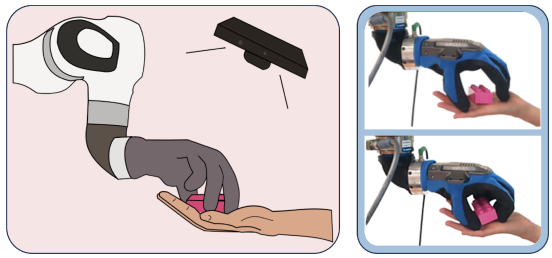


Fig. 1. *Soft* handover: Concept and real implementation. The robotic hand exploits the contact with the human hand to grasp an object by sliding the fingers over the palm.

However, less attention is devoted to the intrinsic safety of the adopted end-effector. Most of the existing solutions use rigid grippers and focus on avoiding contact between the human and the robot during the handover [3], [4]. As a result, the proposed approaches are functional only when applied to objects that are big enough to prevent the robot from touching the person's hand during the grasping motion. Thanks to their intrinsic passive compliance, soft grippers represent a safer choice for human-robot handover [5], [6], [7].

The literature on soft grippers, however, has mostly focused on autonomous manipulation, with the introduction of the so-called "environmental constraints exploitation" strategies, i.e., new bioinspired grasping strategies in which the robotic hand purposefully interacts with the constraints present in the surrounding environment (e.g., tables, walls, edges) to robustly envelop the object [8], [9].

In the context of human-to-robot handovers, the surface of the human hand itself could be considered as an environmental constraint that can be leveraged to better grasp objects. Based on this intuition, this work presents a solution for open-palm

Received 29 July 2024; accepted 15 November 2024. Date of publication 4 December 2024; date of current version 18 December 2024. This article was recommended for publication by Associate Editor Z. Erickson and Editor J. Borràs Sol upon evaluation of the reviewers' comments. This work was supported in part by the European Union through the Next Generation EU project ECS17 "THE-Tuscany Health Ecosystem" (PNRR MUR M4 C2 Inv. 1.5, CUP B63C22000680007, Spoke 9: Robotics and Automation for Health) and in part by the Horizon Europe project "HARIA - Human-Robot Sensorimotor Augmentation - Wearable Sensorimotor Interfaces and Supernumerary Robotic Limbs for Humans with Upper-limb Disabilities" under Grant 101070292. *(Chiara Castellani and Enrico Turco contributed equally to this work.) (Corresponding author: Maria Pozzi.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by CAREUS Univ. of Siena under Application No. 13/2023.

Chiara Castellani, Enrico Turco, Valerio Bo, Domenico Prattichizzo, and Maria Pozzi are with the Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy, and also with the Department of Humanoids and Human Centered Mechatronics, Istituto Italiano di Tecnologia, 16163 Genoa, Italy (e-mail: chiara.castellani@iit.it; enrico.turco@iit.it; valerio.bo@iit.it; domenico.prattichizzo@unisi.it; maria.pozzi@unisi.it).

Monica Malvezzi is with the Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy (e-mail: monica.malvezzi@unisi.it).

Gabriele Costante is with the Department of Engineering, University of Perugia, 06125 Perugia, Italy (e-mail: gabriele.costante@unipg.it).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2024.3511415, provided by the authors.

Digital Object Identifier 10.1109/LRA.2024.3511415

human-to-robot *soft* handovers, where a soft robotic hand deliberately and compliantly enters in contact with the human hand using impedance control and grasps the given object by sliding the fingers over the palm (Fig. 1), exploiting the physical interaction to improve the grasp robustness. This goes beyond previous works in which human-robot contact is avoided or merely mediated [2], [5]. To identify the human hand plane, i.e., the "constraint" that the robot should exploit to grasp the object, a robust and versatile visual perception pipeline is proposed and implemented in this paper. The pipeline relies on a single RGB-D camera and also provides the object pose.

To summarize, the key contribution of our work is an open-palm human-to-robot handover strategy, the *soft* handover, in which a soft robotic hand purposefully exploits the contact with the human hand to enhance grasp robustness. The human hand plane and the object point cloud are retrieved by integrating state-of-the-art components into a new robust and versatile perception pipeline based on a single RGB-D camera. The proposed contribution is evaluated through an objective comparison between the soft handover and a baseline method in terms of grasp success across different objects and different human hands. Additionally, a subjective evaluation of users' perception of ease, naturalness, reliability, annoyance, and trust during the handover process is conducted. This comprehensive validation highlights the potential of soft robotic hands to improve handover performance and enhance human-robot interaction through deliberate physical contact, offering a promising alternative to traditional rigid grippers.

II. Related Works

Most of the solutions proposed for human-to-robot handovers using real-time robotic vision involve the use of parallel jaw grippers [3], [4], [10], [11], [12], [13]. In [3], Rosenberger et al., developed an object-independent human-to-robot handover system using a single gripper-mounted RGB-D camera. Their design focused on safe human-robot interaction, avoiding unintended collisions by employing redundant segmentation and excluding grasp points near the human. However, these safety measures frequently led to handover failures, and the real-time execution was computationally demanding. Differently, in [12], [13], skeleton tracking techniques are used to determine the readiness of the human for object handover. Micelli et al. [12] used a clustering algorithm to find the hand in the point cloud and estimated the bounding box of the object. However, their method relied on several assumptions, including the known positioning of objects relative to the hand and the point of view of the camera. Liu et al. [13], instead, relied only on the wrist position and proposed a task-agnostic framework to adapt to different working conditions. This made the robot behave more conservatively in the presence of uncertainties to ensure no collisions with the human.

Among human-robot handover pipelines, one of the most complete works has been developed by Yang et al. [4]. This work presents a human-to-robot handover system that can be applied to diverse unknown objects, achieving accurate hand and object segmentation and temporally consistent grasp generation. Extensions of [4] were presented in [11], where the authors introduced a Model Predictive Control framework for smoother handover and in [14], where a novel framework for learning human-to-robot handovers from point clouds was proposed. However, the employed vision pipeline is based on a neural network that needs multiple data for training and is optimized for a specific handover position. Thus, the hand segmentation might not work when the hand is partially occluded by the object or if the camera position changes. Furthermore, the method is tailored for parallel-jaw grippers.

Only a few previous works have adopted soft or anthropomorphic hands to perform handover tasks [5], [6], [15]. In [6], for example, authors focused on the subjective evaluation of the human-robot interaction and found out that adding a small delay before moving the robot towards the object increases the perceived handover quality. To track the object position and orientation, the authors used an OptiTrack motion capture system, leading to a rather structured environment, and did not allow human-robot contact. Bianchi et al. [5] have used the handover task to apply and evaluate touch-based grasp primitives in which the hand closure and wrist movements were coordinated based on data acquired by Inertial Measurement Units attached to the robot hand fingertips. Nevertheless, the task was simplified as the human handed the object to the robot in a stationary position. In the context of dexterous handovers, Duan et al. [15], investigated learning-based approaches for anthropomorphic robotic hands, allowing for more natural and effective handovers by leveraging human-like dexterity.

In this work, similarly to [4], we introduce a vision-based handover system that is adaptable to a wide range of objects. Differently from [4], our pipeline is specifically designed to explicitly deal with interactions between the human and the soft robotic hand, leading to different perception and grasp planning requirements. Furthermore, our approach uses a single RGB-D camera, focuses on open-palm and challenging handovers (e.g., small objects), and is able to provide a hand plane estimation to plan a coherent grasp. Only a few previous works explicitly consider the case of open-palm handovers. In [10], for example, "on-open-palm" is only one of the considered cases in the proposed classification of human grasp types for handover. Additionally, to the best of our knowledge, there are no works that specifically tackle this problem with soft hands.

III. Methodology

In the proposed solution for open-palm human-to-robot handover, the assumptions related to the operating conditions are kept as loose as possible. Our pipeline employs a single RGB-D camera and, differently from previous works [4], [6], [12], it does not rely on specific tracking systems. The giver is expected to hand over the object with the hand fully open or even partially closed around the object, provided that the object is visible. The palm has to face upwards, and hand inclination is allowed, provided that the object is held stably.

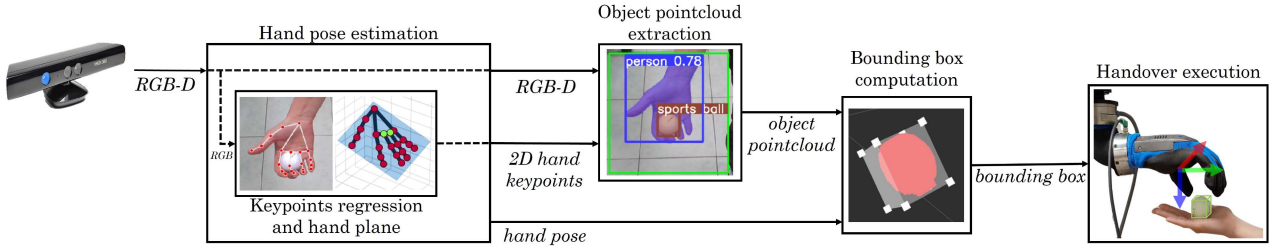


Fig. 2. Vision-based pipeline for soft handover. First, RGB-D data are processed to estimate the human hand pose. Then, the 2D hand keypoints combined with RGB-D information are used to remove the human hand depth information, extracting the object point cloud only. A 3D oriented object bounding box is retrieved by exploiting the object point cloud and the hand pose information. Eventually, a handover is executed when the bounding box estimation is consistent over time.

A. Perception Pipeline

The main steps of the proposed perception pipeline are reported in Fig. 2 and explained in the following.

1) Hand Pose Estimation: The first step consists of the regression of the hand keypoints from the RGB image. In literature, other approaches also exploit depth information [16], [17] to estimate the 3D hand joint locations or to model the interaction with objects [18], [19], but they are very sensible to occlusions. Conversely, MediaPipe library [20] has been proven to perform very well for close-proximity human-robot interaction [21]. For this reason, we chose to adopt the MediaPipe Hands Full model, which uses Convolutional Neural Networks to predict online both the 2D position of the hand keypoints on the image and their 3D coordinates relative to the hand's geometric center.

To estimate the hand pose, we proceeded as follows. First, we compute the 3D hand position (x_h, y_h, z_h) by projecting in 3D the pixel coordinates (u_h, v_h) of the point that is in between the knuckles of the medium and ring fingers (in green in Fig. 2). To this aim, we use the depth information provided by the RGB-D camera as follows: $x_h = \frac{(u_h - c_x) \cdot Z_h}{f_x}$, $y_h = \frac{(v_h - c_y) \cdot Z_h}{f_y}$, $z_h = Z_h$, where Z_h is the depth value at pixel (u_h, v_h), c_x, c_y are the principal point offsets, and f_x, f_y are the camera's focal lengths. Then, we retrieve the 3D hand orientation by computing the orientation of the plane that best fits a subset of 3D hand keypoints. We call this plane "hand plane", and we estimate it based on least-squares fitting and singular value decomposition [22]. The final hand pose is obtained by setting the z-axis normal to the hand plane and the x, y-axes laying on it. Note that to evaluate the hand plane, we do not consider the fingertips keypoints as they might not be always aligned with the hand palm.

2) Object Pointcloud Extraction: To delimit the area where the object is expected to be found, a 3D cubical region is computed around the estimated hand position. The vertices of this 3D region are projected onto the RGB image to infer a coherent 2D region surrounding the hand. Then, instance segmentation is performed on this RGB portion using YOLOv8 [23]. An example of the YOLO output is shown in the block *Object pointcloud extraction* in Fig. 2.

The procedure to obtain the object mask from the cropped RGB image is detailed in Algorithm 1. To make the vision pipeline tolerant to possible YOLO errors we implemented two different strategies. First, we check if each of the instance

Algorithm 1: Procedure to Obtain The Object Mask.

Input: RGB_image, $hand\_keypoints$
Output: obj_mask
1 $obj\_masks, hand\_masks \leftarrow$ YOLO(RGB_image)
2 $obj\_mask \leftarrow$ zeros(size(RGB_image))
3 **if** length$(obj_masks) > 0$ **then**
4 **foreach** $mask \in obj\_masks$ **do**
5 **if** $\#hand\_keypoints\ in\ mask < 15$ **then**
6 $obj\_mask \leftarrow obj\_mask \vee mask$
7 **else**
8 **if** length$(hand_masks) > 0$ **then**
9 $hand\_mask \leftarrow$ zeros(size(RGB_image))
10 **foreach** $mask \in hand\_masks$ **do**
11 $hand\_mask \leftarrow hand\_mask \vee mask$
12 $obj\_mask \leftarrow \neg(hand\_mask)$

segmentation masks classified as objects (*obj_masks*) contains a number of hand keypoints less than a certain threshold (line 5). If this is the case, that mask is considered to contribute to the object mask, and we combine them by performing a bitwise OR operation (line 6). Otherwise, the mask is excluded. In this way, we detect and filter out potential misclassifications of the hand as another object. This strategy improves robustness by minimizing false positives, though it may fail when the object highly occludes the hand. Second, if the object detection fails, but the hand is correctly identified, we use the segmentation mask of the latter to retrieve the object mask (lines 8-12). In other words, we infer the object mask by taking the inverse of the hand mask. This method allows for the identification of objects significantly different from the categories on which YOLO is trained.

The final object segmentation mask is then used to filter the depth image and keep only the depth values pertaining to the object representation. Using this filtered depth data, the point cloud of the object is generated by deprojecting the pixel coordinates, using the equations in Section III-A1.

3) Bounding Box Computation: To compute the oriented bounding box of the object, we follow two steps. First, we reshape the object point cloud by projecting it onto the hand plane and its normal vector. Second, the bounding box of the object is computed using Open3D library which performs the

Principal Component Analysis (PCA) of the convex hull of the reshaped object point cloud [24]. As a result, we obtain a bounding box that has one side parallel to the hand plane. The accuracy and consistency over time of the estimated bounding box are evaluated by monitoring its volume and position during the pre-handover phase. If no sudden changes in their values are detected, the handover can start.

B. Soft Handover Execution

The handover is executed by a robot arm with a soft hand attached to it, as in Fig. 1. Initially, the robot hand is partially closed in a predefined grasp pre-shape to ensure that, above all with small objects, the robotic fingers first enter in contact with the human hand and then close around the object by sliding on the human palm. In other words, the robot hand can perform a surface-constrained grasp [8]. First, the robotic hand is positioned above the object center with a fixed offset. As shown in the *Handover execution* block in Fig. 2, its z-axis (in blue) is perpendicular to the hand plane, while its y-axis (in green) aligns with the shortest side of the bounding box [25]. Then, the soft hand descends along its z-axis, while we monitor the sensed force. When its value exceeds a predefined threshold or the target position is reached, the descending motion stops, and the hand closes. In this phase, a Cartesian impedance controller is employed, designed to have a compliant behavior along the soft hand z-axis and to be stiff along the other directions. This allows a compliant interaction between the robot and the giver's hand.

IV. Experiments

The adopted experimental setup is shown in Fig. 3(a). A desk-mounted Franka Emika Panda collaborative robot arm was used with a Pisa/IIT SoftHand attached to its end-effector [26]. An ATI Gamma Force/Torque sensor was placed at the wrist of the robot to measure the interaction force with the human and to trigger the hand closure. The perception pipeline was developed to work for a generic depth camera, and it was tested here with a Kinect v1 camera.

We conducted three different experiments, where we compared the soft handover with a baseline approach. In the latter, the robotic arm is position-controlled, and the hand is closed to perform a pinch grasp over the object. Similarly to what a parallel-jaw gripper would do, the thumb and index fingers of the soft hand are closed avoiding contact with the human palm. The same robotic hand is adopted both in the soft handover and in the baseline to ensure we compare different grasping strategies keeping the same end-effector.

Experiment 1 compared the grasp success rates of the proposed system when using the soft handover and the baseline method over different objects. In Experiment 2, we measured the robustness of the two control strategies under the occurrence of errors in the object pose estimation. Lastly, Experiment 3 compared the perception of naturalness, reliability, annoyance, and trust during the object transfer with different users and tested the pipeline robustness with various human hands.

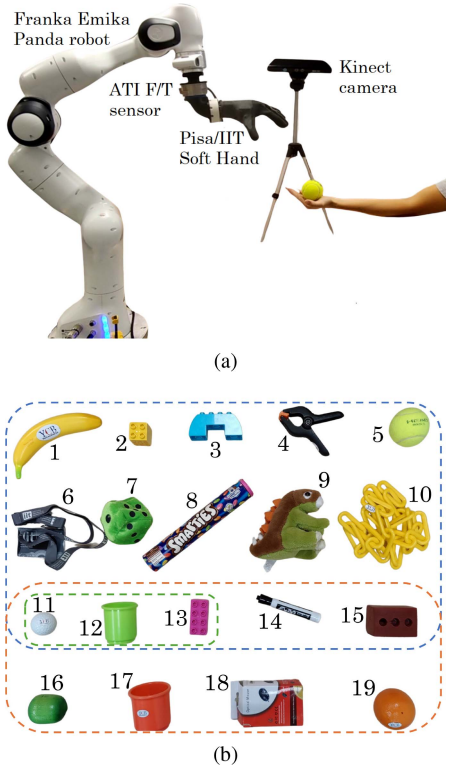
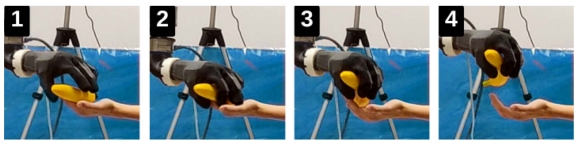


Fig. 3. (a) Experimental setup. (b) Selected objects. Objects outlined in blue are used in experiment 1, those in green in experiment 2, and those in orange in experiment 3.

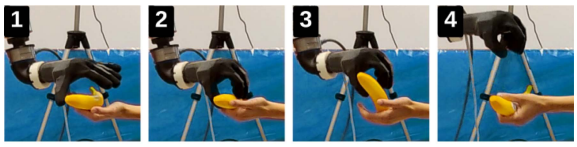
TABLE I
Properties of the Objects Used in the Experiments

ID	Object	Size (cm)	Mass (g)
1	Banana (YCB)	⌀ 3.6 x 19	66
2	Yellow Lego (YCB)	3.2 x 4.3 x 3.2	12.7
3	Blue Lego (YCB)	9.6 x 3.2 x 4.3	26.9
4	Clamps (YCB)	9 x 11.5 x 2.7	59
5	Tennis ball (YCB)	⌀ 6.47	58
6	Badge holder w/ lanyard	8 x 9.7 x 0.2	25
7	Soft dice	5.5 x 5.5 x 5.5	9
8	Smarties	⌀ 1.8 x 22.8	12
9	Toy plush	10 x 12 x 12	50
10	Plastic chain (YCB)	111.5	98
11	Golf ball (YCB)	⌀ 4.27	46
12	Green cup (YCB)	⌀ 6.5 x 6.7	15.1
13	Pink Lego (YCB)	3.2 x 2.3 x 6.4	12.8
14	Marker (YCB)	⌀ 1.15 x 13.7	10
15	Foam brick (YCB)	3 x 2.7 x 16.6	8.7
16	Lime	⌀ 5 x 6.7	70
17	Red cup (YCB)	⌀ 7.5 x 6.8	21
18	Box	4 x 7 x 10.5	27
19	Orange (YCB)	⌀ 7.3	47

In all the experiments, the robot moved towards the object, grasped it, and returned to its initial configuration, where it kept the object for 2 s. The handover started when the system was confident about the bounding box estimation (Section III-A), and was considered successful if the object did not fall until the release. The employed objects are depicted in Fig. 3, with their properties listed in Table I. Most of them belong to the YCB Object Set [27].



(a)



(b)

Fig. 4. Experiment 1: Successful and unsuccessful grasp sequences of the Banana: (a) Soft handover, (b) baseline.

A. Experiment 1: Pipeline Validation

For each control strategy, i.e., the soft handover and the baseline, a set of 10 trials were performed for each of the 15 objects circled in blue in Fig. 3(b). Representative examples of successes and failures observed in Experiment 1 are illustrated in Fig. 4. Notice that when the soft handover is employed (Fig. 4(a)), the robot exploits the contact with the giver's hand to perform the grasp (frame 2), resulting in a direct physical interaction between the robot and the person's hand throughout the grasp. On the contrary, the position-based approach does not involve intentional interaction with the user (Fig. 4(b)). Fig. 5(a) reports the results obtained for the two control strategies in terms of grasp success rate.

B. Experiment 2: Robustness Under Object Pose Uncertainty

For this experiment, we systematically introduced an error in the estimation of the center of the object bounding box while keeping the effective pose of the object constant.

To this aim, we defined a bi-dimensional grid corresponding to gripper displacements along the human palm plane and relative to the center of the object. The grid's vertical and horizontal directions correspond to the object's shortest and longest sides, respectively. We chose a maximum offset from the object center equal to $1/3$ of the object size.

For this experiment, we selected the three objects circled in green in Fig. 3(b), for which we obtained comparable performance using the two control strategies in Experiment 1: Golf ball, Pink Lego, and Green cup. Following the evaluation method proposed in [28], we made two grasp attempts for each position in the grid, and we recorded the success rate. Results are shown in Fig. 5(b).

C. Experiment 3: User Study

Experiment 3 involved 30 participants (20 males and 10 females, aged between 24 and 61 years). The experimental campaign adhered to the Declaration of Helsinki, and all participants provided written informed consent and could withdraw at any time. No compensation was given. The experimental protocol was approved by the Ethical Committee of the University of Siena "CAREUS", no. 12/2023. Users were first introduced to the system with a brief demonstration of the handover process. They were then asked to evaluate the difference in the sensations perceived towards the two handover strategies.

The nine objects circled in orange in Fig. 3(b) were tested. We required participants to deliver each object once per strategy, trying not to move during the handover. For each object, a user tested both strategies in the same position. The experiment consisted of 18 handovers and lasted approximately 30 minutes per user. We randomized the order of objects and strategies to minimize bias and learning effects.

After having tested an object, each participant was asked to evaluate the following statements: S.1: *"It was easy to transfer the object to the robot"*; S.2: *"The interaction with the robot was natural"*; S.3: *"The grasp was found to be robust and reliable"*; S.4: *"The interaction with the robot was annoying"*; S.5: *"I trusted the robot during the transfer of the object"*. The statements from 1 to 4 were adapted from the Robotic Social Attributes Scale (RoSAS) [29], whereas S.5 was adapted from [4]. Each statement was assessed using a 7-point scale ranging from 1 to 7, where values closer to 1 indicated a strong preference for the baseline approach, and values closer to 7 indicated a strong preference for the soft handover strategy. Fig. 6 collects the thirty participants' perceived sensations grouped per tested object. Given that the data are ordinal and come from paired observations, we used the Wilcoxon signed-rank test to compare responses. This test allowed us to determine if there were statistically significant differences in the ratings between the two methodologies across the five questions.

At the end of the complete experiment, each participant filled out a reduced RoSAS questionnaire to evaluate some specific aspects of the users' experience. Participants were asked to rate on a 7-point scale from "not at all" to "very much so" the perceived competence (reliability, responsivity, capability) and discomfort (awkwardness, scariness, danger) (Fig. 7). Lastly, we collected participants' open feedback to determine their overall impressions of the strategies.

V. Discussion

A. Experiment 1: The Proposed Pipeline is Effective Under Different Control Strategies

Overall, the perception pipeline was effective in estimating the pose of the selected objects. However, the following issues were found in some cases: *i)* the object point cloud was not accurate enough because of inaccuracies in the instance segmentation masks, *ii)* depth information was missing on dark surfaces due to excessive absorption of light, and *iii)* the estimation of the giver's hand orientation was wrong. The overall success rate for the soft handover was 89.3%, while the counterpart strategy obtained the success rate of 77.3%. For the majority of the objects, the two strategies showed comparable performance. However, when handing over small or irregularly shaped objects (e.g., yellow Lego, banana, plastic chain, badge holder), the soft handover showed a better performance. Indeed, a purely position-based control strategy requires precise object localization and is more likely to fail when errors occur in the estimation of the object

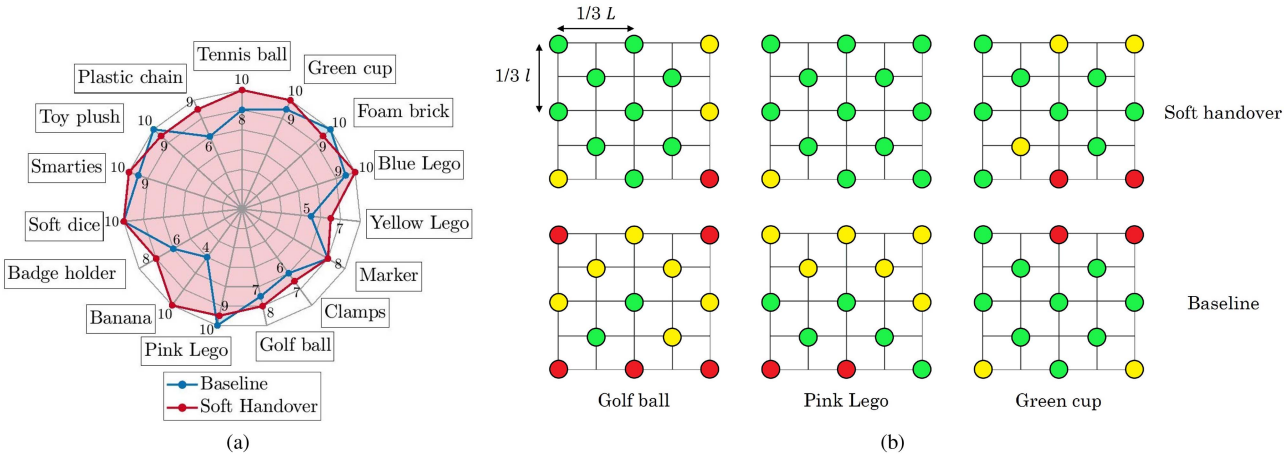


Fig. 5. (a) Experiment 1: Grasp success rate over 15 objects, considering 10 trials per object. (b) Experiment 2: The grasp success rate for different target positions around the object identifies the area in which a certain handover strategy performs better. The size of the grid is proportional to l and L, which indicate the length of the shortest and longest sides of the object, respectively. ●: Two successful grasps; ●: One successful grasp; ●: Two failed grasps.

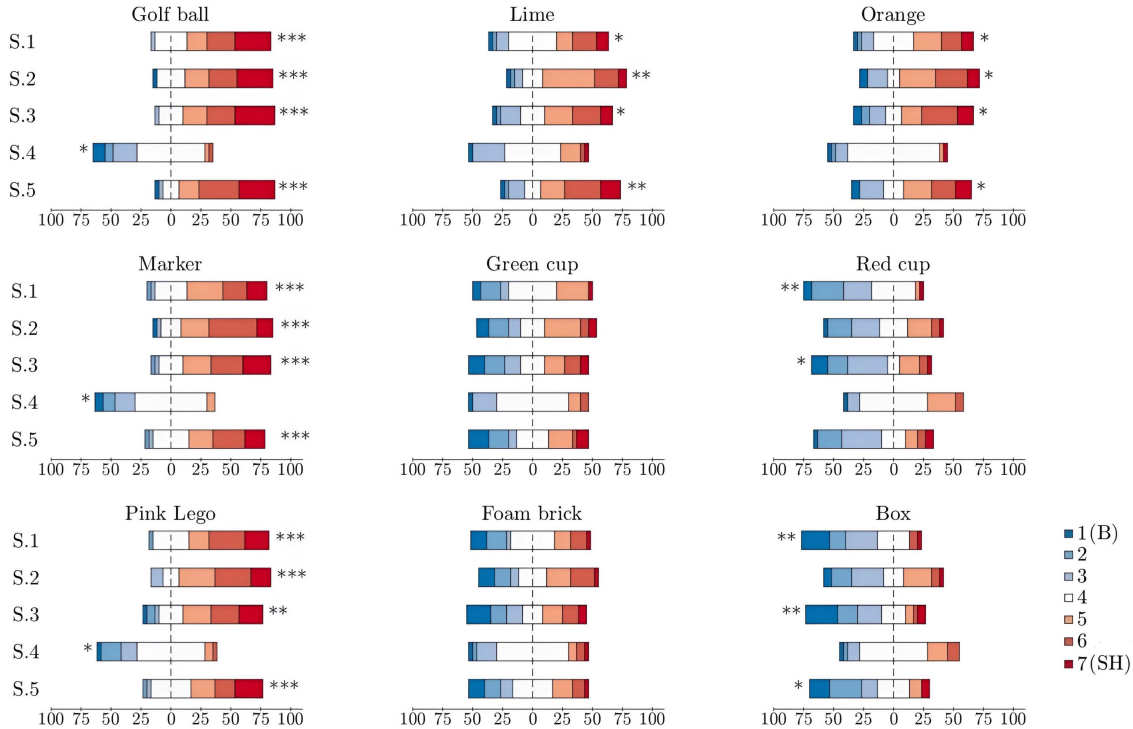


Fig. 6. Experiment 3: Answers of the 30 participants who tested the two handover strategies and evaluated ease, naturalness, reliability, annoyance, and trust during object transfer on a linear scale ranging from 1 (baseline) to 7 (soft handover). Answers are grouped by the tested objects, which are organized by shape in rows (spheres, cylinders, and cuboids), and by size in columns, with size increasing from left to right. The symbols $*$, $**$, and $***$ indicate a p-value lower than 0.05, 0.01, and 0.001, respectively.

position. In contrast, the strategy exploiting the interaction with the human palm increased the chance to perform successful handovers when the object localization was less accurate. For instance, the yellow Lego point cloud was often not accurate due to the lower performance of the segmentation network on this specific object category. The plastic chain posed challenges due to its flexible and irregular structure, whereas the badge holder, being flat and thin, was difficult to be correctly grasped. A different reasoning regards objects like the banana, in which the center of the bounding box falls outside the object itself. The position-based approach failed in these cases since the hand was not well centered with respect to the object (Fig. 4(b)). Conversely, in the soft handover, the soft hand could compensate for the uncertainty by properly exploiting the human palm, demonstrating more robustness in these cases (Fig. 4(a)). How softness can deal with uncertainty is a concept already known in

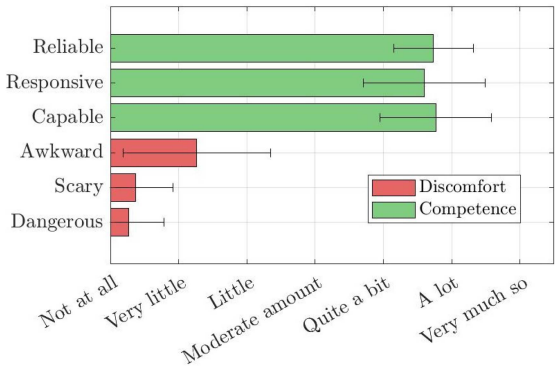


Fig. 7. Experiment 3: Results from RoSAS questionnaire. The bars indicate the mean and the whiskers the standard deviation of participant responses. Green and red bars correspond to competence and discomfort statements, respectively.

the literature and was proven effective in autonomous grasping when employing soft hands [8], [25]. In this paper, we show that this paradigm can also be exploited in human-robot interaction.

B. Experiment 2: The Exploitation of Direct Robot-Human Contact Increases Grasp Robustness

This experiment confirmed what was previously discussed in Section V-A: the soft handover is more robust when there are uncertainties on the object position.

The accuracy of the bounding box prediction is crucial when grasping small objects. For instance, a deviation in the Golf ball position estimation is very likely to cause a failure when employing the position-based strategy (Fig. 5(b)), which results in a grasp success rate of 38%. The soft handover, instead, improves the success rate to 81% by addressing uncertainties through contact with the human hand.

Considering the soft hand aligns with the shortest side l of objects whose l and L present highly different values (e.g. the pink Lego), we notice the two strategies perform likewise when errors are along L. Conversely, when the errors are introduced along l, the soft handover is able to successfully compensate for them, while the baseline fails, especially when the error is equal to $l/3$ (see the red dots in Fig. 5(b) for the Pink Lego grasped with the baseline). The success rate for the Pink Lego was 96% and 73% using the soft handover and the baseline, respectively.

Concerning the green cup, the effectiveness of the soft handover is less evident as errors in the positioning of the hand resulted in similar failures for both approaches. The object was picked up $19/26$ times using the soft handover, whereas it was grasped $20/26$ times with the baseline.

C. Experiment 3: Soft Handover is Perceived as a More Natural and Trustworthy Strategy

The third experiment focused on evaluating users' perception of the two strategies. Analyzing the results of Fig. 6, we can observe some differences in ease, naturalness, reliability, annoyance, and trust perceived during the handover process. For smaller objects (Golf ball, Marker and Pink Lego), the results indicate the soft handover strategy is significantly easier, more natural, and more trustworthy, while being less annoying (the 58.22% of participants preferred the soft handover, voting 5, 6 or 7, the 29.78% were neutral, whereas only 12% preferred the baseline). For medium-sized objects (Lime, Green cup, and Foam brick), the responses indicate a well-balanced perception between the two handover strategies. The soft handover was preferred by 38.44% of participants, while 30% favored the baseline, and 31.56% were neutral. However, results for the Lime show a statistically significant preference for the soft handover (52%). For larger objects (Orange, Red cup, Box), there is a notable shift towards the baseline strategy, which was preferred by 39.78% of participants. In particular, the Red cup and the Box show a significant preference for the baseline, with 47.33% and 50%, respectively. However, the Orange is an exception, where the soft handover is significantly preferred (48%).

The outcomes of this analysis are also reflected by the success rates observed for different objects. The three smallest objects are grasped more successfully using the soft handover (94.4%) compared to the baseline approach (61.1%). In contrast, the baseline is more effective for larger objects (i.e., Foam brick, Red cup, and Box) with success rates of 97.78% compared to 80% for the soft handover. The exception is the Orange, which achieved a success rate of 90% with the soft handover and 73.3% with the baseline. The overall success rates of both strategies align with Experiment 1 (87.4% soft handover, 79.6% baseline).

Many insights were collected from the open question. The soft handover approach was described as more natural by most of the participants. One of them compared it to human-human interaction: *"The soft approach feels similar to shaking hands with a real person."*. Besides, participants often felt that the soft handover offered a more secure and reliable grasp: *"The soft approach seems much more reliable and secure during the grasp, the other one fails more easily."*. The soft handover strategy was often viewed as robust, particularly when the interaction was felt not excessive. Two participants noted that in one of the conditions the robot purposefully pushed against the human hand, and this was sometimes perceived as annoying, even though the current force level was deemed acceptable.

To evaluate the overall handover process, independently from the adopted grasping strategy, we asked participants to rate six selected items from the RoSAS questionnaire (Fig. 7). Answers ranged from "*Not at all*" to "*Very much so*" and were converted to a 7-point scale (from 0 to 6). High ratings were given for reliability (6 ± 0.6), responsiveness (5 ± 0.9), and capability (4.7 ± 0.8), whereas the discomfort ratings were mostly low. Only awkwardness got higher ratings (1.3 ± 1.1) and this could be due to the occasional issues in the handover process, such as failures or excessive interaction force. The very low ratings for fear (0.36 ± 0.55) and danger (0.26 ± 0.52) are positive indicators, showing that participants felt safe and not threatened by the robotic system during the task.

VI. Conclusion and Future Work

This paper presents a vision-based open-palm human-to-robot handover approach that exploits a soft robotic hand. Using a

single RGB-D camera, the perception pipeline estimates the person's hand palm orientation and reconstructs the object point cloud to plan the handover based on the oriented 3D bounding box.

A systematic evaluation to compare the proposed method to a baseline position-based approach showed that exploiting the explicit human-robot physical interaction during the hand closure increases the grasp success rate and the robustness to uncertainties on object positioning. A user study showed that when the robot-human interaction is more useful and evident (i.e., with small objects), users prefer the soft handover in terms of ease, naturalness, reliability, and trust.

The novel handover strategy proposed in this work assumes that the user's hand remains static once the robot motion starts. Further work is needed to develop a reactive handover system that dynamically adapts to human actions during the task. This would require managing occlusions that occur when the robotic hand is above the human hand. In addition, in this work, the soft handover is implemented assuming an open-palm delivery of the object. Future work will be focused on extending the applicability of the pipeline to different human grasps, still considering the contact with the human hand not as something to avoid, but rather as an aid that can facilitate the robot grasp. For example, different constraint exploitation strategies like the slide-to-edge grasp [9] will be investigated.

REFERENCES

[1] R. Liu, R. Chen, A. Abuduweili, and C. Liu, "Proactive human-robot co-assembly: Leveraging human intention prediction and robust safe control," in *Proc. 2023 IEEE Conf. Control Technol. Appl.*, 2023, pp. 339–345.

[2] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulić, "Object handovers: A review for robotics," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1855–1873, Dec. 2021.

[3] P. Rosenberger et al., "Object-independent human-to-robot handovers using real time robotic vision," *IEEE Robot. Automat. Lett.*, vol. 6, no. 1, pp. 17–23, 2020.

[4] W. Yang, C. Paxton, A. Mousavian, Y.-W. Chao, M. Cakmak, and D. Fox, "Reactive human-to-robot handovers of arbitrary objects," in *Proc. 2021 IEEE Int. Conf. Robot. Automat.*, 2021, pp. 3118–3124.

[5] M. Bianchi et al., "Touch-based grasp primitives for soft hands: Applications to human-to-robot handover tasks and beyond," in *Proc. 2018 IEEE Int. Conf. Robot. Automat.*, 2018, pp. 7794–7801.

[6] M. K. Pan, E. Knoop, M. Bächer, and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," in *Proc. 2019 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 6735–6741.

[7] G. Salvietti, Z. Iqbal, and D. Prattichizzo, "Bilateral haptic collaboration for human-robot cooperative tasks," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 3517–3524, Apr. 2020.

[8] C. Eppner, R. Deimel, J. Alvarez-Ruiz, M. Maertens, and O. Brock, "Exploitation of environmental constraints in human and robotic grasping," *Int. J. Robot. Res.*, vol. 34, pp. 1021–1038, 2015.

[9] J. Bimbo et al., "Exploiting robot hand compliance and environmental constraints for edge grasps," *Front. Robot. AI*, vol. 6, 2019, Art. no. 135.

[10] W. Yang, C. Paxton, M. Cakmak, and D. Fox, "Human grasp classification for reactive human-to-robot handovers," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2020, pp. 11123–11130.

[11] W. Yang et al., "Model predictive control for fluid human-to-robot handovers," in *Proc. 2022 Int. Conf. Robot. Automat.*, 2022, pp. 6956–6962.

[12] V. Micelli, K. Strabala, and S. S. Srinivasa, "Perception and control challenges for effective human-robot handoffs," in *Proc. Robot., Sci. Syst. Workshop RGB-D Cameras*, 2011.

[13] R. Liu, R. Chen, and C. Liu, "Task-agnostic adaptation for safe human-robot handover," *IFAC-PapersOnLine*, vol. 55, no. 41, pp. 175–180, 2022.

[14] S. Christen, W. Yang, C. Pérez-D'Arpino, O. Hilliges, D. Fox, and Y.-W. Chao, "Learning human-to-robot handovers from point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 9654–9664.

[15] H. Duan, P. Wang, Y. Li, D. Li, and W. Wei, "Learning human-to-robot dexterous handovers for anthropomorphic hand," *IEEE Trans. Cogn. Develop. Syst.*, vol. 15, no. 3, pp. 1224–1238, Sep. 2023.

[16] L. Ge, Y. Cai, J. Weng, and J. Yuan, "Hand PointNet: 3D hand pose estimation using point sets," in *Proc. 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8417–8426.

[17] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "Robust 3D hand pose estimation from single depth images using multi-view CNNs," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4422–4436, Sep. 2018.

[18] Y. Hasson et al., "Learning joint reconstruction of hands and manipulated objects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11799–11808.

[19] S. Hampali, M. Rad, M. Oberweger, and V. Lepetit, "Honnotate: A method for 3D annotation of hand and object poses," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3196–3206.

[20] A. Vakunov, C. Chang, F. Zhang, G. Sung, M. Grundmann, and V. Bazarevsky "Mediapipe hands: On-device real-time hand tracking," 2020, *arXiv:2006.10214*.

[21] J. Docekal, J. Rozlivek, J. Matas, and M. Hoffmann, "Human keypoint detection for close proximity human-robot interaction," in *Proc. 2022 IEEE-RAS 21st Int. Conf. Humanoid Robots*, Nov. 2022, pp. 450–457, doi: 10.1109%2Fhumanoids53995.2022.10000133.

[22] I. Söderkvist, "Using SVD for some fitting problems," University Lecture, Tech. Rep., no. 2, pp. 2–5, 2009.

[23] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," Version 8.0.0, Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[24] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," 2018, *arXiv:1801.09847*.

[25] M. Pozzi, S. Marullo, G. Salvietti, J. Bimbo, M. Malvezzi, and D. Prattichizzo, "Hand closure model for planning top grasps with soft robotic hands," *Int. J. Robot. Res.*, vol. 39, no. 14, pp. 1706–1723, 2020.

[26] M. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi, "Adaptive synergies for the design and control of the Pisa/IIT softhand," *Int. J. Robot. Res.*, vol. 33, pp. 768–782, 2014.

[27] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the Yale-CMU-Berkeley object and model set," *IEEE Robot. Automat. Mag.*, vol. 22, no. 3, pp. 36–52, Sep. 2015.

[28] B. S. Homberg, R. K. Katzschmann, M. R. Dogar, and D. Rus, "Robust proprioceptive grasping with a soft robot hand," *Auton. Robots*, vol. 43, pp. 681–696, 2019.

[29] M. K. Pan, E. A. Croft, and G. Niemeyer, "Evaluating social perception of human-to-robot handovers using the robot social attributes scale (ROSAS)," in *Proc. 2018 ACM/IEEE Int. Conf. Hum.-Robot Interaction*, 2018, pp. 443–451.

Open Access funding provided by 'Università degli Studi di Siena' within the CRUI CARE Agreement