IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

A nonlinear MPC for Physical Human-Aerial Robot Interaction in Collaborative Transportation Tasks

A. Gonzalez-Morgado*, J. Soueidan*, G. Heredia, A. Ollero, P. Fraisse, M. Tognon, M. Cognetti

***Abstract*— Aerial robots are transitioning from traditional surveillance and monitoring roles to more advanced tasks involving physical interaction. Despite this progress, physical Human-Aerial Robot Interaction remains largely underexplored due to the complexity and stability-related issues of such platforms. This paper introduces a novel control framework that enables an aerial platform to cooperatively transport an object with a human operator. The control approach is built on a nonlinear model predictive control (NMPC), integrating the dynamic models of the human, the aerial robot, and the transported object. To ensure safe and robust physical interaction, the NMPC is combined with a compliant controller. Additionally, our controller prioritizes forward motion over lateral movements to accommodate the human's natural direction of motion. We validate this framework through indoor flight experiments, demonstrating how a human operator and a fully actuated hexarotor can effectively collaborate to transport a bar. The results highlight the aerial robot's ability to assist the human during physical transportation tasks, enhancing efficiency and comfort.**

NOMENCLATURE

$\mathcal{F}_{\mathrm{W}}, \mathcal{F}_{*}$	World frame, $*$ frame
$\mathbf{p}_{*}, \mathbf{v}_{*}$	Position and velocity of $*$ w.r.t. $\mathcal{F}_{\mathrm{W}}$
${}^{\mathrm{W}}\mathbf{R}_{*}$	Rotation matrix of $*$ w.r.t. $\mathcal{F}_{\mathrm{W}}$
$\boldsymbol{\eta}_{*}, \boldsymbol{\omega}_{*}$	Euler angles and angular velocity of $*$ w.r.t. $\mathcal{F}_{\mathrm{W}}$
$m_{*}, \mathbf{J}_{*}$	Mass and inertia matrix of $*$
$\mathbf{f}^{\mathrm{I}}_{\mathrm{R}}, \mathbf{f}^{\mathrm{I}}_{\mathrm{H}}$	Interaction force applied on the object by the robot and the human, expressed in $\mathcal{F}_{\mathrm{O}}$
$\mathbf{u}_{\mathrm{H}}, \mathbf{u}_{\mathrm{A}}$	Input of the human, virtual input of the robot
$\mathbf{M}_{\mathrm{A}}, \mathbf{B}_{\mathrm{A}}$	Virtual inertia and damping of the robot
$\mathbf{x}, \mathbf{u}$	State and control vector in the NMPC
$\mathbf{W}_{j}$	Weighting matrix $j \in \{\mathbf{p}_{\mathrm{O}}, \boldsymbol{\eta}_{\mathrm{O}}, \mathbf{v}_{\mathrm{O}}, \boldsymbol{\omega}_{\mathrm{O}}, \mathbf{u}\}$

Where $* \in \{\mathrm{R}, \mathrm{O}, \mathrm{H}\}$, with R: robot, O: object, H: human. Whenever used, the superscript $^{\mathrm{ref}}$ denotes a reference value.

I. INTRODUCTION

Aerial robots are increasingly used not only for contact-free tasks like navigation or surveillance, but also for physical interaction tasks, such as industrial inspection and maintenance [1], precision agriculture [2], and disaster response [3]. They have also been utilized to assist humans in various tasks, thanks to their unlimited workspace and ability to access hard-to-reach areas. They can deliver packages through windows in urban environments [4] or provide tools to operators working in elevated spaces [5], reducing delays and physical strain. Furthermore, they can act as guides, offering force-based communication through a tethered connection with humans to safely guide them toward a goal position [6].

Fig. 1. Collaborative transportation task of an object between a human operator and an aerial robot.

Despite their potential, the use of aerial robots in collaborative transportation tasks with humans remains relatively under-exploited. This is largely due to challenges such as ensuring safe and intuitive human-robot interaction, managing dynamic stability during physical contact, and developing control strategies that accommodate both human intent and environmental uncertainties. In contrast, significant progression has been achieved in other robotic domains, such as humanoid robots, which have already been used to jointly carry a table [7]. Furthermore, other control frameworks propose the use of humanoids in a proactive manner, where human intentions are estimated during the task to enhance efficiency and coordination [8]. These estimations of human intentions are based on extensive research on human locomotion [9]. However, ground robots are limited in terms of traversability and workspace. On the other hand, aerial robots offer the potential to carry out collaborative tasks in environments that are inaccessible or impractical for ground robots, particularly in areas with uneven or cluttered terrain.

In recent years, some works have involved aerial robots for collaborative object transportation with humans. For example, [10] proposed a custom-built sensor to estimate human commands and provide feedback, while [11] introduced a human-multi-robot aerial system for cooperative object transportation tasks. However, even if it offers advantages like transporting heavy payloads, the multi-robot solution often requires complex control strategies and additional constraints, like maintaining cable tension throughout the task. Unlike approaches with humanoid robots, which already incorporate findings from human locomotion studies to improve coordination and intention recognition [9], these aerial robot approaches have not yet fully integrated this knowledge. As a result, there is still significant potential to enhance aerial-human collaboration by integrating insights from human motion studies.

* A. Gonzalez-Morgado and J. Soueidan equally contributed to the paper.

A. Gonzalez-Morgado (corresponding author, e-mail: mantonio@us.es), G. Heredia and A. Ollero are with the GRVC Robotics Lab, Sevilla, Spain. J. Soueidan and M. Cognetti are with the LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France. P. Fraisse is with the LIRMM, Montpellier, France. M. Tognon is with Univ Rennes, Inria, Rennes, France.

A. Contributions and novelties

In this work, we introduce a control framework for collaborative transportation tasks involving humans and aerial robots. Our approach combines a compliant controller and a non-linear model predictive controller (NMPC) to generate optimal control inputs for the robot in order to transport an object to a target location. The proposed framework explicitly considers the dynamics of the human, the aerial robot, and the transported object while respecting actuation limits for both the human and the robot. The compliant controller serves as a safety layer, limiting the interaction forces in case the human actions are different from the ones predicted by the NMPC. To improve human comfort, our approach prioritizes forward motion over lateral movements, inspired by studies on human locomotion [9], [12]. This is achieved by dynamically adjusting the orientation-related weights in the NMPC cost function. In summary, we propose the following contributions:

- A two-level control framework that integrates a compliant controller and an NMPC, enabling effective human-robot collaboration in transportation tasks.
- Experimental validation of the proposed framework using a fully actuated aerial platform.

The remainder of the paper is organized as follows. Section II presents the modeling of the aerial robot, the human, and the transported object. Section III details the control framework developed for cooperative transportation tasks. The implementation of this control framework, along with its experimental validation through indoor flight tests, is discussed in Section IV. Finally, the conclusions of this work and potential future research directions are outlined in Section V.

II. Modeling

Figure 2 represents the transportation task, depicting the geometric quantities used to formulate the problem. The entire system is composed of three sub-systems: (i) the transported object (Sect. II-A); (ii) the aerial manipulator (Sect. II-B); (iii) the human (Sect. II-C). To describe them, we define the following reference frames: a world reference frame, $\mathcal{F}_\mathrm{W}$, with its origin O_W placed arbitrarily, and axes $\mathbf{x}_\mathrm{W}$, $\mathbf{y}_\mathrm{W}$, and $\mathbf{z}_\mathrm{W}$, such that $\mathbf{z}_\mathrm{W}$ is opposite to gravity; a robot frame $\mathcal{F}_\mathrm{R}$, located at the center of mass (CoM) of the aerial robot O_R, with axes $\mathbf{x}_\mathrm{R}$, $\mathbf{y}_\mathrm{R}$, and $\mathbf{z}_\mathrm{R}$; a human frame $\mathcal{F}_\mathrm{H}$, rigidly attached to the transported object at the human's grasping position, O_H, with axes $\mathbf{x}_\mathrm{H}$, $\mathbf{y}_\mathrm{H}$, and $\mathbf{z}_\mathrm{H}$; and an object frame $\mathcal{F}_\mathrm{O}$, located at the CoM of the object O_O, with axes $\mathbf{x}_\mathrm{O}$, $\mathbf{y}_\mathrm{O}$, and $\mathbf{z}_\mathrm{O}$. In addition, $\mathbf{x}_\mathrm{O}$ is parallel to the line connecting O_H and O_R, and points in the same direction as $\mathbf{x}_\mathrm{H}$. Furthermore, $\mathbf{x}_\mathrm{H}$ represents the human heading direction, pointing to O_R, while $\mathbf{z}_\mathrm{H}$ and $\mathbf{z}_\mathrm{O}$ are aligned with gravity. The relevant variables used in this article are summarized in the Nomenclature section.

A. Transported Object

The position and linear velocity of the object with respect to (w.r.t.) the world frame $\mathcal{F}_\mathrm{W}$ are denoted by $\mathbf{p}_\mathrm{O} = [x_O, y_O, z_O]^\top \in \mathbb{R}^3$, and $\mathbf{v}_\mathrm{O} \in \mathbb{R}^3$, respectively. The orientation of the object is represented by the rotation matrix ${}^\mathrm{W}\mathbf{R}_\mathrm{O} \in \mathrm{SO}(3)$, which describes the rotation of $\mathcal{F}_\mathrm{O}$ w.r.t. $\mathcal{F}_\mathrm{W}$. This orientation can also be described using the Euler angles $\boldsymbol{\eta}_\mathrm{O} = [\phi_O, \theta_O, \psi_O]^\top \in \mathbb{R}^3$. The angular velocity of the object w.r.t. $\mathcal{F}_\mathrm{W}$ expressed in $\mathcal{F}_\mathrm{O}$ is defined by $\boldsymbol{\omega}_\mathrm{O} \in \mathbb{R}^3$. We consider that the object is grasped by the human at $\mathbf{b}_\mathrm{H} \in \mathbb{R}^3$ and by the aerial robot at $\mathbf{b}_\mathrm{R} \in \mathbb{R}^3$, both expressed w.r.t. $\mathcal{F}_\mathrm{O}$.

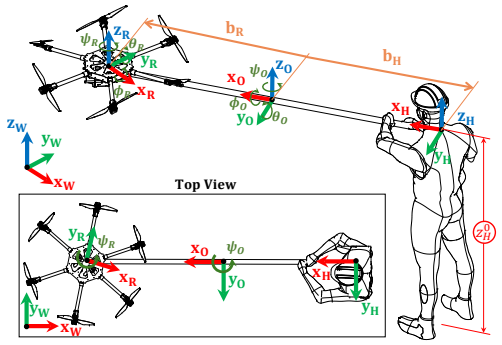

Fig. 2. Visual representation of a collaborative object transportation task between a human and an aerial robot.

Assumption 1. *We assume that both the aerial robot and the human are connected to the transported object through spherical joints. In the case of the aerial robot, a dedicated mechanical joint is implemented, while for the human, the wrist naturally serves this function. This assumption ensures that no interaction torques are transmitted to the object.*

Due to this assumption, only two interaction forces appear: (i) the interaction force exerted by the robot on the object, $\mathbf{f}_\mathrm{R}^\mathrm{I} = [f_\mathrm{R}^{\mathrm{I,x}}, f_\mathrm{R}^{\mathrm{I,y}}, f_\mathrm{R}^{\mathrm{I,z}}]^\top \in \mathbb{R}^3$, and (ii) the interaction force exerted by the human on the object, $\mathbf{f}_\mathrm{H}^\mathrm{I} = [f_\mathrm{H}^{\mathrm{I,x}}, f_\mathrm{H}^{\mathrm{I,y}}, f_\mathrm{H}^{\mathrm{I,z}}]^\top \in \mathbb{R}^3$. Both forces are expressed in $\mathcal{F}_\mathrm{O}$. With these definitions, the dynamic model of the transported object is expressed as:

$$\dot{\mathbf{p}}_\mathrm{O} = \mathbf{v}_\mathrm{O}, \tag{1a}$$

$$\dot{\boldsymbol{\eta}}_\mathrm{O} = \mathbf{T}_\mathrm{O}\boldsymbol{\omega}_\mathrm{O}, \tag{1b}$$

$$m_\mathrm{O}\dot{\mathbf{v}}_\mathrm{O} = {}^\mathrm{W}\mathbf{R}_\mathrm{O}\mathbf{f}_\mathrm{H}^\mathrm{I} + {}^\mathrm{W}\mathbf{R}_\mathrm{O}\mathbf{f}_\mathrm{R}^\mathrm{I} - m_\mathrm{O}g\mathbf{z}_\mathrm{O}, \tag{1c}$$

$$\mathbf{J}_\mathrm{O}\dot{\boldsymbol{\omega}}_\mathrm{O} = \mathbf{b}_\mathrm{R} \times \mathbf{f}_\mathrm{R}^\mathrm{I} + \mathbf{b}_\mathrm{H} \times \mathbf{f}_\mathrm{H}^\mathrm{I} - \boldsymbol{\omega}_\mathrm{O} \times \mathbf{J}_\mathrm{O}\boldsymbol{\omega}_\mathrm{O}, \tag{1d}$$

where $\mathbf{T}_\mathrm{O} \in \mathsf{SO}(3)$ relates the angular velocity $\boldsymbol{\omega}_\mathrm{O}$ with the angle derivatives $\dot{\boldsymbol{\eta}}_\mathrm{O}$, $\mathbf{J}_\mathrm{O} \in \mathbb{R}^{3\times3}$ is the inertia matrix of the object, expressed in $\mathcal{F}_\mathrm{O}$, and $m_\mathrm{O} \in \mathbb{R}^{*+}$ its mass.

B. Aerial Robot

The position of $\mathcal{F}_\mathrm{R}$ w.r.t. $\mathcal{F}_\mathrm{W}$ is denoted as $\mathbf{p}_\mathrm{R} = [x_R, y_R, z_R]^\top \in \mathbb{R}^3$. Its orientation relative to $\mathcal{F}_\mathrm{W}$ is represented by ${}^\mathrm{W}\mathbf{R}_\mathrm{R} \in \mathsf{SO}(3)$, which can also be parameterized using Euler angles $\boldsymbol{\eta}_\mathrm{R} = [\phi_R, \theta_R, \psi_R]^\top \in \mathbb{R}^3$. The linear velocity, expressed in $\mathcal{F}_\mathrm{W}$, is denoted as $\mathbf{v}_\mathrm{R} \in \mathbb{R}^3$, while the angular velocity w.r.t. $\mathcal{F}_\mathrm{W}$, denoted as $\boldsymbol{\omega}_\mathrm{R} \in \mathbb{R}^3$, is expressed in the body frame $\mathcal{F}_\mathrm{R}$. Assuming a fully-actuated vehicle [13], we consider as control inputs the total force $\mathbf{f}_\mathrm{c} \in \mathbb{R}^3$

and torque $\boldsymbol{\tau}_{\mathrm{c}} \in \mathbb{R}^3$ applied by the propellers to the CoM of the aerial robot, both expressed in $\mathcal{F}_{\mathrm{R}}$. With these definitions, the dynamics of the aerial platform can be expressed as:

$$\dot{\mathbf{p}}_{\mathrm{R}} = \mathbf{v}_{\mathrm{R}}, \tag{2a}$$
$$\dot{\boldsymbol{\eta}}_{\mathrm{R}} = \mathbf{T}_{\mathrm{R}} \boldsymbol{\omega}_{\mathrm{R}}, \tag{2b}$$
$$m_{\mathrm{R}} \dot{\mathbf{v}}_{\mathrm{R}} = {}^{\mathrm{W}}\mathbf{R}_{\mathrm{R}} \mathbf{f}_{\mathrm{c}} - m_{\mathrm{R}} g \mathbf{z}_{\mathrm{W}} - {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}} \mathbf{f}_{\mathrm{R}}^{\mathrm{I}}, \tag{2c}$$
$$\mathbf{J}_{\mathrm{R}} \dot{\boldsymbol{\omega}}_{\mathrm{R}} = \boldsymbol{\tau}_{\mathrm{c}} - \boldsymbol{\omega}_{\mathrm{R}} \times \mathbf{J}_{\mathrm{R}} \boldsymbol{\omega}_{\mathrm{R}}, \tag{2d}$$

where the matrix $\mathbf{T}_{\mathrm{R}} \in \mathsf{SO}(3)$ relates the angular velocity $\boldsymbol{\omega}_{\mathrm{R}}$ to the angle derivatives $\dot{\boldsymbol{\eta}}_{\mathrm{R}}$, while $m_{\mathrm{R}} \in \mathbb{R}^{*+}$ represents the robot mass and $\mathbf{J}_{\mathrm{R}} \in \mathbb{R}^{3\times3}$ its inertia matrix in $\mathcal{F}_{\mathrm{R}}$. The transported object is connected to the robot via a spherical joint at its CoM, ensuring that no interaction torques are transmitted between the object and the robot, as shown in Eq. (2d).

C. Human

The human is modeled as a point mass attached to the object, with its heading direction $\mathbf{x}_{\mathrm{H}}$ aligned with $\mathbf{x}_{\mathrm{O}}$. This assumption is motivated by the fact that, during cooperative transportation tasks (such as carrying a table), the human body naturally orients itself toward the direction of motion, as shown in [12]. Furthermore, we denote by $\mathbf{p}_{\mathrm{H}} = [x_H, y_H, z_H]^\top \in \mathbb{R}^3$ the grasping position of the human while $\mathbf{v}_{\mathrm{H}} \in \mathbb{R}^3$ is their linear speed, both expressed in $\mathcal{F}_{\mathrm{W}}$.

Assumption 2. *We consider that the human grasping position is constrained to a constant height z_H^0. This is because, during a cooperative transportation task, height variations are unlikely unless obstacles or other factors require them. The consideration of such factors, including obstacle avoidance or environmental interactions, is beyond the scope of this paper.*

Thus, the human model can be defined as:

$$\dot{\mathbf{p}}_{\mathrm{H}} = \mathbf{v}_{\mathrm{H}}, \tag{3a}$$
$$m_{\mathrm{H}} \dot{\mathbf{v}}_{\mathrm{H}} = {}^{\mathrm{W}}\mathbf{R}_{\mathrm{H}} \mathbf{u}_{\mathrm{H}} - m_{\mathrm{H}} g \mathbf{z}_{\mathrm{H}} - {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}} \mathbf{f}_{\mathrm{H}}^{\mathrm{I}}, \tag{3b}$$
$$z_H = z_H^0, \tag{3c}$$

where $m_{\mathrm{H}} \in \mathbb{R}^+$ is the mass of the human. Additionally, $\mathbf{u}_{\mathrm{H}} = [u_{H_x}, u_{H_y}, u_{H_z}]^\top \in \mathbb{R}^3$ represents the human input expressed in the human reference frame $\mathcal{F}_{\mathrm{H}}$.

III. Control framework

The proposed control framework enables the aerial robot to perform a cooperative transportation task with the human operator. The objective of the cooperative transportation task is to reach a final desired state for the object, described by $\mathbf{x}^{\mathrm{ref}} = [{\mathbf{p}_{\mathrm{O}}^{\mathrm{ref}}}^\top, {\boldsymbol{\eta}_{\mathrm{O}}^{\mathrm{ref}}}^\top, {\mathbf{v}_{\mathrm{O}}^{\mathrm{ref}}}^\top, {\boldsymbol{\omega}_{\mathrm{O}}^{\mathrm{ref}}}^\top]^\top \in \mathbb{R}^{12}$. In particular, we consider rest-to-rest motions, where both the initial and final velocities are zero. This assumption does not limit the generality of our approach.

We propose a framework that is composed of two main components: (i) an inner-level compliant controller (see Sec. III-A), which simplifies the aerial platform's dynamics to that of a mass-damper system by effectively canceling its dynamics; and (ii) an outer-loop NMPC (see Sec. III-B), responsible for online human-aware motion planning.

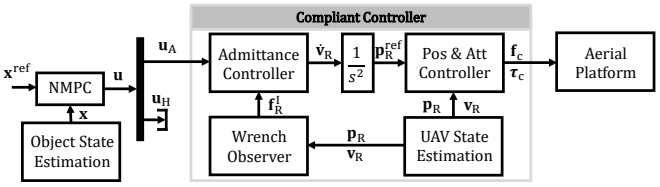

Fig. 3. Control framework of the aerial manipulator for the collaborative task. The human control action $\mathbf{u}_{\mathrm{H}}$ is not used by any inner controller.

Assumption 3. *The proposed NMPC determines both the inputs for the compliant controller and the human that are optimal according to a cost function, while respecting some input/state constraints. Clearly, human inputs are not as controllable as those of the robot. However, this is done to predict the behavior of an ideal collaborative human. We assume that the human operator is aware of the task objectives and that he/she tries to solve a similar optimization problem, where the cost is similar to the one derived in the MPC. This results in human inputs similar to the human control actions determined by the NMPC. In practice, this is facilitated by informing the human about the desired object state $\mathbf{x}^{\mathrm{ref}}$.*

It is important to note that Assumption 3 is not particularly restrictive. In fact, the robot's compliant behavior enhances robustness to discrepancies between predicted and actual human actions. The compliant controller limits the interaction forces with the human, contributing to safe behavior even in the presence of trajectory mismatches. Moreover, the system remains safe in practice, as the control input is actively bounded by the high-level NMPC. Additionally, the iterative nature of NMPC allows it to adapt continuously to human behavior at each control step.

A. Compliant controller

For collaborative tasks with humans, the preferred control methods found in the literature are either *impedance control* [14](which commands forces based on position deviations) or *admittance control* [6] (which commands a position in response to external forces). These methods cancel the robot dynamics, imposing a desired mass-damper behavior for the robot. The resulting closed-loop robot dynamics are (see, e.g., [15] for a formal derivation):

$$\dot{\mathbf{p}}_{\mathrm{R}} = \mathbf{v}_{\mathrm{R}}, \tag{4a}$$
$$\dot{\mathbf{v}}_{\mathrm{R}} = \mathbf{M}_{\mathrm{A}}^{-1}(-\mathbf{B}_{\mathrm{A}} \mathbf{v}_{\mathrm{R}} - {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}} \mathbf{f}_{\mathrm{R}}^{\mathrm{I}} + \mathbf{u}_{\mathrm{A}}), \tag{4b}$$

where $\mathbf{M}_{\mathrm{A}}, \mathbf{B}_{\mathrm{A}} \in \mathbb{R}^{3\times3}$ are positive definite diagonal matrices representing the robot's virtual inertia and damping, respectively. Furthermore, the term $-{}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}} \mathbf{f}_{\mathrm{R}}^{\mathrm{I}}$ is the force felt by the robot, expressed in $\mathcal{F}_{\mathrm{W}}$, and $\mathbf{u}_{\mathrm{A}} = [u_{\mathrm{A}}^{\mathrm{x}}, u_{\mathrm{A}}^{\mathrm{y}}, u_{\mathrm{A}}^{\mathrm{z}}]^\top \in \mathbb{R}^3$ is the virtual input of the mass-damper system, which represents the force this system applies to the object in $\mathcal{F}_{\mathrm{W}}$. Note that there is no stiffness term in this model, since no desired position is defined in the admittance filter. The

apparent stiffness is instead introduced by the NMPC, which generates the desired force based on a desired state.

In this paper, we chose an admittance controller as we wish to control the robot position during the task. Its block diagram is illustrated in Fig. 3. The controller implements Eq. (4) taking as inputs $\mathbf{u}_\mathrm{A}$ and the interaction force $\mathbf{f}_\mathrm{R}^\mathrm{I}$ and generating a reference acceleration $\dot{\mathbf{v}}_\mathrm{R}$ for the robot. Specifically, $\mathbf{u}_\mathrm{A}$ is computed by the NMPC (see Sec. III-B), which is responsible for generating the robot's control actions to accomplish the transportation task. Moreover, the interaction force $\mathbf{f}_\mathrm{R}^\mathrm{I}$ can be obtained either through direct measurement using a force sensor or, as in our case, by estimation using a model-based observer, as demonstrated in [16]. The admittance controller's output, $\dot{\mathbf{v}}_\mathrm{R}$, is double-integrated to produce the reference position trajectory $\mathbf{p}_\mathrm{R}^\mathrm{ref}$ for the robot. This trajectory is then tracked by a low-level controller, which also utilizes the estimated robot's current position and velocity provided by a state estimator. In our case, the low-level controller is the one described in [17], while the state estimator is a standard Unscented Kalman Filter. Note that the admittance controller is only applied to the translational motion, since the spherical joint prevents the transmission of interaction torques (see Assumption 1). Therefore, attitude compliance is not needed because no interaction torques are applied to the robot.

B. NMPC for cooperative transportation

To compute the optimal control input for the robot ($\mathbf{u}_\mathrm{A}$), we employ an NMPC method, where the input for the human ($\mathbf{u}_\mathrm{H}$) is also considered, to find the best prediction of the human motion (see Assumption 3). In the following, we describe the components of the NMPC in detail.

1) Optimal Control Problem: Our NMPC is formulated as a discrete-time Optimal Control Problem (OCP), discretized over N shooting points. The controller solves the OCP at each sampling instant t^*, considering a prediction horizon of duration T_H. As this OCP is solved online at every control cycle, it allows the re-plan of the robot control actions, thereby reducing the impact of potential inaccuracies between the human control actions and the computed ones. Formally, our NMPC solves the following OCP at each sampling instant t^*:

$$\min_{\substack{\mathbf{x}_0 \dots \mathbf{x}_N \\ \mathbf{u}_0 \dots \mathbf{u}_{N-1}}} J(\mathbf{X}, \mathbf{U}) \tag{5a}$$

$$\text{s.t.} \quad \mathbf{x}_0 = \mathbf{x}(t^*), \tag{5b}$$

$$\mathbf{f}_\mathrm{impl}^\mathrm{D}(\mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{u}_k) = \mathbf{0}_{12\times 1}, \tag{5c}$$

$$\left(\mathbf{p}_{\mathrm{O},k} + {}^\mathrm{W}\mathbf{R}_{\mathrm{O},k}\mathbf{b}_\mathrm{H}\right)\mathbf{z}_\mathrm{W}^\top = z_H^0, \tag{5d}$$

$$\underline{\mathbf{u}}_\mathrm{A} \le \mathbf{u}_{\mathrm{A},k} \le \overline{\mathbf{u}}_\mathrm{A}, \tag{5e}$$

$$\underline{\mathbf{u}}_\mathrm{H} \le \mathbf{u}_{\mathrm{H},k} \le \overline{\mathbf{u}}_\mathrm{H}, \tag{5f}$$

$$\underline{\mathbf{v}}_\mathrm{H} \le {}^\mathrm{W}\mathbf{R}_{\mathrm{O},k}^\top \mathbf{v}_{\mathrm{H},k} \le \overline{\mathbf{v}}_\mathrm{H}. \tag{5g}$$

The state vector is defined as the state of the object to be transported $\mathbf{x} = [\mathbf{p}_\mathrm{O}^\top, \boldsymbol{\eta}_\mathrm{O}^\top, \mathbf{v}_\mathrm{O}^\top, \boldsymbol{\omega}_\mathrm{O}^\top]^\top \in \mathbb{R}^{12}$, while the control vector is defined as $\mathbf{u} = [\mathbf{u}_\mathrm{H}^\top, \mathbf{u}_\mathrm{A}^\top]^\top \in \mathbb{R}^6$, where $\mathbf{u}_\mathrm{A}$ represents the robot control inputs, serving as references for the compliant controller, and $\mathbf{u}_\mathrm{H}$ corresponds to the human inputs. The objective function is denoted by $J(\mathbf{X}, \mathbf{U})$, where $\mathbf{X} = [\mathbf{x}_0, \dots, \mathbf{x}_N] \in \mathbb{R}^{12\times(N+1)}$ and $\mathbf{U} = [\mathbf{u}_0, \dots, \mathbf{u}_{N-1}] \in \mathbb{R}^{6\times N}$. Regarding the constraints, Eq. (5b) imposes that the initial NMPC condition is the current object state $\mathbf{x}(t^*)$, which is provided by an object state estimator, as shown in Fig. 3. Furthermore, Eq. (5c) enforces the system dynamics for the robot, the object, and the human through the function $\mathbf{f}_\mathrm{impl}^\mathrm{D}$. Additionally, Eq. (5d) enforces a constant height for the human end-effector, as discussed in Assumption 2. Eq. (5e-5f) define the bounds on the control actions, while Eq. (5g) limits the human local velocities to promote forward motions. All these elements will be described in detail in the following sections.

2) Cost Function: The goal of the cooperative transportation task is to ensure that the object reaches a desired state, $\mathbf{x}^\mathrm{ref} = [\mathbf{p}_\mathrm{O}^{\mathrm{ref}\top}, \boldsymbol{\eta}_\mathrm{O}^{\mathrm{ref}\top}, \mathbf{v}_\mathrm{O}^{\mathrm{ref}\top}, \boldsymbol{\omega}_\mathrm{O}^{\mathrm{ref}\top}]^\top$. To achieve this, we aim to minimize the following cost function $J(\mathbf{X}, \mathbf{U})$:

$$\begin{aligned} J(\mathbf{X}, \mathbf{U}) = \sum_{k=0}^{N-1} \|\mathbf{u}_k\|^2_{\mathbf{W}_\mathbf{u}} + \sum_{k=0}^{N} \Big(\|\mathbf{p}_{\mathrm{O},k} - \mathbf{p}_\mathrm{O}^\mathrm{ref}\|^2_{\mathbf{W}_{\mathbf{p}_\mathrm{O}}} + \\ + \|\mathbf{v}_{\mathrm{O},k} - \mathbf{v}_\mathrm{O}^\mathrm{ref}\|^2_{\mathbf{W}_{\mathbf{v}_\mathrm{O}}} + \|\boldsymbol{\eta}_{\mathrm{O},k} - \boldsymbol{\eta}_\mathrm{O}^\mathrm{ref}\|^2_{\mathbf{W}_{\boldsymbol{\eta}_\mathrm{O}}} + \\ + \|\boldsymbol{\omega}_{\mathrm{O},k} - \boldsymbol{\omega}_\mathrm{O}^\mathrm{ref}\|^2_{\mathbf{W}_{\boldsymbol{\omega}_\mathrm{O}}} \Big), \end{aligned} \tag{6}$$

which is composed of two main terms: (i) minimize the control effort (first summation term); and (ii) minimize the tracking error (second summation term) in terms of linear and angular position and velocity. This cost function was chosen to reflect a reasonable human behavior in a collaborative task: reaching a target location while minimizing the effort. During the experimental validation, the human is informed about the desired object's target location before the experiment. In addition, a visual marker is placed at the target position (see the accompanying video), which helps to guarantee Assumption 3.

Each term is weighted by an appropriate weighting diagonal matrix, denoted in Eq. (6) with $\mathbf{W}_j \in \mathbb{R}^{3\times3}$, $j \in \{\mathbf{p}_\mathrm{O}, \boldsymbol{\eta}_\mathrm{O}, \mathbf{v}_\mathrm{O}, \boldsymbol{\omega}_\mathrm{O}, \mathbf{u}\}$. It is important to highlight that, due to the spherical joint on the robot's side, the weights associated with the object's roll (i.e., the first elements of $\mathbf{W}_{\boldsymbol{\eta}_\mathrm{O}}$ and $\mathbf{W}_{\boldsymbol{\omega}_\mathrm{O}}$) are set to zero, as this angle is uncontrollable. As a result, the human has full control over this degree of freedom.

3) Equality constraints due to the system dynamics: This section derives the expression for $\mathbf{f}_\mathrm{impl}^\mathrm{D}$ in Eq. (5c), which is used to enforce the dynamics of the system within our NMPC. The goal is to express the model introduced in Eq. (1) for the object as a function of the optimization variables $(\mathbf{x}, \mathbf{u})$ for well-posing our OCP. We recall that, in our framework, $\mathbf{x} = [\mathbf{p}_\mathrm{O}^\top, \boldsymbol{\eta}_\mathrm{O}^\top, \mathbf{v}_\mathrm{O}^\top, \boldsymbol{\omega}_\mathrm{O}^\top]^\top$ is the object state and $\mathbf{u} = [\mathbf{u}_\mathrm{H}^\top, \mathbf{u}_\mathrm{A}^\top]^\top$ are the inputs for the human and the robot. The model in Eq. (1) captures the dynamic behavior of the object when the two forces $\mathbf{f}_\mathrm{R}^\mathrm{I}, \mathbf{f}_\mathrm{H}^\mathrm{I}$ act on it. However, there is no explicit dependency on the optimization variables.

To this end, we can rearrange Eq. (3) and Eq. (4) to obtain:

$$\mathbf{f}_{\mathrm{R}}^{\mathrm{I}} = {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}}^{\top}(\mathbf{u}_{\mathrm{A}} - \mathbf{B}_{\mathrm{A}}\mathbf{v}_{\mathrm{R}} - \mathbf{M}_{\mathrm{A}}\dot{\mathbf{v}}_{\mathrm{R}}), \tag{7a}$$

$$\mathbf{f}_{\mathrm{H}}^{\mathrm{I}} = {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}}^{\top}({}^{\mathrm{W}}\mathbf{R}_{\mathrm{H}}\mathbf{u}_{\mathrm{H}} - m_{\mathrm{H}}\dot{\mathbf{v}}_{\mathrm{H}} - m_{\mathrm{H}}g\mathbf{z}_{\mathrm{H}}). \tag{7b}$$

As it is evident from Eq. (7), this equation depends on human-related ($\dot{\mathbf{v}}_{\mathrm{H}}$) and robot-related ($\mathbf{v}_{\mathrm{R}}, \dot{\mathbf{v}}_{\mathrm{R}}$) variables. To properly define our OCP, these variables must be expressed as functions of the object's state. To this aim, notice that the system kinematics impose a constraint on the robot-object relative position, which can be expressed as $\mathbf{p}_{\mathrm{R}} = \mathbf{p}_{\mathrm{O}} + \mathbf{r}_{\mathrm{R}}$, where $\mathbf{r}_{\mathrm{R}} = {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}}\mathbf{b}_{\mathrm{R}}$. A similar constraint holds for the human-object relative position, which is written as $\mathbf{p}_{\mathrm{H}} = \mathbf{p}_{\mathrm{O}} + \mathbf{r}_{\mathrm{H}}$, where $\mathbf{r}_{\mathrm{H}} = {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}}\mathbf{b}_{\mathrm{H}}$. By deriving these expressions, we rewrite $\mathbf{v}_{\mathrm{R}}, \dot{\mathbf{v}}_{\mathrm{R}}, \mathbf{v}_{\mathrm{H}}$ and $\dot{\mathbf{v}}_{\mathrm{H}}$ as a function of the object's state:

$$\mathbf{v}_{\mathrm{R}} = \mathbf{v}_{\mathrm{O}} + \boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{R}}, \tag{8a}$$

$$\dot{\mathbf{v}}_{\mathrm{R}} = \dot{\mathbf{v}}_{\mathrm{O}} + \dot{\boldsymbol{\omega}}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{R}} + \boldsymbol{\omega}_{\mathrm{O}} \times (\boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{R}}), \tag{8b}$$

$$\mathbf{v}_{\mathrm{H}} = \mathbf{v}_{\mathrm{O}} + \boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{H}}, \tag{8c}$$

$$\dot{\mathbf{v}}_{\mathrm{H}} = \dot{\mathbf{v}}_{\mathrm{O}} + \dot{\boldsymbol{\omega}}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{H}} + \boldsymbol{\omega}_{\mathrm{O}} \times (\boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{H}}). \tag{8d}$$

By plugging Eq. (8) in Eq. (7), $\mathbf{f}_{\mathrm{R}}^{\mathrm{I}}$ and $\mathbf{f}_{\mathrm{H}}^{\mathrm{I}}$ can be expressed in the function of the optimization variables through:

$$\begin{aligned}\mathbf{f}_{\mathrm{R}}^{\mathrm{I}} =& {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}}^{\top}(\mathbf{u}_{\mathrm{A}} - \mathbf{B}_{\mathrm{A}}(\mathbf{v}_{\mathrm{O}} + \boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{R}}) \\ & - \mathbf{M}_{\mathrm{A}}(\dot{\mathbf{v}}_{\mathrm{O}} + \dot{\boldsymbol{\omega}}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{R}} + \boldsymbol{\omega}_{\mathrm{O}} \times (\boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{R}})),\end{aligned} \tag{9a}$$

$$\begin{aligned}\mathbf{f}_{\mathrm{H}}^{\mathrm{I}} =& {}^{\mathrm{W}}\mathbf{R}_{\mathrm{O}}^{\top}\big({}^{\mathrm{W}}\mathbf{R}_{\mathrm{H}}\mathbf{u}_{\mathrm{H}} - m_{\mathrm{H}}g\mathbf{z}_{\mathrm{H}} \\ & - m_{\mathrm{H}}(\dot{\mathbf{v}}_{\mathrm{O}} + \dot{\boldsymbol{\omega}}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{H}} + \boldsymbol{\omega}_{\mathrm{O}} \times (\boldsymbol{\omega}_{\mathrm{O}} \times \mathbf{r}_{\mathrm{H}}))\big).\end{aligned} \tag{9b}$$

By substituting Eq. (9) into Eq. (1), the object model becomes explicitly dependent on the optimization variables $(\mathbf{x}, \mathbf{u})$ and their derivatives. In other words, the object model can be expressed as $\mathbf{f}_{\mathrm{impl}}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}) = \mathbf{0}_{12\times 1}$, whose expression is omitted for the sake of space. This function defines the differential equations of the human-object-robot system in an implicit form. As a final step, $\mathbf{f}_{\mathrm{impl}}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u})$ is discretized using the classical implicit Runge-Kutta method, obtaining the function $\mathbf{f}_{\mathrm{impl}}^{\mathrm{D}}(\mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{u}_k)$ that is used in Eq. (5c).

As an additional equality constraint, the end-effector of the human is constrained at a constant height z_H^0, as mentioned in Assumption 2, through Eq. (5d).

4) Inequality constraints due to control actions: The inequality constraints in Eq. (5e-5f) represent the physical actuation limitation of the human and the robot. In particular, $\underline{\mathbf{u}}_{\mathrm{A}} \in \mathbb{R}^3$ and $\overline{\mathbf{u}}_{\mathrm{A}} \in \mathbb{R}^3$ represent the lower and upper bounds of the robot's control actions, and $\underline{\mathbf{u}}_{\mathrm{H}} \in \mathbb{R}^3$ and $\overline{\mathbf{u}}_{\mathrm{H}} \in \mathbb{R}^3$ denote the corresponding bounds for the human counterpart.

5) Inequality constraints for the human velocities: Additionally, as humans have a limited movement speed, we imposed the constraint in Eq. (5g), where $\mathbf{v}_{\mathrm{H}}$ is related to the system states through Eq. (8c). The vectors $\overline{\mathbf{v}}_{\mathrm{H}} = [v_H^f, v_H^l, v_H^u]^{\top} \in \mathbb{R}^3$ and $\underline{\mathbf{v}}_{\mathrm{H}} = [v_H^b, v_H^r, v_H^d]^{\top} \in \mathbb{R}^3$ define the upper and lower bounds of the human velocities, expressed in $\mathcal{F}_{\mathrm{H}}$. These bounds give the possibility to define different limits for the human by choosing v_H^f (forward), v_H^b (backward), v_H^l (left), and v_H^r (right). Sec. III-B.6 details how to set them in order to achieve a comfortable movement for the human during the collaborative transportation task. We consider symmetric lateral velocity limits given by $v_H^l = v_H^{\mathrm{lat}}$ and $v_H^r = -v_H^{\mathrm{lat}}$, where $v_H^{\mathrm{lat}} > 0$ is the human's lateral velocity limit. Furthermore, v_H^u and v_H^d limit the upward and downward velocities of the human. In practice, these velocities will be constrained to $v_H^u = v_H^d = 0\ \mathrm{m\,s^{-1}}$, as the human is constrained to z_H^0.

6) Human-Awareness Features: To account for human behavior during the transportation task, we integrate specific features into our NMPC. Specifically, humans typically favor forward movements over lateral ones [9], [12], using lateral motions primarily for specific actions. To capture this preference, we constrain the lateral velocity to be lower than the forward velocity, ensuring $v_H^{\mathrm{lat}} < v_H^f$, where both velocities are defined in $\mathcal{F}_{\mathrm{H}}$. Notably, $v_H^{\mathrm{lat}} > 0$ is permitted, as lateral movements may be required at certain moments during the collaborative task. To facilitate human forward movement, the NMPC dynamically adjusts the orientation of the transported object throughout the motion. Specifically, when the object's positional error $e_p = \|\mathbf{p}_{\mathrm{O}} - \mathbf{p}_{\mathrm{O}}^{\mathrm{ref}}\|_2$ exceeds a predefined threshold $e_p \geq r_{out}$, the NMPC relaxes the enforcement of the desired final orientation $\boldsymbol{\eta}_{\mathrm{O}}^{\mathrm{ref}}$. This design choice reflects natural human behavior during collaborative tasks, where people tend to prioritize reaching the target location quickly and comfortably, adjusting the final orientation only when close to the goal. Consequently, this enables humans to move forward in a more natural and comfortable way without being constrained by strict orientation requirements at early stages. Once the object is within a threshold $e_p \leq r_{in}$, the NMPC enforces $\boldsymbol{\eta}_{\mathrm{O}}^{\mathrm{ref}}$. A smoothing function is applied between r_{out} and r_{in} to ensure a seamless transition. To implement this behavior, we adjust the weighting matrix associated with the heading direction during the trajectory. In particular, we modulate the weights $w_{\psi_O} \in \mathbb{R}$ and $w_{\dot{\psi}_O} \in \mathbb{R}$, which correspond to the bottom-right components of $\mathbf{W}_{\boldsymbol{\eta}_{\mathrm{O}}}$ and $\mathbf{W}_{\boldsymbol{\omega}_{\mathrm{O}}}$, respectively. These weights are set to low values when the object is far from the goal object position, and they gradually increase as it approaches the target. This is achieved with:

$$w_i = \begin{cases} w_i^{min} & \text{if } e_p \geq r_{out}, \\ w_i^{max} + a_i(e_p - r_{in}) & \text{if } r_{in} < e_p < r_{out}, \\ w_i^{max} & \text{if } 0 \leq e_p \leq r_{in}, \end{cases} \tag{10}$$

where w_i^{min} and w_i^{max} are the weighting values associated with the state $i \in \{\psi_O, \dot{\psi}_O\}$ when the object is far from and close to the reference $\mathbf{p}_{\mathrm{O}}^{\mathrm{ref}}$, respectively. Additionally, $a_i = \frac{w_i^{min} - w_i^{max}}{r_{out} - r_{in}}$ is the slope of the line that interpolates between the transition distances r_{in} and r_{out}.

IV. Results

A. Experimental Setup

The framework was validated with the setup in Fig. 1 where the goal is to transport an object to a desired location. The object is a lightweight, 2-meter-long carbon fiber bar. The robot is a fully-actuated hexarotor with tilted propellers, connected to the object via a spherical joint. The bar has a uniform mass, with its CoM at its geometric center.

Consequently, the attachment points for the human and the robot are defined as $\mathbf{b}_{\mathrm{H}} = -\frac{l_{\mathrm{O}}}{2}\mathbf{x}_{\mathrm{O}}$ and $\mathbf{b}_{\mathrm{R}} = \frac{l_{\mathrm{O}}}{2}\mathbf{x}_{\mathrm{O}}$, where l_{O} is the length of the bar. The reference attitude for the robot, decoupled from its reference position given by the admittance controller, is set to $\phi_R^{\mathrm{ref}} = 0^\circ$ and $\theta_R^{\mathrm{ref}} = 0^\circ$, as we use a fully-actuated vehicle. The robot yaw angle is aligned with that of the object to maintain a practical range of motion for the spherical joint, i.e., $\psi_R^{\mathrm{ref}} = \psi_O$. Although the current implementation exploits full actuation to decouple rotation from translation, the proposed control framework remains applicable to standard underactuated aerial robots. In such cases, roll and pitch would be implicitly defined by the position controller.

All components of the control framework presented in Sec. III are implemented using the software framework *TeleKyb*. The admittance controller, the wrench observer, the robot position controller, and the robot state estimator operate on an onboard Odroid computer running on the aerial robot. Conversely, the NMPC runs on a ground computer at 50 Hz, and communicates with the onboard computer via WiFi. The NMPC is implemented using acados [18], with a prediction horizon of $T_H = 3.0$ s and $N = 6$ shooting nodes. This configuration represents a compromise between prediction quality and computational efficiency, ensuring that the OCP can be solved within the control frequency. Although the number of shooting nodes is quite low, the goal is not to obtain highly precise trajectory predictions, but rather to capture the intended cooperative behavior. The long horizon allows the optimizer to anticipate meaningful progress toward the goal, while the reduced number of nodes limits computation time, making real-time implementation feasible. Furthermore, the object state estimator, which also runs on the ground computer, uses the motion capture system to estimate the state of the carbon fiber bar. The parameters used in our experiments are detailed in Tab. I. A special mention should be provided for the weights w_{ψ_O} and $w_{\dot{\psi}_O}$ associated with the yaw and for r_{in} and r_{out}, which have been adjusted for each scenario and will be further discussed in Sec. IV-B.

The human is aware of the desired state $\mathbf{x}^{\mathrm{ref}}$ thanks to a vertical support that is placed in the desired location where the bar should be placed (see the accompanying video), validating Assumption 3. Finally, as discussed in Assumption 2, the motion naturally remains in the $\mathbf{x}_{\mathrm{W}} - \mathbf{y}_{\mathrm{W}}$ plane. Therefore, the pitch limits of the spherical joint are never reached during operation, which inherently prevents violations of Assumption 1 in the pitch rotation. In addition, to simplify the presentation and improve readability, we report data only on this plane.

TABLE I

MAIN PARAMETERS FOR THE EXPERIMENTS.

Parameter	Value
m_{O} / l_{O}	0.232 kg / 2.0 m
$\underline{\mathbf{u}}_{\mathrm{A}}$ / $\overline{\mathbf{u}}_{\mathrm{A}}$	$-[6, 6, 6]^\top$ N / $[6, 6, 6]^\top$ N
$\underline{\mathbf{u}}_{\mathrm{H}}$ / $\overline{\mathbf{u}}_{\mathrm{H}}$	$-[4, 4, 4]^\top$ N / $[4, 4, 4]^\top$ N
$\mathbf{W}_{\mathbf{u}}$	diag([10, 10, 10, 10, 10, 10])
$\mathbf{W}_{\mathbf{p}_{\mathrm{O}}}$ / $\mathbf{W}_{\mathbf{v}_{\mathrm{O}}}$	diag([10, 10, 10]) / diag([1, 1, 1])
$\mathbf{W}_{\boldsymbol{\eta}_{\mathrm{O}}}$ / $\mathbf{W}_{\boldsymbol{\omega}_{\mathrm{O}}}$	diag([0, 10, w_{ψ_O}]) / diag([0, 1, $w_{\dot{\psi}_O}$])
v_H^f / v_H^b / v_H^{lat}	2.0 m s^{-1} / -2.0 m s^{-1} / 0.5 m s^{-1}

B. Experimental Validation

The framework has been validated on three scenarios with two distinct human operators. In all scenarios, the objective is to transport the object to a target location and then return it to its initial position. The human operator controls the transition between goal locations using a switch on a remote control. In total, 13 experiments were conducted In the following sections, we present the results for each scenario.

1) Forward and backward movements: The first experiment aims to validate the control framework in a simple scenario, where the bar is requested to move along its main axis without inducing lateral movements in the human operator. To achieve this, the object is commanded to move from $\mathbf{p}_{\mathrm{O}}^{\mathrm{ref}} = [-1.4,\ 0.1,\ 1.5]^\top \mathrm{m}$ to $\mathbf{p}_{\mathrm{O}}^{\mathrm{ref}} = [1.4,\ 0.0,\ 1.5]^\top \mathrm{m}$, and then return to the original location, while maintaining $\psi_O^{\mathrm{ref}} = 0^\circ$. Within this scenario, we impose constant values for the weights associated with the yaw angle. Specifically, $w_{\psi_O} = 10$ and $w_{\dot{\psi}_O} = 1$. Keeping these weights fixed prevents the object from reorienting. This choice aligns with the simplicity of this scenario, where the object movement occurs along a specific direction, and no reorientation is needed. The results for this scenario are illustrated in Fig. 4. Specifically, the first column of Fig. 4 illustrates the pose of the transported object, the robot's horizontal control inputs, and the human's lateral velocities, expressed in the human frame $\mathcal{F}_{\mathrm{H}}$. As expected, our control framework successfully generates robot motions that guide the object to its desired goal position while satisfying the OCP constraints in (5b–5g). This is further validated by the top plot in Fig. 5, which presents the one-way trajectory of the object in the $\mathbf{x}_{\mathrm{W}} - \mathbf{y}_{\mathrm{W}}$ plane.

It is worth highlighting that the NMPC predicts the system's behavior with reasonable accuracy, particularly over short time horizons (see the NMPC predictions in Fig. 4). For example, in the x_O position – which represents the main movement across all experiments – the predicted trajectories align closely with the measured ones, minimizing the tracking error. In contrast, discrepancies can emerge in other components, such as the y_O position. Since the intended motion in y_O is to remain approximately constant, any tracking errors are relatively small and less perceptible. As a result, the operator may not actively correct for these minor deviations, leading to a mismatch between the actual behavior and the NMPC predictions in that direction. However, due to the presence of the admittance controller and the iterative nature of the NMPC, our control framework adapts to the human behavior.

2) Lateral movements: After validating longitudinal motions, this experiment aims to confirm that lateral movements are uncomfortable for humans, as noted in [9], [12] for non-aerial robots. To this end, the bar is commanded to move laterally from $\mathbf{p}_{\mathrm{O}}^{\mathrm{ref}} = [-2.1,\ -1.0,\ 1.5]^\top \mathrm{m}$ to $\mathbf{p}_{\mathrm{O}}^{\mathrm{ref}} =$

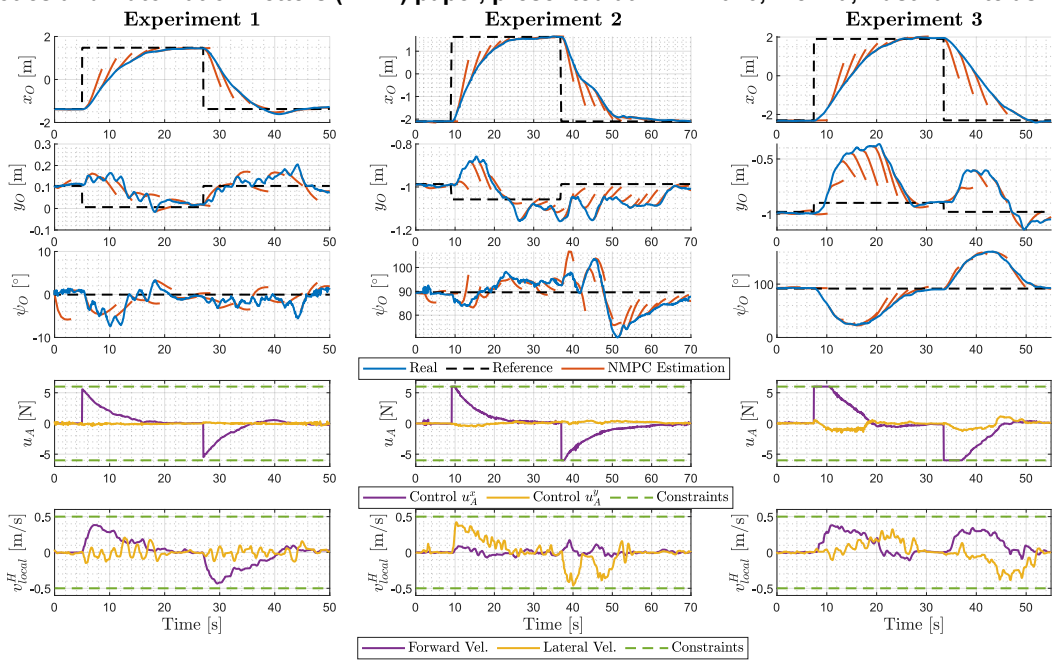

Fig. 4. The results of the three experiments are visualized as follows: the forward/backward scenario (left column), the lateral movement scenario (central column), and the reorientation scenario (right column). For each scenario, the 2D pose of the object – represented by x_O, y_O, and ψ_O – is shown in the first three rows. Specifically, the desired object pose is indicated by black dashed lines, while the actual object pose is represented by solid blue lines. NMPC predictions, plotted every three seconds for clarity, are shown in red. In the fourth row, the robot control inputs are reported in purple and yellow, together with their limits, depicted in green. Lastly, the human velocities, expressed in $\mathcal{F}_{\mathrm{H}}$, are displayed in the last row. Specifically, the human's forward velocity is depicted in purple, and the lateral velocity is represented in yellow. For the sake of clarity, only the lateral velocity bounds are represented (dashed green line), while the forward/backward velocity bounds are set to $v_H^f = 2\ \mathrm{m\,s^{-1}}$ and $v_H^b = -2\ \mathrm{m\,s^{-1}}$.

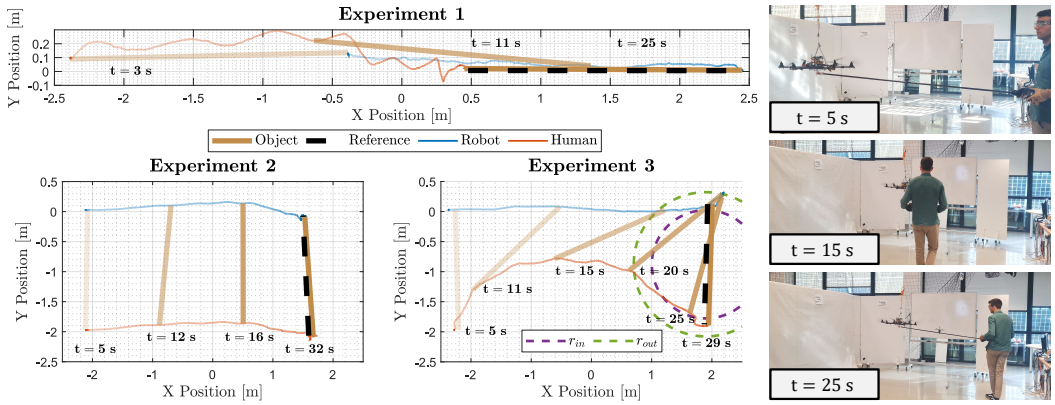

Fig. 5. Left: Trajectories of the robot (blue), the human (red), and the object (orange) in the $\mathbf{x}_{\mathrm{W}} - \mathbf{y}_{\mathrm{W}}$ plane for the three experimental scenarios. To enhance clarity, only the forward movement is shown, omitting the trajectories for returning to the initial position. The transparency effect indicates the temporal progression of the trajectories, transitioning from transparent (initial) to solid (final) lines. In Experiment 3, the transition zone, defined by r_{in} and r_{out} in Eq. (10), is represented by the purple and green dashed circles, respectively. Right: Three different snapshots from Experiment 3.

$[1.6,\ -1.05,\ 1.5]^\top$m, and then return to the original position, while maintaining $\psi_O^{\mathrm{ref}} = 90°$. No orientation correction mechanism is applied – achieved by setting $w_{\psi_O} = 10$ and $w_{\dot{\psi}_O} = 1$ – to ensure that our control framework favors lateral movements. The central column of Fig. 4 illustrates the experimental results over time, while the bottom-left section of Fig. 5 presents the object's trajectory in the $\mathbf{x}_{\mathrm{W}} - \mathbf{y}_{\mathrm{W}}$ plane. As evident from these figures, our framework successfully generates the optimal inputs for the robot to complete the task while satisfying all constraints, including the human lateral velocity limit. However, the latter is often close to its limits, as shown in Fig. 4. In addition, when moving laterally, humans typically alternate between short steps and brief pauses, rather than maintaining a constant velocity. This can be observed between 35 s and 55 s in Fig. 4, where the human lateral velocity fluctuates significantly. This results in noticeable deviations in the yaw angle of the object, reaching up to 20° during these intervals. This, coupled with the need for the human to continuously glance sideways while keeping track of the robot (see the accompanying video), confirms that lateral motions are uncomfortable for the human and should be

minimized whenever possible, as in the next scenario.

3) Reorientation of the object for human comfort: In this experiment, we validate the reorientation mechanism described in Eq. (10). The results are reported in the last column of Fig. 4, which shows the evolution in time of the object in the $\mathbf{x}_{\mathrm{W}} - \mathbf{y}_{\mathrm{W}}$ plane, as well as the robot inputs and the human velocity in $\mathcal{F}_{\mathrm{H}}$. The transportation task is correctly executed by reaching the desired object state, as confirmed by the bottom-right plot of Fig. 5 that shows the experiment in the $\mathbf{x}_{\mathrm{W}} - \mathbf{y}_{\mathrm{W}}$ plane. Referring to Fig. 4, the orientation error initially increases at the start of the trajectory. This is due to the reorientation mechanism in Eq. (10), which prioritizes human forward movement when the object is far from the desired goal location, even at the cost of a temporary rise in the orientation error. As the object approaches the goal, the orientation error begins to decrease, driven by adjustments in the weights w_{ψ_O} and $w_{\dot{\psi}_O}$ in Eq. (10). In the transition zone, the bar is reoriented toward the desired final orientation through lateral movements of both the human and the aerial robot. In other words, the human first reorients to move in their forward direction, as shown by the velocity plot in the last column of Fig. 4 between 8 and 18 seconds. Upon approaching the target location, they complete the task using lateral movements to fine-tune the object's orientation, as observed in the same plot between 18 and 28 seconds. These movements are shown on the right side of Fig. 5, which depicts three different time instants of the experiment. To obtain this behavior, the following parameters in Eq. (10) were chosen: $w_{\psi_O}^{max} = 10$, $w_{\psi_O}^{min} = 1$, $w_{\dot{\psi}_O}^{max} = 0.01$, $w_{\dot{\psi}_O}^{min} = 0.001$, $r_{in} = 0.9$ m, and $r_{out} = 1.2$ m.

This experiment demonstrates how the proposed framework facilitates transportation tasks by prioritizing forward movements, resulting in a behavior reminiscent of two humans collaborating. Specifically, during long-distance transportation tasks, our framework initially encourages forward motions, with object reorientation primarily occurring near the final position, mirroring natural human collaboration strategies.

V. CONCLUSIONS

This work presents a control framework for collaborative transportation tasks involving a human operator and an aerial robot. The proposed approach integrates two controllers: (i) a compliant controller that reduces the aerial robot to a mass-damper system, and (ii) a Nonlinear Model Predictive Control (NMPC) that generates optimal inputs for the compliant controller for transporting an object to a target location. Our NMPC explicitly considers the dynamics of the human, the robot, and the object as equality constraints, as well as it imposes limits for the control variables. To enhance human comfort, the framework promotes forward motion over lateral movement by restricting the human's lateral velocities and enabling object reorientation, thereby facilitating a more natural forward movement for the human operator. The framework is validated through experiments with a fully actuated hexarotor. The results demonstrate the framework's capability to generate real-time inputs for the robot, consistently prioritizing forward motion for the human whenever feasible. Future works will focus on the automatic tuning of the weights of the OCP cost function, allowing the aerial robot to dynamically adapt its behavior. To enable deployment outside lab environments, robust state estimation for the robot and the bar and safety mechanisms in unstructured settings will be required.

REFERENCES

[1] A. Ollero, G. Heredia, A. Franchi, *et al.*, "The AEROARMS Project: Aerial Robots with Advanced Manipulation Capabilities for Inspection and Maintenance," *IEEE Robotics Automation Magazine*, vol. 25, no. 4, pp. 12–23, 2018.

[2] C. Geckeler, S. E. Ramos, M. C. Schuman, *et al.*, "Robotic volatile sampling for early detection of plant stress: Precision agriculture beyond visual remote sensing," *IEEE Robotics & Automation Magazine*, pp. 2–12, 2023.

[3] N. Nikhil, S. M. Shreyas, G. Vyshnavi, *et al.*, "Unmanned aerial vehicles (uav) in disaster management applications," in *2020 Third Int. Conf. on Smart Systems and Inventive Technology*, 2020, pp. 140–148.

[4] A. Suarez, A. Gonzalez, C. Alvarez, *et al.*, "Through-window home aerial delivery system with in-flight parcel load and handover: Design and validation in indoor scenario," *Int. Journal of Social Robotics*, pp. 1–24, 2024.

[5] G. Corsini, M. Jacquet, H. Das, *et al.*, "Nonlinear model predictive control for human-robot handover with application to the aerial case," in *2022 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2022, pp. 7597–7604.

[6] M. Tognon, R. Alami, and B. Siciliano, "Physical human-robot interaction with a tethered aerial vehicle: Application to a force-based human guiding problem," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 723–734, 2021.

[7] D. J. Agravante, A. Cherubini, A. Bussy, *et al.*, "Collaborative human-humanoid carrying using vision and haptic sensing," in *2014 IEEE Int. Conf. on Robotics and Automation*, 2014, pp. 607–612.

[8] A. Bussy, P. Gergondet, A. Kheddar, *et al.*, "Proactive behavior of a humanoid robot in a haptic transportation task with a human partner," in *2012 IEEE RO-MAN: The 21st IEEE Int. Symposium on Robot and Human Interactive Communication*, 2012, pp. 962–967.

[9] G. Arechavaleta, J.-P. Laumond, H. Hicheur, *et al.*, "On the nonholonomic nature of human locomotion," *Auton. Robots*, vol. 25, pp. 25–35, 08 2008.

[10] P. Prajapati and V. Vashista, "Aerial physical human robot interaction for payload transportation," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4903–4910, 2023.

[11] G. Li, X. Liu, and G. Loianno, "Human-aware physical human–robot collaborative transportation and manipulation with multiple aerial robots," *IEEE Transactions on Robotics*, vol. 41, pp. 762–781, 2025.

[12] G. Arechavaleta, J.-P. Laumond, H. Hicheur, *et al.*, "The nonholonomic nature of human locomotion: a modeling study," in *The First IEEE/RAS-EMBS Int. Conf. on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006.*, 2006, pp. 158–163.

[13] M. Hamandi, F. Usai, Q. Sablé, *et al.*, "Design of multirotor aerial vehicles: A taxonomy based on input allocation," *The Int. Journal of Robotics Research*, vol. 40, no. 8-9, pp. 1015–1044, 2021.

[14] F. Ficuciello, L. Villani, and B. Siciliano, "Variable impedance control of redundant manipulators for intuitive human–robot physical interaction," *IEEE Transactions on Robotics*, vol. 31, no. 4, pp. 850–863, 2015.

[15] M. Ryll, G. Muscio, F. Pierri, *et al.*, "6d physical interaction with a fully actuated aerial robot," in *2017 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2017, pp. 5190–5195.

[16] T. Tomić, C. Ott, and S. Haddadin, "External wrench estimation, collision detection, and reflex reaction for flying robots," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1467–1482, 2017.

[17] M. Faessler, A. Franchi, and D. Scaramuzza, "Differential flatness of quadrotor dynamics subject to rotor drag for accurate tracking of high-speed trajectories," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, p. 620–626, Apr. 2018.

[18] R. Verschueren, G. Frison, D. Kouzoupis, *et al.*, "acados—a modular open-source framework for fast embedded optimal control," *Mathematical Programming Computation*, vol. 14, no. 1, pp. 147–183, 2022.