2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

A technical integration framework for human-like motion generation in symmetric dual arm picking of large objects

Marco Baracca[1,*], Luca Morello[2] and Matteo Bianchi[1]

***Abstract*— Dual arm robot picking of large objects is a common task in industrial settings, which is often accomplished besides a human operator, as a part of a more complex execution pipeline. This not only requires the simultaneous control of multiple arms to achieve the desired motion of the object and the maintenance of the right amount of force to ensure a stable grasp, but it has also to guarantee a safe and trustworthy human-robot interaction. One way to achieve the latter requirements is to ensure the execution of human-like robot motions, which can be easily understood and predicted by humans. In this paper, we present a technical framework that upon a passivity-based adaptive force-impedance control for modular multi-manual object manipulation, integrating it with a vision-based system to increase the effectiveness and generalizability of the manipulative action, as well as with a human-like Cartesian motion planning algorithm, to enable dual arm picking of large objects. We tested our approach in experiments with real manipulators during different types of large objects picking.**

I. INTRODUCTION

Multi arm robotic manipulation is a key feature to perform cobotics-mediated tasks in industrial settings, such as dual arm picking and handling of large objects, besides and together with the human operator [1]. This is a multi-faceted problem, which requires not only (1) to manage the coordination and the interaction with the environment of multiple manipulators [2], [3], but also (2) to ensure that such interaction is performed in unstructured environment, in a way that is safe and predictable for humans. For what concerns point (1), which deals with moving the object in the desired configuration and holding it firmly during the motion, different solutions have been presented in literature, ranging from the definition of a low-level control law to the design of a high-level motion planning algorithm [4]. Among the most recent results, it is worth mentioning [5], where the authors proposed a motion generation method for dual arm fast grabbing and tossing of boxes, and [6], where a framework for bimanual grasping of large momentum objects was developed. However, despite the significant theoretical value, these methods were usually deployed in controlled environments, where the model of the object is known in advance, and the system can have access to precise information regarding the position and orientation of the object. In [7] the authors developed a modular hybrid force/impedance control suitable for multi-arm manipulation, which can be applied to a system with an arbitrary number of manipulators deployed in a general configuration. However, even though the authors provided strong results from a control point of view, the assumptions made in this work in terms of knowledge of the geometry of the object and the contact point locations prevent an easy implementation in real unstructured scenarios.

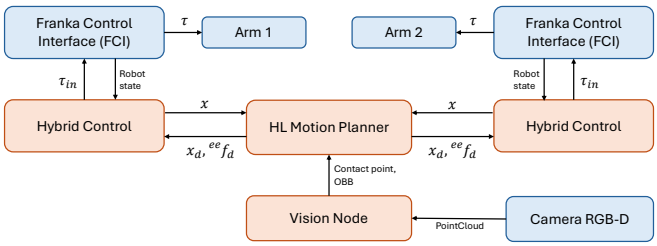

Fig. 1. General scheme of the ROS framework. In orange the blocks represent the different ROS nodes implementing the different parts presented in our work while in blue the blocks represent the external hardware.

For what concerns point (2), a key enabler for the deployment of multi-arm manipulator systems besides humans is safety. One of the solutions to achieve this goal is to ensure that robot movements can be easily predictable by the human operator, i.e. the movements should embed and exhibit human-like characteristics [8]. In [9], the authors analyzed human motion during dual arm picking, using Principal Component Analysis, and exploited these results for motion generation and control, targeting dual arm robotic object manipulation. However, this method is based on the definition of specific points of interest in the human kinematic chain (e.g. shoulder, the tip of the thumb, middle and little fingers), whose translation to generic robotic kinematic chains is not straightforward, limiting the number of artificial systems to which it can be applied. Furthermore, the method requires recording two specific human datasets (one for the reaching task and one for the dual arm manipulation) to extract the needed Principal Component representation.

A possible approach to ensure human-like motions of manipulators that can be kinematically dissimilar from the human example is the one that we proposed in [10]. More specifically, we leveraged a representation of the human hand motion based on functional Principal Component Analysis to design a motion planning algorithm in the Cartesian domain. However, the approach was designed for single-arm systems and a strategy to extend it to multi-arm manipulation is still missing.

In this work, we present a technical integration framework implemented in Robot Operating System (ROS), which takes

This work was supported by European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 101017274 (DARKO); the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab and FoReLab project (Departments of Excellence).

[1]Research Center E. Piaggio and Department of Information Engineering, University of Pisa, Pisa, Italy; [2]Department of Mechanical Engineering, KU Leuven, Ghent, Belgium. *Corresponding author: marco.baracca@phd.unipi.it

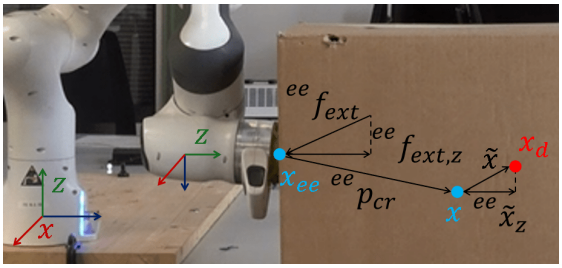

Fig. 2. Scheme representing the formulation of the modular control for dual-arm manipulation. This figure shows the two main parts of the control law: 1) the mapping between the object's centroid position x and the end-effector position x_{ee}; and 2) the force exchanged between the object and the manipulator.

advantage from our human-like motion generation approach in [10] and the outcomes from [7], for dual-arm grasping and picking of large boxes. To enable a proper integration of the two methods, we exploited the fact that the reference motion in [7] is defined in the Cartesian domain and it permits the integration of a large number of planners without requiring particular adaptation of the motion planned to the kinematic of the system. The framework is complemented with a vision component, which exploits the data of an RGB-D camera to estimate the position and dimensions of the box to be manipulated and infer the possible contact point location to perform grasping (relaxing part of the hypothesis of the original work and improving the deployment capability in unstructured environments).

An overview of the proposed integration framework is presented in Fig. 1. The manuscript is structured as follows: first, we provide a brief explanation from a theoretical point of view of the single parts composing the overall motion and control framework, and a description of the ROS technical integration framework. Afterwards, we describe the experimental setup we developed to test the framework, reporting and discussing the output of the experiments. Finally, we discuss the future steps to be addressed.

II. METHODS

In this section, we report a description of the different building blocks composing the framework. First, we start describing the hybrid force/impedance control scheme presented in [7], which is the base upon which we develop the entire system. The two main reasons behind this choice are: 1) the modularity nature of its design permits it to be applied to an arbitrary number of manipulators without increasing the complexity and 2) its definition in the Cartesian domain permits easy integration with generic planning algorithms without any requirement to adapt them to the kinematic structure of the system. After that, we move to describe the human-like motion planning algorithm used to compute the reference motion to be fed to the control law. In the end, we describe the vision pipeline implemented to make the system able to perceive the position and the dimension of the box to be picked.

A. Modular Multi-arm Control

In this section, we will briefly report the structure of the hybrid force/impedance control framework applied in this framework. For further detail regarding the theory behind this approach we refer the interested reader to [7].

The main concept behind this control approach is to define a single point of interest $\boldsymbol{x}$ of the object and use it as a reference for the controller of each manipulator involved. Assuming that the manipulators are in a stable contact configuration with the box, we can define the vector ${}^{ee}\boldsymbol{p}_{cr}$ to express the relative position of the point $\boldsymbol{x}$ with respect to the end effector position $\boldsymbol{x}_{ee}$ (see Fig. 2). With this information, we can compute the Jacobian matrix $\boldsymbol{J}_{cr} \in \mathbb{R}^{6\times6}$ as:

$$\boldsymbol{J}_{cr} = \begin{bmatrix} \boldsymbol{I}_3 & [{}^{ee}\boldsymbol{p}_{cr}]^T_\times \\ \boldsymbol{0}_3 & \boldsymbol{I}_3 \end{bmatrix}, \tag{1}$$

where $[\cdot]_\times$ is the skew-symmetric operator. In this way, the manipulator joint velocities $\dot{\boldsymbol{q}} \in \mathbb{R}^n$ can be mapped onto the Cartesian velocity of the desired object point of interest through:

$$\dot{\boldsymbol{x}} = \boldsymbol{J}_{cr}\boldsymbol{J}_{ee}(\boldsymbol{q})\dot{\boldsymbol{q}} = \boldsymbol{J}(\boldsymbol{q})\dot{\boldsymbol{q}}. \tag{2}$$

At last, the dynamics of the manipulator in Cartesian space can be defined with respect to $\boldsymbol{x}$ in the following way:

$$\boldsymbol{M}_C(\boldsymbol{q})\ddot{\boldsymbol{x}} + \boldsymbol{C}_C(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{x}} + \boldsymbol{f}_g(\boldsymbol{q}) = \boldsymbol{J}^{-T}(\boldsymbol{q})\boldsymbol{\tau}_{in} + \boldsymbol{J}_{cr}^{-T}\boldsymbol{w}_{ext}, \tag{3}$$

where $\boldsymbol{M}_C(\boldsymbol{q})$, $\boldsymbol{C}_C(\boldsymbol{q},\dot{\boldsymbol{q}}) \in \mathbb{R}^{6\times6}$ and $\boldsymbol{f}_g(q) \in \mathbb{R}^6$ are the inertia, Coriolis and gravitational terms reported in the Cartesian domain, $\boldsymbol{\tau}_{in} \in \mathbb{R}^n$ is the torque command and $\boldsymbol{w}_{ext} \in \mathbb{R}^6$ is the external wrench applied at the robot end-effector. Please note that if the Jacobian matrix $\boldsymbol{J}(\boldsymbol{q}) \in \mathbb{R}^{6\times n}$ is not invertible, its pseudoinverse can be used.

Moving to the definition of the torque control input, during non-prehensile multi-arm manipulation, the controller has two main objectives: 1) apply the desired force ${}^{ee}f_d$ on the surface of the object to ensure a stable grasp and 2) move the object to the desired position $\boldsymbol{x}_d$. To fulfil both tasks, a hybrid force/impedance controller was used. The control torque can be expressed as $\boldsymbol{\tau}_{in} = \boldsymbol{\tau}_{imp} + \boldsymbol{\tau}_{frc}$ where $\boldsymbol{\tau}_{imp}$, $\boldsymbol{\tau}_{frc} \in \mathbb{R}^n$ stand for the input torques for motion control and contact force control, respectively.

1) Variables Impedance Control: To avoid the generation of high-intensity forces during the interaction with the environment, an impedance control structure was used for motion control. The control law can be defined at the center of rotation of the object with the following equation:

$$\boldsymbol{\tau}_{imp} = \boldsymbol{J}^T(\boldsymbol{q})(\boldsymbol{K}_C\tilde{\boldsymbol{x}} + \boldsymbol{D}_C\dot{\tilde{\boldsymbol{x}}} + \boldsymbol{M}_C(\boldsymbol{q})\ddot{\boldsymbol{x}}_d + \boldsymbol{C}_C(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{x}}_d) \tag{4}$$

with $\tilde{\boldsymbol{x}} = \boldsymbol{x}_d - \boldsymbol{x}$. In the above formulation, the stiffness and damping matrices are represented by $\boldsymbol{K}_C$, $\boldsymbol{D}_C \in \mathbb{R}^{6\times6}$, and $\boldsymbol{x}_d$ is the desired object pose expressed to the robot frame.

To address the conflict between desired motion and force, in [7] the authors proposed an adapting stiffness matrix ${}^{ee}\boldsymbol{K}_C \in \mathbb{R}^{6\times6}$ defined in the robot end-effector frame. The following is a description of the adaptive stiffness policy:

$${}^{ee}\boldsymbol{K}_C = diag([k_{t,x}, k_{t,y}, \rho_{imp}k_{t,z}, k_{r,x}, k_{r,y}, k_{r,z}]) \tag{5}$$

$$\rho_{imp} = \begin{cases} 1 & \text{If} \quad \delta_{imp} \leq {}^{ee}\tilde{x}_z \\ 0.5(1-\cos(\pi\frac{{}^{ee}\tilde{x}_z}{\delta_{imp}})) & \text{If} \quad 0 \leq {}^{ee}\tilde{x}_z < \delta_{imp} \\ 0 & \text{Otherwise} \end{cases} \tag{6}$$

The parameters $k_i \, \mathbb{R}_{\geq 0}$ represent the default stiffness coefficients for translation and rotation, while ${}^{ee}\tilde{x}_z$ define the pose error along the end-effector's z-axis frame. The distance δ_{imp} denotes the threshold for stiffness adaptation along the z-direction of the end-effector. When desired motion conflicts with a secure grasp, this adaptation prioritizes contact forces, hence the overall stability.

2) Force control: Each manipulator must apply a contact force ${}^{ee}f_d$ along the z-direction of the end-effector frame when performing object-grabbing manipulation tasks. The contact force operating along the z-direction of the end-effector is represented by ${}^{ee}f_{ext,z} \in \mathbb{R}$ in the force control law. As an outcome, the following represents the expression of the proposed control law:

$$\boldsymbol{\tau}_{frc} = \boldsymbol{J}_{ee}^{T}(q)[{}^{ee}\boldsymbol{R}, \boldsymbol{0}_{3\times 3}]^{T}[0,0,\rho_{frc}{}^{ee}f_{frc}]^{T} \tag{7}$$

$${}^{ee}f_{frc} = {}^{ee}f_d + k_p{}^{ee}\tilde{f}_{ext} + k_i\int {}^{ee}\tilde{f}_{ext}dt + k_d{}^{ee}\dot{\tilde{f}}_{ext} \tag{8}$$

where $\tilde{f}_{ext} = {}^{ee}f_d + {}^{ee}f_{ext,z}$. The force controller output f_{frc} is transformed from the end-effector frame ${}^{ee}f_{frc}$ to the robot frame. The controller's behaviour is shaped by the proportional, integral, and derivative gains k_p, k_i, and k_d. When the manipulator deviates significantly from the set-point in the z-direction of the end-effector's frame, a control variable labelled ρ_{frc} as described below disables the force controller, preventing unwanted motion, particularly in contact loss scenarios.

$$\rho_{frc} = \begin{cases} 0 & \text{if} \quad 2\delta_{frc} \leq |{}^{ee}\tilde{x}_z| \\ 0.5(1+\cos(\pi(\frac{{}^{ee}\tilde{x}_z}{\delta_{frc}}-1))) & \text{if} \quad \delta_{frc} \leq |{}^{ee}\tilde{x}_z| < 2\delta_{frc} \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

with $\delta_{frc} \in \mathbb{R}_{>0}$ being the threshold at which the force controller gets disabled.

B. External Force Estimation

To permit to our framework to use force control, the value of the force applied by the robot on the grasped object is needed. The easiest way to gather this information is to introduce a Force/Torque sensor at the contact point and directly measure the interaction force. However, reliable F/T sensors are usually very expensive and fragile, and the complexity of the entire setup would increase dramatically.

In [11], a set of solutions to estimate external torque applied to a manipulator exploiting only proprioceptive information and the knowledge of the dynamic model of the robot are presented. Among these, in our work, we decided to use the formulation based on the generalized momentum observer which does not require joint acceleration and the inversion of the inertia matrix. For the sake of brevity, we report here only the final formulation of the estimated joint torque $\boldsymbol{r}(t) \in \mathbb{R}^n$, referring the reader to the previously cited paper for further details:

$$\boldsymbol{r}(t) = \boldsymbol{K}_O\left(\boldsymbol{M}(\boldsymbol{q})\dot{\boldsymbol{q}} - \int_0^t (\boldsymbol{\tau}_m + \hat{\boldsymbol{\beta}}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{r})ds - \boldsymbol{p}(0)\right), \tag{10}$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{q},\dot{\boldsymbol{q}}) = \hat{\boldsymbol{g}}(\boldsymbol{q}) - \hat{\boldsymbol{C}}^{T}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}$.

Under the assumption that the manipulator configuration is not a kinematic singularity and knowing the location of the contact point, which in our case corresponds to the tip of the end-effector, we can compute the estimation of the generalized external force as $\hat{\boldsymbol{w}}_{ext} = (\boldsymbol{J}_{ee}(\boldsymbol{q})^{T})^{+}\boldsymbol{r}$, where $+$ denotes the generalized pseudoinverse [12].

C. Human-like Motion Planning Algorithm

One of the main goals of this work is to enrich the framework in [7] by enabling the generation of human-like motions during dual arm-picking. This should likely favour the deployment of the manipulator systems besides humans. To this aim, we exploited the results of our previous work [10] for the implementation of the motion planning algorithm of the overall framework. The main reason behind this choice is that our approach is defined in the Cartesian space, permitting an easy integration with the control framework reported in [7].

To quickly summarise the theoretical base of the planner, we exploited functional Principal Components Analysis (fPCA) to extract the main characteristics of human upper limb motions and we embedded them in a motion planning algorithm. For the sake of space in the following, we report only a brief explanation of the approach while, for a more detailed explanation, we refer the interested reader to [13] for the results of fPCA on the Cartesian human hand motion and to [10] for the details of the motion planning algorithm.

Taking into account the single DoF, the fPCA permits the reconstruction of a generic motion $x(t)$ as a weighted sum of functional Principal Components (fPCs) previously extracted analysing a dataset of recorded movement as $x(t) \approx \bar{x} + S_0(t) + \sum_{i=1}^{s_{max}} \alpha_i S_i(t)$, where $\bar{x}$ denotes the average pose of the hand, $S_0(t)$ is the average trajectory observed over all trajectories in the dataset, α_i is the weight associated to the i-th basis element $S_i(t)$ and s_{max} is the number of components used. The peculiarity of the basis of functions obtained through fPCA is that is ordered in terms of the explained variance that each element accounts for. In this way, the minimum number of element can be used to achieve a certain level of representation of the original dataset.

Starting from this result, if we have a set of constraints for our desired trajectory (for example initial and final position, velocity and acceleration), we can define an equation system to find the coefficients $\bar{x}$ and α_i as

$$\begin{bmatrix} 1 & S_1(t_0) & \dots & S_5(t_0) \\ 1 & S_1(t_f) & \dots & S_5(t_f) \\ 0 & \dot{S}_1(t_0) & \dots & \dot{S}_5(t_0) \\ 0 & \dot{S}_1(t_f) & \dots & \dot{S}_5(t_f) \\ 0 & \ddot{S}_1(t_0) & \dots & \ddot{S}_5(t_0) \\ 0 & \ddot{S}_1(t_f) & \dots & \ddot{S}_5(t_f) \end{bmatrix} \begin{bmatrix} \bar{x} \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} = \begin{bmatrix} x(t_0) - S_0(t_0) \\ x(t_f) - S_0(t_f) \\ \dot{x}(t_0) - \dot{S}_0(t_0) \\ \dot{x}(t_f) - \dot{S}_0(t_f) \\ \ddot{x}(t_0) - \ddot{S}_0(t_0) \\ \ddot{x}(t_f) - \ddot{S}_0(t_f) \end{bmatrix}. \tag{11}$$

With the obtained weights, the desired trajectory can be computed by exploiting the weighted sum defined in the fPCA formulation to compute the desired motion as $x(t) = \bar{x} + S_0(t) + \sum_{i=1}^{5} \alpha_i S_i(t)$. Given the Cartesian definition of this planning algorithm, inside the proposed framework can be used both for the planning of the single-arm reaching motion

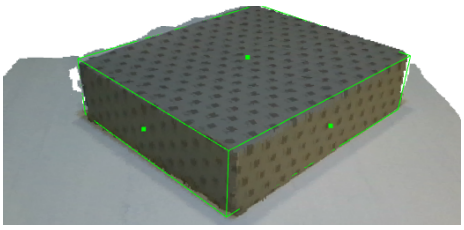

Fig. 3. An example of the output of the vision pipeline proposed.

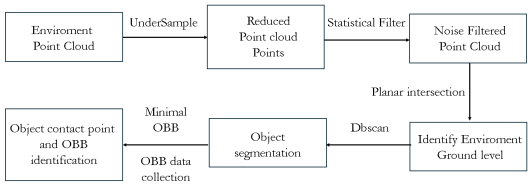

Fig. 4. Block diagram of the visual pipeline

and the desired motion of the box in the multiarm control scheme chosen.

D. Vision Pipeline

To accomplish the manipulation task, robots need to be able to make contact with the object. This calls for precise information about the Cartesian positions of the contact points. In the literature, there are a large number of different approaches for object pose estimations ranging from standard image analysis to deep learning techniques [14]. In this section, we will describe the solution implemented to test the framework. However, it is important to underline that any method capable of returning a bounding box of the object to be grasped can be used instead of the proposed one

The first step involves uniform downsampling, resulting in lower computational costs while preserving the point cloud's essential features. A statistical filter is then used to correct distortions caused by noise and imperfect segmentation. The mean distance between a set of points is determined and compared to a threshold based on the standard deviation from the mean of the average distances to remove outliers. The implementation of this type of filtering is motivated by the incomplete knowledge of the object's surface. The sampled parts exhibit common defects caused by the real sensor, namely imperfect homogeneity in point density and the presence of ripples on the object's surface. The filter in question, considering an average distance between points, proves to be more efficient in areas where point scarcity is lower, such as at the edges of the sensor's known surface. In these areas, not only is the point scarcity lower, but there is also more noise due to the imperfect segmentation of the object. After that, we apply a density-based clustering algorithm, which is especially useful for identifying clusters of arbitrary shapes. DBSCAN [15] has significant advantages over methods such as k-means because it can identify clusters of various shapes and distinguish noise points from clusters. Finally, the minimum bounding box containing the cluster obtained in the previous step is defined. This allows us to easily determine the centroid and, using the object's dimensions, the required contact points to perform dual arm grasping. In Fig. 3 we show an example of the results obtained with our vision pipeline while in Fig. 4 a schematic representation of the pipeline described above is depicted.

E. ROS Integration Framework

For the integration of the different parts, we exploited as base ROS Noetic. More precisely, we developed a set of ROS nodes implementing the different blocks and then we put them in communication using ROS topics. In Fig. 1 we can observe an overall scheme representing how the different parts are connected. In our framework 3 different types of nodes can be identified:

1) *Vision Node*: it takes as input the point cloud from the camera and, exploiting the method described in II-D, returns the Oriented Bounding Box (OBB) and the desired contact point for each manipulator.

2) *HL Motion Planner*: it takes the information provided by the vision node and the actual end-effector pose of the manipulators to compute the desired trajectory using the algorithm described in II-C. After that, it manages the sending of the pose and force references to the hybrid control nodes of each manipulator.

3) *Hybrid Control*: it receives as input the desired pose and force from the planner and the actual state of the robot from the Franka Control Interface (FCI) and computes the desired torque command following the law described in II-A. This command is sent to the FCI which manages the low-level control of the hardware. The force estimation algorithm presented in II-B is integrated in this node to avoid any delay between these two parts.

III. EXPERIMENTAL VALIDATION

A. Experimental Setup

The experimental validation consists of a set of pick-up actions of a large box (40.5x42x31cm, 0.780 Kg). With respect to [7], where the tests started with the object already grasped, in our experimental validation the system also performs the reaching motion from the starting configuration to the desired contact position estimated through the camera. This was done to test the capability of the proposed framework to establish a firm grasp even not knowing the position and dimension of the box in advance. The two main assumptions made are: 1) the box has a homogenous density, which permits us to approximate the centroid found with the vision with the centre of mass of the object; and 2) we know in advance the force required to perform a non-prehensile picking, which depends on the weight of the object and the friction between its surface and the end-effectors of the robots.

The experimental setup includes two Franka manipulators positioned with a displacement of $1.3m$ along the local frames' y-axis, while sharing the same local orientation. A soft-rubber hemispherical tip (similarly to [7]) is mounted at the end-effector of both robots. This helps the robot to achieve soft contact with the object and increase the friction with the box surface, preventing it from overtaking the force limits of the manipulators. The control parameters used for these tests are $k_{t,x},\ k_{t,y},\ k_{t,z} = 600N/m$, $k_{r,x},\ k_{r,y},\ k_{r,z} = 20N/rad$, $\delta_{imp} = 1cm$, $\delta_{frc} = 5cm$, ${}^{ee}f_d =$

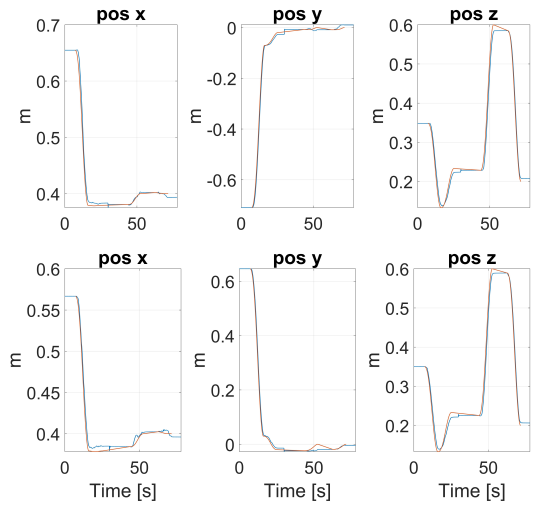

Fig. 5. Comparison of the manipulator's actual trajectories (blue) and the desired ones generated by the human-like planner (orange). The top row represents Arm 1 and the bottom row Arm 2

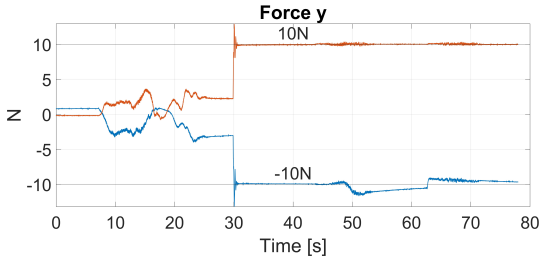

Fig. 6. The estimated external forces at the tip of the end-effector are shown in the global frame (Arm 1 in blue and Arm 2 in orange).

$10N$, $[k_p,\ k_i,\ k_d]^T = [1.5,\ 0.3,\ 0]^T$ and D_C is chosen to achieve critical damping behaviour. To implement the vision part of the pipeline we used an Intel® RealSense™ Depth Camera D415 placed behind the two manipulators. We calibrated the relative pose between the camera and the robots through a custom procedure exploiting the AprilTag library.

B. Pick-up Results

A total of 18 pick-up tests were conducted with different initial conditions. Out of these, 17 experiments yielded successful outcomes, achieving the task of picking up the box without any issues related to contact loss. In the one remaining experiment, the result was unsuccessful during the reaching phase due to the joint limits of the manipulators. A set of snapshots depicting one of the tests performed can be found in Fig. 7.

To evaluate the trajectory tracking precision, we computed the Root Mean Square Error for each trial performed. The means and the standard deviations over all the tests performed for each arm are respectively $0.022 \pm 5.44 \cdot 10^{-04} m$ and $0.022 \pm 5.47 \cdot 10^{-04} m$. In Fig. 5 an example of the trajectory performed is depicted.

In Fig. 6, we can see the estimated external force at the end-effector. From this plot, we can observe that after the first part during the reaching motion where the classical impedance control is activated and the sensed forces are related mainly to joint friction not represented in the dynamic model, the hybrid control is activated and is able to set the applied force to the desired one. What happens around $t = 50s$ is worthy of interest and deserves a more in-depth discussion. We can observe a deviation from the desired force of arm 1. This is caused by a position error of arm 2 which tries to push against the box. However, the system can compensate for this error and maintain a firm grasp of the object.

C. Human-likeness

After evaluating the effectiveness of the proposed approach in accomplishing the task, we move to evaluate the Human-Likeness of the produced motion. Several peculiar features were identified in human motion, such as arm postures and kinematic temporal behaviour [16]. However, most of these metrics are mainly connected to an anthropomorphic kinematic structure. For this reason, we focus on evaluating the jerk of the motion produced by our framework. In fact, in the literature, the minimum jerk behaviour of human motions was extensively proven [17].

The averages of the median jerk at the Cartesian level obtained for each arm are respectively $(1.02 \pm 0.05) \cdot 10^{-05} m/s^3$ and $(1.19 \pm 0.03) \cdot 10^{-05} m/s^3$ (for the planned trajectory is $(0.19 \pm 0.03) \cdot 10^{-05} m/s^3$). We can observe that, even in the presence of a hybrid control which manages the interaction between the two robots, we achieve a low level of jerk in the resultant motion of the two arms. Furthermore, even though the tasks involved are different, the jerk obtained is comparable with previous results [18].

IV. LIMITATION, CONCLUSIONS AND FUTURE WORKS

In this work we propose a ROS framework that enables to integrate the human-like motion generation in [10] and the modular hybrid force/impedance control in [7], for the autonomous pick-up of large boxes. The target scenario is dual-arm robotic manipulation in unstructured environments, besides and together with human operators, to ensure not only the effectiveness, but also the safety and predictability of the motion for humans. The framework integrates also a vision layer, which exploits data from an RGB-D camera to estimate the position and dimensions of the box to be manipulated and infer the possible contact point location to perform grasping, relaxing the hypothesis in [7] about the knowledge of the location and geometry of the object. We tested the framework with real manipulators, proving its capability to accomplish the desired task under reasonable assumptions. Furthermore, we evaluated the human-likeness of the produced movement proving that the system can maintain the desired characteristics. We acknowledge that this is only the first step toward a completely autonomous dual-arm system able to operate in daily living environments, and it should be developed and improved under several aspects.

To this aim, our future effort will be devoted to relax the hypothesis on the knowledge of the force required to hold firmly the object. This is related mainly to the inertia of the

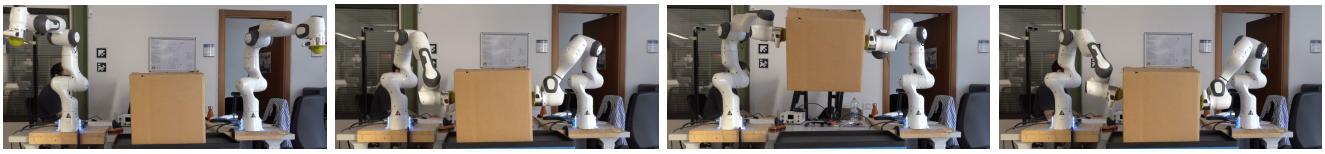

Fig. 7. Snapshot of a pick-up task. The robots start from the initial configuration and reach the estimated contact points provided by the vision layer. Then the framework starts to apply force and lift the box. In the end, the system put down the box in the original position.

object and the friction between the surfaces of the object and the tip of the end effector. For inertial parameters, during the testing of the overall framework, we performed some preliminary evaluation on the feasibility of estimating object mass exploiting the generalized momentum observer [12], showing promising results. However, also the knowledge of the center of mass location and inertial tensor plays an important role in ensuring a stable grasp, also with objects with not homogeneous density. A possible solution for this problem could be found in the active sensing literature [19], where the problem of optimal motion generation to minimize parameter estimation uncertainty is addressed. In this way, the system could perform exploratory movements and use the contact forces gathered through the momentum observer already implemented to estimate the inertial parameters required. Regarding the friction between surfaces instead, an intriguing approach could be the integration of soft optical tactile sensors as the tips of the end effector [20].

Another point to be addressed is the generalization of this approach to objects different from boxes. In this sense, an interesting approach for this step could be [21], where the authors developed a framework capable of generating a feasible grasp for different grippers relying on the decomposition of general shapes into boxes. The extension of our framework to multi-arm (i.e. more than two) object picking is also envisioned, having the possibility of designing systems with higher manipulation capability, and test them in more unstructured environments, e.g. further developing the vision system to deal with cluttered environments [22], and addressing the generation of more complex tasks [23].

REFERENCES

[1] M. Raessa, J. C. Y. Chen, W. Wan, and K. Harada, "Human-in-the-loop robotic manipulation planning for collaborative assembly," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 4, pp. 1800–1813, 2020.

[2] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, "Shared control–based bimanual robot manipulation," *Science Robotics*, vol. 4, no. 30, p. eaaw0955, 2019.

[3] F. Krebs and T. Asfour, "A bimanual manipulation taxonomy," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 031–11 038, 2022.

[4] C. Smith, Y. Karayiannidis, L. Nalpantidis, X. Gratal, P. Qi, D. V. Dimarogonas, and D. Kragic, "Dual arm manipulation—a survey," *Robotics and Autonomous systems*, vol. 60, no. 10, pp. 1340–1353, 2012.

[5] M. Bombile and A. Billard, "Dual-arm control for coordinated fast grabbing and tossing of an object: Proposing a new approach," *IEEE Robotics & Automation Magazine*, vol. 29, no. 3, pp. 127–138, 2022.

[6] L. Yan, T. Stouraitis, J. Moura, W. Xu, M. Gienger, and S. Vijayakumar, "Impact-aware bimanual catching of large-momentum objects," *IEEE Transactions on Robotics*, 2024.

[7] E. Shahriari, S. A. B. Birjandi, and S. Haddadin, "Passivity-based adaptive force-impedance control for modular multi-manual object manipulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2194–2201, 2022.

[8] A. M. Zanchettin, L. Bascetta, and P. Rocco, "Acceptability of robotic manipulators in shared working environments through human-like redundancy resolution," *Applied ergonomics*, vol. 44, no. 6, pp. 982–989, 2013.

[9] S. Y. Shin and C. Kim, "Human-like motion generation and control for humanoid's dual arm object manipulation," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2265–2276, 2014.

[10] M. Baracca, G. Averta, and M. Bianchi, "A general approach for generating artificial human-like motions from functional components of human upper limb movements," *Control Engineering Practice*, vol. 148, p. 105968, 2024.

[11] S. Haddadin, A. De Luca, and A. Albu-Schäffer, "Robot collisions: A survey on detection, isolation, and identification," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1292–1312, 2017.

[12] M. Iskandar, O. Eiberger, A. Albu-Schäffer, A. De Luca, and A. Dietrich, "Collision detection, identification, and localization on the dlr sara robot with sensing redundancy," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3111–3117.

[13] M. Baracca, P. Bonifati, Y. Nisticò, V. Catrambone, G. Valenza, A. Bicchi, G. Averta, and M. Bianchi, "Functional analysis of upper-limb movements in the cartesian domain," in *Converging Clinical and Engineering Research on Neurorehabilitation IV: Proceedings of the 5th International Conference on Neurorehabilitation (ICNR2020), October 13–16, 2020*. Springer, 2022, pp. 339–343.

[14] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[15] L. Ni and S. Jinhang, "The analysis and research of clustering algorithm based on pca," in *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. IEEE, 2017, pp. 361–365.

[16] A. Meixner, M. Carl, F. Krebs, N. Jaquier, and T. Asfour, "Towards unifying human likeness: Evaluating metrics for human-like motion retargeting on bimanual manipulation tasks," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 13 015–13 022.

[17] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *Journal of neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[18] G. Averta, D. Caporale, C. Della Santina, A. Bicchi, and M. Bianchi, "A technical framework for human-like motion generation with autonomous anthropomorphic redundant manipulators," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3853–3859.

[19] N. Mavrakis and R. Stolkin, "Estimation and exploitation of objects' inertial parameters in robotic grasping and manipulation: A survey," *Robotics and Autonomous Systems*, vol. 124, p. 103374, 2020.

[20] W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, and S. J. Redmond, "Tactile sensors for friction estimation and incipient slip detection—toward dexterous robotic manipulation: A review," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9049–9064, 2018.

[21] A. Palleschi, F. Angelini, C. Gabellieri, L. Pallottino, A. Bicchi, M. Garabini *et al.*, "Grasp it like a pro 2.0: A data-driven approach exploiting basic shape decomposition and human data for grasping unknown objects," *IEEE Transactions on Robotics*, 2023.

[22] S. D'Avella, M. Bianchi, A. M. Sundaram, C. A. Avizzano, M. A. Roa, and P. Tripicchio, "The cluttered environment picking benchmark (cepb) for advanced warehouse automation: Evaluating the perception, planning, control, and grasping of manipulation systems," *IEEE Robotics & Automation Magazine*, pp. 2–15, 2023.

[23] J. Gao, X. Jin, F. Krebs, N. Jaquier, and T. Asfour, "Bi-kvil: Keypoints-based visual imitation learning of bimanual manipulation tasks," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 16 850–16 857.