Door-to-Door Parcel Delivery from Supply Point to Users Home with Heterogeneous Robot Team: euROBIN First Year Robotics Hackathon

Alejandro Suarez[1], Rainer Kartmann[2], Daniel Leidner[3], Luca Rossini[4], Johann Huber[5], Carlos Azevedo[6], Quentin Rouxel[7], Marko Bjelonic[8], Antonio Gonzalez-Morgado[1], Christian Dreher[2], Peter Schmaus[3], Arturo Laurenzi[4], François Hélénon[5], Rodrigo Serra[6], Jean-Baptiste Mouret[7], Lorenz Wellhausen[8], Vicente Perez-Sanchez[1], Jianfeng Gao[2], Adrian Simon Bauer[3], Alessio De Luca[4], Mouad Abrini[5], Rui Bettencourt[6], Olivier Rochel[7], Joonho Lee[8], Pablo Viana[1], Christoph Pohl[2], Nesrine Batti[3], Diego Vedelago[4], Vamsi Krishna Guda[5], Alexander Reske[8], Carlos Alvarez[1], Fabian Reister[2], Werner Friedl[3], Corrado Burchielli[4], Aline Baudry[5], Fabian Peller-Konrad[2], Thomas Gumpert[3], Luca Muratore[4], Philippe Gauthier[5], Franziska Krebs[2], Sebastian Jung[3], Lorenzo Baccelliere[4], Hippolyte Watrelot[5], Andre Meixner[2], Anne Köpken[3], Mohamed Chetouani[5], Pascal Weiner[2], Florian Lay[3], Felix Hundhausen[2], Anne Reichert[3], Noémie Jaquier[2], Florian Schmidt[3], Marco Sewtz[3], Freek Stulp[3], Lioba Suchenwirth[3], Rudolph Triebl[3], Xuwei Wu[3], Begoña Arrue[1], Rebecca Schedl-Warpup[3], Marco Hutter[8], Serena Ivaldi[7], Pedro U. Lima[6], Stéphane Doncieux[5], Nikos Tsagarakis[4], Tamim Asfour[2], Anibal Ollero[1], and Alin Albu-Schäffer[3]

***Abstract*—Logistics and service operations involving parcel preparation, delivery, and unpacking from a supply point to the user's home could be carried out completely by robots in the near future, taking benefit of the capabilities of the different robot morphologies for the logistics, outdoors, and domestic environments. The use of robots for parcel delivery can contribute to the goals of sustainability and reduced emissions by exploiting the different locomotion modalities (wheeled, legged, and aerial). This paper reports the development and results obtained from the first robotics hackathon celebrated as part of the European Robotics and Artificial Intelligence Network (euROBIN) involving eight robotic platforms in three domains: 1) an industrial robotic arm for parcel preparation at the supply point, 2) a Centauro robot, a dual-arm aerial manipulator, and a wheeled-legged quadruped for parcel transportation, and 3) two humanoid robots and two commercial mobile manipulators for parcel delivery and unpacking in domestic scenarios. The paper describes the joint operation and the evaluation scenario, the features and capabilities of the robots, particularly those involved in the realization of the tasks, and the lessons learned.**

***Index Terms*—robotics hackathon, heterogeneous robots, parcel delivery.**

MULTIMEDIA

The video of the euROBIN Robotics Hackathon 2023 can be found in https://www.youtube.com/watch?v=vrRwY7f0g8I.

This work is supported by euROBIN, the European ROBotics and AI Network (Grant agreement ID: 101070596) funded by the European Commission.

Manuscript received January XX, 2024; revised March XX, 2024.

[1]University of Seville (USE), GRVC Robotics Lab.

[2]Karlsruhe Institute of Technology (KIT).

[3]German Aerospace Center (DLR).

[4]Fondazione Istituto Italiano di Tecnologia (IIT).

[5]Sorbonne University (SU), ISIR, CNRS.

[6]Instituto Superior Técnico, U. Lisboa (IST).

[7]Institut National de Recherche en Informatique et Automatique (INRIA), CNRS, Université de Lorraine.

[8]Eidgenoessische Technische Hochschule Zuerich (ETHZ).

I. INTRODUCTION

It is expected that in the next years robots will be increasingly adopted in logistics, delivery, and domestic scenarios, being able to conduct autonomously complete operations from the supply point to the user's home, involving the preparation, transportation, and unpacking of parcels with goods.

Several works have studied the use of autonomous robots for last-mile delivery [1], [2] in urban environments [3], proposing different strategies to optimize the cost or minimize the tardiness , taking also benefit of the different locomotion capabilities provided by wheeled, legged, and aerial robots intended to deliver the parcel directly to the user's home. However, the next level of autonomy in supply operations requires the interaction between different robots located in the three domains (packaging, transportation, and domestic) and the experimental evaluation to demonstrate the feasibility of conducting this operation without human intervention.

Robotics hackathons [4], competitions [3], and challenges have served to demonstrate the ability of robots to conduct certain tasks in a cooperative way in close to real-world scenarios, requiring diverse capabilities to accomplish the task, including perception, navigation, exploration, and manipulation, among others.

As part of the European Robotics and Artificial Intelligence Network, the First euROBIN Event was celebrated on 15-19 May 2023 in Seville, comprising a robotics hackathon with the participation of eight teams from research groups in Europe: the German Aerospace Center (DLR), Karlsruhe Institute of Technology (KIT), Institut National de Recherche en Informatique et Automatique (INRIA), Sorbonne University (SU), Instituto Superior Técnico, U. Lisboa (IST), Eidgenoessische Technische Hochschule Zuerich (ETHZ), Universidad de Sevilla (USE, organizer of the event), and Fondazione

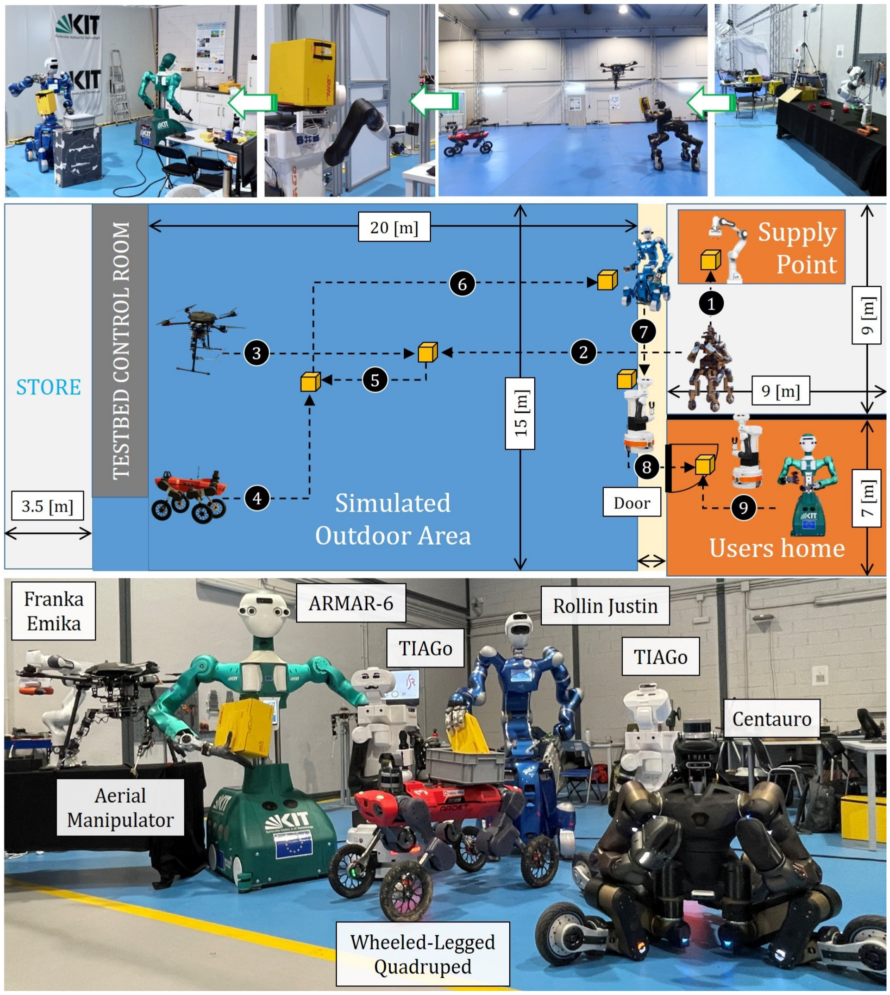

Fig. 1. Scenario of the euROBIN robotics hackathon and involved robots.

Istituto Italiano di Tecnologia (IIT). The robots participating in the hackathon are shown in Figure 1.

The main contribution of this paper is the experimental demonstration of a cooperative parcel delivery task conducted by a team of eight different robots in a representative indoor scenario depicted in Figure 1, involving an industrial robotic arm for preparing the parcel at the supply point, which is transported by three outdoor robots (Centauro, aerial manipulator, and wheeled-legged quadruped) and delivered by three mobile manipulators at the user's home where another humanoid robot unpacks the parcel. The operation involves the visual detection and bimanual grasping of the parcel with contact force control, the aerial hand-over and parcel drop on a carrier box, the mapping and autonomous navigation of the scenario for ringing a doorbell, opening the door of the user's home, and unpacking and storing the objects in a kitchen scenario. The paper presents an overview of the robots, their capabilities, and functionalities, as well as several lessons learned from the experimental demonstration.

The key achievements derived from this work are:

- The realization of the complete operation involving eight different robots without direct physical human intervention but combining teleoperation and autonomous robot behaviors using the same parcel as benchmarking item in the three considered domains (logistics, transportation, and home service).
- The first aerial hand-over operation between a human-like dual-arm aerial manipulation robot and a Centauro robot, and the following parcel drop on a wheeled-legged quadruped equipped with a carrier box.
- The cooperative robot-to-robot parcel delivery at user's home, involving ringing and opening the door, and unpacking the objects of the parcel.
- A list of lessons learned, including practical aspects and scientific topics to be addressed in order to improve the capabilities and performance of robot teams in applications as the one considered here.

The rest of the paper is organized as follows. Section II describes the intended operation, the scenario, and some considerations. The robotic platforms involved in the hackathon are introduced in Section III, whereas Section IV describes the functionalities of each robot involved in the demonstration. Experimental results are reported in Section V, remarking the conclusions and lessons learned in Section VII.

II. Problem Statement

A. Task Definition

The euROBIN project considers three representative application domains for the robots participating in the hackathon, shown in Figure 1:

1) Robotic Manufacturing for a Circular Economy.
2) Robots for Enhanced Quality of Life and Well-Being.
3) Outdoor Robots for Sustainable Communities.

The hackathon is considered here as a cooperative scenario in which all robots have to participate to solve the challenge consisting in the preparation, transportation, delivery, and unpacking of a parcel, from the supply point where the industrial robotic arm drops the objects in the box, to a mockup scenario of the user's home involving the parcel handover between robots on the way, as well as ringing, opening, and passing through a door. The choice of this task is motivated by two main reasons. On the one hand, as a milestone, it represents a complete logistics operation carried out only by robots, with partial human support only through teleoperation. On the other hand, it involves several skills (navigation, manipulation, perception) of robots with very different morphologies, promoting interactions between them that in some cases have not been explored so far, such as the aerial-ground robot handover.

The design of the challenge, using a standard parcel as benchmarking object shared among the different robots in a chain, facilitates the definition of evaluation metrics such as the execution time of each of the phases and interactions between robots, or the identification of reliability issues and requirements or constraints imposed by the robots. The use of the standardized object was done also with the aim of promoting in a future hackathon transfer of solutions between robots, for example for object detection, localization, grasping and manipulation. Furthermore, the challenge involves the interaction with a sensorized robotic door comprising a doorbell button, a handle, and a sensor to evaluate the door opening.

B. Scenario Description

The robotics hackathon was organized in the GRVC Aerial Robotics Laboratory of the University of Seville. This facility, with dimensions of $36\,\mathrm{m}$ length by $20\,\mathrm{m}$ width by $10\,\mathrm{m}$ height, provides space enough for the eight robots and the 68 people participating in the event. Figure 1 shows the distribution of the robots and the three main areas: the Supply Point, the Simulated Outdoor Area, and the Home environment. The facility includes a $20\,\mathrm{m}$ by $15\,\mathrm{m}$ testbed equipped with 28 Opti-Track cameras and a safety net for flying the Aerial Manipulation robot in safe conditions.

The robots are located in the four areas identified in Figure 1, indicating the main dimensions as well as the paths followed by the robots and the locations of the handover:

1) The Supply Point where the Franka Emika robotic arm loads the parcel retrieved by the Centauro robot.
2) The Simulated Outdoor Area, corresponding to the flight arena, where the Aerial Manipulator takes the parcel from Centauro and drops it on the ANYmal on Wheels.

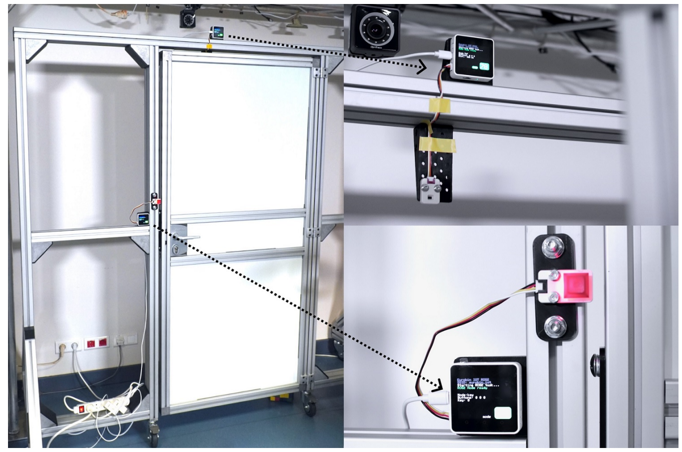

Fig. 2. Sensorized door, with two sensors (the doorbell and a time-of-flight sensor on top of the frame) that publish the state of the door on ROS2 topics.

3) The hallway between the blue and orange areas, where Rollin' Justin takes the parcel delivered by the Wheeled-Legged Quadruped and puts it on the back of TIAGo-1 robot that navigates to the door.
4) The User's Home mockup scenario, accessed through the robotic door which is opened by TIAGo-2 robot, and where ARMAR-6 takes the parcel, puts it on the table, and stores the objects on the drawers.

The User's home scenario includes a sensorized door and a kitchen, shown in Figure 2. It is constructed using aluminum profiles to facilitate disassembly and replication while closely adhering to the dimensions and handle mechanism typical of standard doors. It is also equipped with wheels so that it can be easily moved in the lab. The sensor suite is based on the M5Stack Core 2 product family, which centers around a Core housing the WiFi-enabled ESP32 CPU, a touch screen, and an extensive collection of pre-packaged sensors. Communication between these sensors and robotic systems relies on the Micro-ROS project[1], which provides a ROS2 [5] interface for microcontrollers. The connection can be established either directly by WiFi, or through an agent running on a more powerful board and connected by WiFi to the M5Stack Cores.

During the hackathon, two sensors were used: (1) a doorbell press button, which triggers both an audible alert and activates a light while also publishing ROS2 messages, and (2) a time-of-flight sensor that publishes the distance between the top of the door and the door frame, allowing for an assessment of the door's degree of openness. Overall, this sensor system prefigures a ROS2-based Internet of Things (IoT) that can be extended with other sensors (for example, for sensing the opening of a drawers or weighing laundry) and fully integrated with the robotic ecosystem. In particular, the self-discovery abilities of ROS2 make it possible to connect sensors on the WiFi network and see them appear as ROS2 topics. The code for interfacing a set of sensors with Micro-ROS on the M5Stack is available online[2].

[1] https://micro.ros.org/
[2] https://github.com/hucebot/ros_m5core

TABLE I
MAIN FEATURES OF THE EIGHT ROBOTIC PLATFORMS INVOLVED IN THE HACKATHON.

Platform	Mobile base Num. actuators	Manipulator DoFs*	Weight kg	Payload kg	Power W	Positioning method
Franka Emika FR3	-	7	18	3	80	Joint encoders, eye-in-hand camera
Centauro	4×6	2×6 + 1	117	2 × 10	750	Encoders + IMU + odometry
Aerial Manipulator	4	2×4	7	2 × 0.5	1000	Opti-Track + IMU + encoders
ANYmal on Wheels	4×4	None	50	10	450	Joint encoders, IMU, LiDAR
Rollin' Justin	4×3	2×7 + 2×13 + 5	180	2×15	1000	Visual SLAM, joint encoders
TIAGo-1	2	7 + 1	70	3	60	AMCL with Laser Scans
TIAGo-2	2	7 + 1 + 1	70	3	60	Human visual feedback
ARMAR-6	4	2×8 + 2×2 + 1	160	2×10	460	Laser scans, joint encoders

*Including arm, hands, and torso.

TABLE II
CAPABILITIES AND PURPOSE OF THE INVOLVED ROBOTS IN THE HACKATHON.

Platform	Capabilities	Role in the hackathon	Location
Franka Emika FR3	Vision-based grasping, hand-guided manipulation	Dropping objects on parcel, closing box flaps	Supply Point
Centauro	Wheeled-legged locomotion, bimanual manipulation	Parcel grasping, transportation, handover	Supply P. - SOA*
Aerial Manipulator	Hover flight, bimanual manipulation, teleoperation	Aerial parcel grasping and drop on quadruped	SOA
ANYmal on Wheels	Hybrid locomotion and autonomous navigation	Parcel transportation, delivery to Rollin' Justin	SOA - Hallway
Rollin' Justin	Supervised autonomy, force-feedback telepresence	Parcel grasping and handover to TIAGo-1	Hallway
TIAGo-1	Autonomous navigation and object manipulation	Parcel transport, ring doorbell, access home	Hallway - User's H.
TIAGo-2	Teleoperation, mobile navigation	Door opening, let TIAGo-1 enter home	User's Home
ARMAR-6	Autonomous bimanual mobile manipulation	Parcel grasping, unpacking and storing objects	User's Home

*Simulated Outdoor Area.

C. Considerations

In order to facilitate the grasping and handover for some of the robots, the standard cardboard parcel was slightly modified by incorporating a handle. The parcel is loaded with common and representative objects of the domestic environment, including in this case two plastic bottles and a cup, weighing 0.5 kg total. Note that the payload capacity of the dual arm aerial manipulation robot, around 1 kg, determines the maximum weight that can be handled.

The Outdoor Area was simulated in the indoor testbed since current regulation does not allow flying drones weighing more than 250 grams in urban environments. Also, sharing a common space contributes to promoting the collaboration between robots and teams, simplifies the logistics, and avoids possible weather inconveniences.

As it will be described later, some of the robots were teleoperated in the navigation or manipulation tasks for simplicity reasons. Note that no integration work was done before the hackathon, since the challenge was defined at the beginning of the event.

III. ROBOTIC PLATFORMS

This section presents a short overview of the different robotic platforms involved in the door-to-door parcel delivery operation, indicating the main features and capabilities along with the interactions with the other robots. The functionalities employed in the operation are detailed in the next Section. In the following, robots are presented in their order of participation in the hackathon. The main features of these robotic platforms are collected in Table I, summarizing in Table II the involved capabilities and role of each robot in the hackathon.

A. Industrial Robotic Arm

Franka Emika FR3 is a 7 Degrees of Freedom (DoF) robot system tailored to both research and industry. The arm is equipped with torque sensors at each joint. A single DoF parallel gripper is used for grasping and dropping the objects within the parcel, and for closing the flaps of the box before the Centauro robot takes it away. This manipulator guarantees an industrial-grade pose repeatability of ±0.1 mm, 855 mm reach, and a workspace coverage of 94.5%. It has a payload of 3 kg. The FR3 was augmented with a wrist-mounted RGB-D camera (Realsense D435i) for perception of the environment. The robot was controlled through its ROS1 interface.

B. Centauro

Executing loco-manipulation tasks involving the manipulation and navigation of obstacles within intricate environments demands the capabilities of a humanoid robot. The Centauro robot [6], distinguished by its hybrid wheeled-legged quadrupedal design, exhibits robust mobility in diverse settings, particularly in unstructured environments. With a total of 38 DoFs, this robot boasts four legs and two arms, each driven by six individual joints. Further, it incorporates a torso yaw mechanism to expand its manipulation range and a head pitch joint that enhances the camera's field of view mounted on its head.

In its pursuit of achieving autonomous behavior, Centauro is equipped with a pair of Intel RealSense cameras for capturing short-range environmental data and a LiDAR sensor for a more expansive observational scope. The platform exhibits the capability to manipulate items weighing up to 10 kg per arm, rendering it apt for tasks such as transporting parcels from loading areas to dual-arm aerial manipulators, moving in

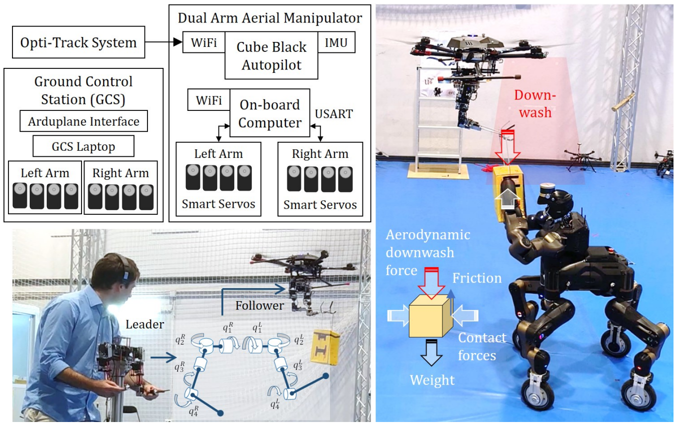

Fig. 3. Components and architecture of the dual-arm aerial manipulator with leader-follower kinaesthetic teleoperation interface (left). Identified forces acting on the parcel during the aerial handover between Centauro and the Aerial Manipulator (right).

environments characterized by real-world obstacles like stairs or intricate terrains.

C. Dual-Arm Aerial Manipulator

The aerial manipulation robot is intended to conduct the grasping, transportation, and delivery of parcels in areas that cannot be reached by ground robots, for example delivering in the roof of buildings, or overcoming the dense traffic of the roads. The platform consists of a medium capacity multi-rotor (4 kg payload, excluding the dual arm, with 5-10 min flight time) equipped with a lightweight and compliant anthropomorphic dual-arm manipulator (LiCAS). A modified version of this platform was used in [7] for the realization of maintenance operations on power lines. The low weight of the arms (2.5 kg) allows their integration in aerial platforms, whereas the human-size and human-like design results in a natural replication of the human arms motion. Each of the arms provides four joints for end effector positioning [7], three at the shoulder and the elbow. The end effector consists of a simple passive gripper used for grasping the parcel from a handle. Wrist joints are not necessary, reducing the total mass and inertia. A spring-lever transmission mechanism is introduced between the servo horn and the links to provide mechanical joint compliance, which allows to protect the actuators from impacts and overloads, and provides a certain level of passive accommodation of the aerial robot to interaction wrenches exerted on flight [8]. Figure 3 represents the components and architecture of the aerial manipulator, including the leader-follower kinaesthetic teleoperation interface that allows a human operator to transfer his bimanual manipulation capabilities to the aerial robot, exploiting the human-like kinematics of both LiCAS arms[3].

D. Wheeled-Legged Quadruped

The ANYmal on Wheels wheeled-legged robot [9], [10] seamlessly integrates the benefits of both wheels and legs. This design enables the robot to efficiently traverse long distances and adeptly handle complex urban challenges such as steps and stairs. Given this fusion of range and versatility, the robot emerges as an ideal candidate for last-mile delivery operations, addressing the limitations of conventional wheeled platforms that often struggle in the final meters. To manage its 16 DoFs, the robot employs reinforcement learning to determine the optimal mobility mode. Consequently, the robot can autonomously navigate upstairs and roll across flat terrains without human input. The robot is equipped with a carrier box on its top where the aerial manipulator will drop the parcel.

[3]https://www.youtube.com/watch?v=Goya5JNdnqA

E. Rollin' Justin

Rollin' Justin, an advanced wheeled humanoid robot by DLR [11], features human-like dexterity and mobility. Its core components include two DLR lightweight arms with DLR-II hands that are interconnected via a lightweight torso. With 44 controllable DoFs, it can reach objects on various surfaces. An RGB-D sensor and a stereo camera pair in its head allows the robot to perceive its environment and enable remote operation.

Supervised autonomy, a core element of Rollin' Justin control framework, involves equipping the robot with a certain level of decision-making capabilities [12]. Thus, Justin can perform tasks autonomously based on predefined code snippets, so-called Action Templates [13], while still maintaining oversight and intervention from human operators. This approach aims to strike a balance between human guidance and robotic autonomy, effectively streamlining task execution and enhancing operational efficiency.

By integrating tactile feedback mechanisms into the control interface, operators can feel and respond to the environment in real time [14]. This not only enhances their situational awareness but also enables precise manipulation of objects and interaction with the surroundings. The combination of supervised autonomy and direct haptic telepresence revolutionizes the way humans interact with the remote robot, offering a seamless blend of human expertise and robotic capabilities [15]. This is especially relevant for the deployment of robots in partially unknown environments as it is the goal of euROBIN. The application of this technology was tested for the first time in a terrestrial scenario during the hackathon as described in Section IV.

F. Mobile Manipulator-1

The IST TIAGo depicted in Figure 4 (denoted as TIAGo-1) is a customized version of the TIAGo Steel robot developed by PAL Robotics, improved for the domestic environment. The IST TIAGo is a mobile manipulator robot with a differential-drive base, a lifting torso, a 2-DoF pan-tilt head, and a 7-DoF arm with 3 kg maximum payload.

In indoor domestic scenarios, the two Hokuyo laser range-finders mounted in the base and a downwards-facing RGB-D camera in the head provide adequate localization and navigation with 3D obstacle avoidance. A second Orbbec Astra S RGB-D camera mounted in the head enables object detection, 6D pose estimation, and object tracking. A ReSpeaker microphone array with noise canceling and an 8W speaker enables human-robot interaction via speech.

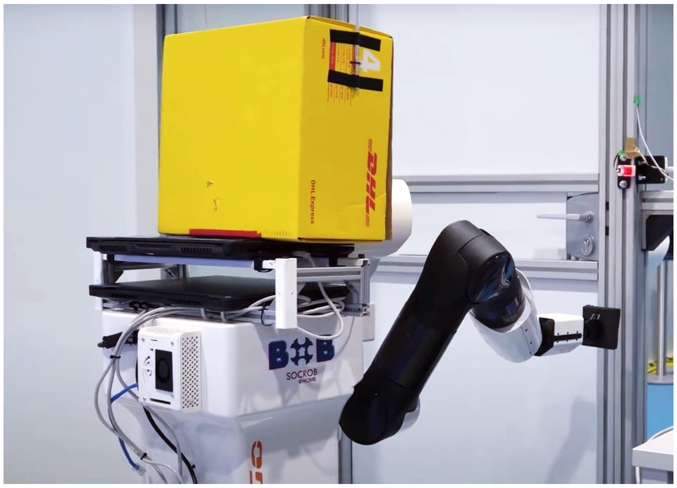

Fig. 4. IST TIAGo-1 carrying the parcel inside the house scenario.

Fig. 5. INRIA TIAGo-2 teleoperated with the HTC Vive 6-DoFs pose controller to open the door.

The implemented algorithms exchange data through ROS1 and are part of a larger framework that uses Petri nets for knowledge representation, task planning, and execution.

G. Mobile Manipulator-2

The INRIA TIAGo robot shown in Figure 5 (denoted as TIAGo-2) is a single-arm TIAGo Steel robot from PAL Robotics. It uses a custom software stack for direct teleoperation, utilizing both proprioceptive and visual feedback. The arm, head, and gripper joints operate in position-control mode, while the two wheel joints on the mobile base are velocity-controlled. A camera on the robot's head captures visual feedback. To mitigate the latency due to the video stream encoded in H264, the camera was connected to an Nvidia Jetson Nano board, which employs hardware acceleration to encode the video stream in H264 before transmission.

On the operator's side, the human commands the robot hand 6-DoF pose using an HTC Vive controller that is tracked with millimeter accuracy by 2 laser-based "lighthouses". The mobile base navigation commands are input using a 3D Mouse by 3DConnexion. The operator views video feedback on a laptop screen and the robot's proprioceptive state is represented using a 3D model rendering.

The robot and operator's station are linked via a local WiFi network, employing either 2.4 GHz or 5 GHz channels to mitigate network congestion at the test site. The video stream uses gstreamer, which offers a versatile video pipeline for encoding and decoding. Communication for all operator commands, proprioception, and visual feedback relies on the UDP transport protocol, ensuring low-latency real-time communication within a client-server architecture detailed in [16].

H. ARMAR-6

ARMAR-6 [17] is a high-performance humanoid robot designed for human-robot collaborative tasks. It is equipped with versatile sensorimotor and cognitive capabilities integrated into a functional architecture allowing its use in various real-world scenarios. In particular, the robot integrates perception, mobile manipulation, compliant motion execution, grasping and manipulation of heavy objects, among others. Moreover, as ARMAR-6 was initially developed with the aim of providing a second pair of hands to support a human worker, it evolves in human-centered environments and can manipulate a wide range of everyday objects. It is also equipped with various cognitive abilities ranging from natural language understanding, reasoning about spatial object relations, recognition of human actions and intentions and learning task models from human demonstrations and observations.

ARMAR-6 has 27 actuated DoFs and features an anthropomorphic upper body with 8-DoF torque-controlled arms and two underactuated five-finger hands. The two-DoF head includes a visual perception system comprising an Azure Kinect RGB-D, a Roboception rc_visard 160, and a wide-baseline passive stereo camera system. The dual-arm system together with the height-adjustable torso results in a workspace of $10.7\,\mathrm{m}^3$ and a maximum height of $192\,\mathrm{cm}$. Its holonomic platform hosts battery packs, power management systems, two laser scanners, and four computers.

The interaction of numerous software components on ARMAR-6 is facilitated by the robot software framework ArmarX [18], which also allows the seamless integration and interchange of third-party contributions. ArmarX provides a three-level functional cognitive architecture [19] consisting of (1) a sub-symbolic low level (e. g., sensorimotor control), (2) a symbolic high level (e. g., language and scene understanding, task planning and execution monitoring), (3) and a mid-level in the form of a memory system which mediates between the low-level and high-level abilities.

IV. ROBOTIC FUNCTIONALITIES

This section describes the functionalities and methods implemented by the different robots to fulfil each of the tasks involved in the door-to-door parcel delivery operation, again, presented in the order of participation of the robots.

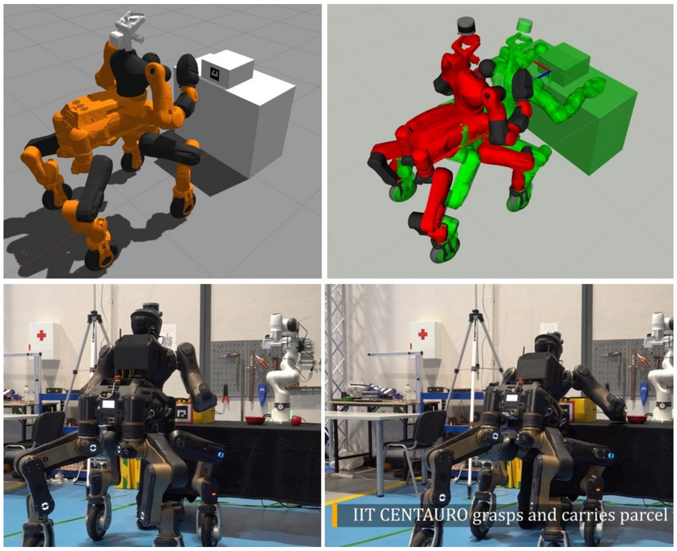

Fig. 6. Generation of whole-body trajectory for Centauro in the bimanual parcel grasping task. Simulated environment with the robot in front of the box placed on a table with the visible Aruco marker used to plan the collision-free goal pose (top), and replication with the real robot (bottom).

A. Parcel Load with Visual Object Detection

The parcel load consists of a sequence of open-loop reach-and-grasp trajectories automatically generated. A simulated scene containing the Franka Emika FR3 arm and some YCB objects is built using Pybullet. An evolutionary algorithm that optimizes both diversity and quality [20] is applied to generate a large set of diverse and high-performing grasps for each considered object. The best-performing trajectories in terms of fitness (defined as a mixture of energy consumption and contact point variance) are deployed into the real world. The code is available on Github[4].

The objects to be placed in the parcel have been positioned similarly to the simulated scene used for the generation of trajectories. The object state alignment has been carried out using point cloud from a RealSense D435i RGB-D camera. A closed-loop approach that uses camera images to automatically adapt the trajectories learned in simulation to object location in the real world was later developed [21]. A human operator selects the objects to be grasped, which determines the trajectory of the robotic arm for grasping the requested items. The end effector is moved to the top of the parcel using standard motion planning, as it can be seen in Figure 8-1). Collision avoidance is done by considering a cuboid bounding box around the object. The flaps of the parcel are then closed by playing back a trajectory recorded through kinesthetic demonstration (see Figures 8-2).

B. Force Controlled Parcel Grasping and Transportation

The Centauro robot was in charge of transferring the parcel from the loading area to the aerial manipulation robot. This robot replicates the human capability of grasping an object while moving in any kind of environment relying on its hybrid wheel-legged nature. Despite the robot is provided with an autonomous locomotion framework [22], the experimental scenario was a simple structured and obstacle-free environment, and the locomotion task has been executed tele-operating the robot using a position controller for the steering and rolling joints. On the contrary, the manipulation task (i.e., grasping the box from the loading station and holding it in an accessible configuration for the aerial robot) was executed in a fully autonomous way. The robot detects the parcel and the loading area through an Aruco marker drawn on one side of the box. Once Centauro reaches an appropriate position in front of the parcel at the supply point, the robot plans a bi-manual approaching trajectory placing the two hands around the box using a whole-body sample-based trajectory generator. The planning framework, dubbed *cartesio planning*[5], augments the PlanningScene from the MoveIt! Framework to consider also the floating-base, thus enlarging the manipulation workspace of the robot [23]. This framework takes as input the start and a goal configurations. The first corresponds with the current robot's configuration, while the second is chosen in such a way that the two arms are placed at a user-defined distance from the two sides of the box. Collisions with the environment are avoided including two boxes embedding the parcel and the loading station in the PlanningScene, starting from the pose of the parcel given by the Aruco marker (see Figure 6). The grasping is accomplished by sending a Cartesian velocity reference trajectory for the hands in the direction opposite to the normal of the parcel's surface through an impedance controller. The motion continues until a contact force threshold is reached, ensuring that the parcel is firmly grasped during its transportation and retrieval by the aerial robot.

C. Aerial Robot Position Control and Teleoperation

The aerial parcel retrieval from the Centauro robot and the parcel drop on the carrier box of the wheeled-legged quadruped was done relying on the position control of the aerial platform and a kinesthetic teleoperation scheme of the anthropomorphic dual arm, as depicted in Figure 3. The flight controller of the aerial platform, implemented by the ArduPilot software, consists of a four-layer cascade controller: angular rate (lowest level), attitude, velocity and position. The state of the platform is obtained from an Extended Kalman filter that takes as input the accelerometer, magnetometer and gyroscope data from the inertial measurement unit (IMU) along with the position measurement from an Opti-Track positioning system available in the flight area. Note that, since the length of the arms is around $50\,\mathrm{cm}$, with an effective reach around $30\,\mathrm{cm}$, it is necessary that the position and control error of the aerial platform is less than 10% so the arms can easily reach the handle of the parcel [24].

The kinesthetic teleoperation interface consists of a joint-to-joint mapping between the leader dual arm (LDA) handled by the human operator, and the follower dual arm (FDA) integrated on the aerial platform. The torque control of the LDA servos is disabled so the human operator can move easily the joints by compensating a small friction of the gearbox. The internal encoder of the servos is used to measure the LDA joint

[4]https://gitlab.isir.upmc.fr/l2g/qd_grasp

[5]https://github.com/ADVRHumanoids/cartesio_planning

position vector used as desired reference, sent to the on-board computer board of the aerial robot through UDP sockets. The FDA servos take as reference position the corresponding LDA position feedback, provided at a 50 Hz.

D. Energy Efficient Wheeled-Legged Navigation

Traditional legged robots are limited in range, typically covering only a few kilometers on a single battery charge, even though flat terrains often do not necessitate stepping. In contrast, during our trials, the wheeled-legged quadruped primarily employed its wheels on predominantly flat terrains, reserving leg actuation solely for obstacles. Previous research [10], [25] indicated that driving the robot can significantly reduce its cost of transport. This fusion of capabilities ensures both efficiency and adaptability, presenting a compelling alternative to other mobility concepts.

E. Parcel Retrieval and Delivery to Robot

Rollin' Justin showcased a combined control approach in a task involving the retrieval of a parcel delivered by the ANYmal on Wheels wheeled-legged quadruped robot and its subsequent handover to TIAGo-1. This mixed control mode encompassed supervised autonomy and telepresence control. Initiating the task in supervised autonomy mode, Rollin' Justin autonomously approached the ANYmal on Wheels using predefined Action Templates for the localization and lifting of the box, efficiently retrieving the parcel. This setup enabled the robot to interact with its environment with minimal human intervention, streamlining the parcel retrieval process as it is shown in Figure 8 tile 7. Transitioning to telepresence control mode, the operator assumed direct control over Rollin' Justin. The robot was maneuvered towards TIAGo-1 in preparation for transferring the parcel. The placement of the parcel was conducted by using haptic telepresence control. This mode allowed precise manipulation of the parcel, showcasing the benefits of real-time force feedback as the parcel had to be placed securely on an elevated surface on top of the TIAGo-1 robot as it is visible in tile 8 of Figure 8. This accomplishment demonstrates the potential of combining supervised autonomy and telepresence control for seamless human-robot collaboration. The successful handover of the parcel underscores the practical application of this mixed control approach.

F. Navigation and Doorbell Ringing

The TIAGo-1 robot navigates fully autonomously through the hallway to the door of the customer's home, allowing the robot to move in a known environment with dynamic obstacles. This capability is based on the open-source move_base ROS package, using a Dijkstra planner and a PAL robotics guidance algorithm built and tuned to the TIAGo robot. The robot localizes itself in a previously obtained occupancy grid 2D map using the odometry of the wheeled base and LiDAR data. This sensor data is used by an Adaptive Monte Carlo Localization algorithm to maintain a probabilistic estimate of the robot's pose that can cope with changes in the environment and odometry errors. Additionally, dynamic obstacles are added to the local map of the environment to avoid collisions. These dynamic obstacles are obtained from LiDAR readings but also from the 3D data provided by the RGB-D camera of the robot. The 3D points are filtered to detect obstacles above the ground. To do this, a region of the pointcloud is converted to the 2D plane and fed as obstacles to the navigation pipeline. The obstacles update the environment global map used in the path planning and guidance steps, allowing the robot to avoid obstacles in real-time.

Doorbell detection and segmentation are obtained using the Detectron2 model [6] trained in a small doorbell dataset acquired at the home area represented in Figure 1. Then, the doorbell's 3D position is estimated from the depth image, using the doorbell mask and the Polylabel algorithm [7] to find the mask center. The 3D pose estimate is then converted to a 6D pose estimate by adding an orientation orthogonal to the doorbell plane. Finally, the robot performs a three-step movement to ring the doorbell. First, the end-effector is moved to a pose identical to the object pose but shifted by 10 cm along the axis orthogonal to the doorbell plane. Then the end-effector is moved to the doorbell pose and moved back 2 seconds later. The inverse kinematics is computed using MoveIt! [26], which also enables collision avoidance of obstacles perceived using the camera depth sensor. The end-effector was endowed with a passively compliant material to prevent excessive force when pushing the doorbell button.

G. Door Handle Opening

The mission of TIAGo-2 robot is to open the sensorized door to allow the delivery robot, TIAGo-1, to enter the mockup home environment. Direct teleoperation is effective in addressing contact-rich loco-manipulation tasks, leveraging the problem-solving abilities of humans who understand, for instance, how to utilize the entire geometry of the robot as a potential contact surface.

A whole-body controller calculates onboard commands for all position-controlled joints at 100 Hz. This controller is based on a sequential quadratic programming formulation [27] to retarget the hand pose command into the robot's morphology in real-time. At each time step, a QP problem is solved to compute changes in the desired posture and joint torque configuration using the QuadProg[8] solver. To enhance safety and robustness against potential operator errors, QP's inequality constraints are utilized to enforce both joint position and maximum actuator torque limits while remaining resilient to kinematic singularities. All commands issued by the operator are subjected to low-pass filtering to eliminate potential noise introduced by data packet loss, and the velocity and acceleration of the target motion are bound .

The TIAGo-2 robot uses position-controlled joints, but the current limits of the actuators can be dynamically set and changed. By tuning manually these parameters, this feature enables force-like control to be applied to the gripper fingers and the hand's wrist. This scheme enhances resilience to minor command errors that may occur when opening the door.

[6]https://github.com/facebookresearch/detectron2
[7]https://github.com/mapbox/polylabel
[8]https://github.com/stack-of-tasks/eiquadprog

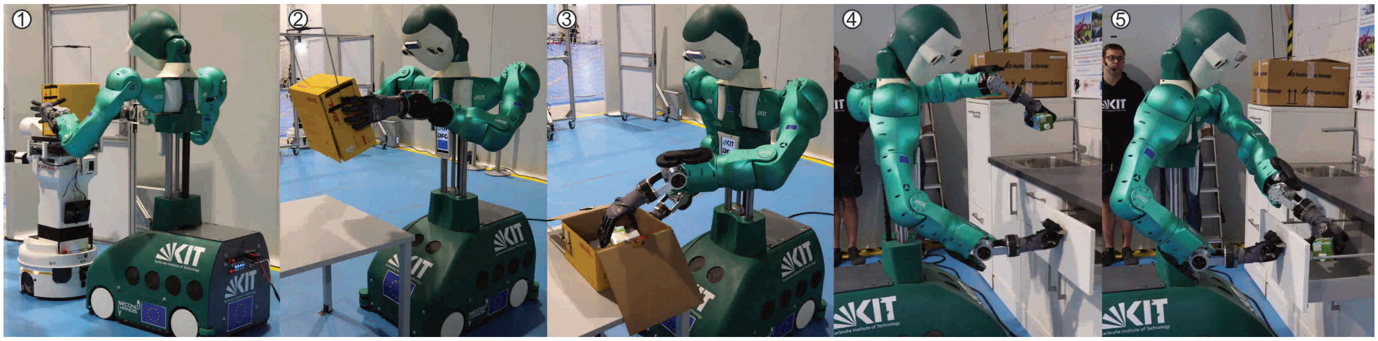

Fig. 7. Parcel unboxing with the humanoid robot ARMAR-6. ARMAR-6 receives the parcel from TIAGo-1 ① and places it on the table ②. It then proceeds to unpack the contents of the box by grasping the unknown objects from the opened parcel ③. ARMAR-6 then drives to the kitchen counter, opens a drawer ④, and places the grasped object inside ⑤. All of these tasks, except for opening the parcel, were performed autonomously without human intervention.

H. Parcel Unboxing

At the last stage of the parcel delivery operation, ARMAR-6's goals were to retrieve the parcel from the back of the TIAGo-1 mobile manipulator, place it on a table, unbox it, and arrange its contents in a kitchen drawer. Figure 7 displays an overview of this last stage. It is important to emphasize that the entire process was achieved *autonomously* by the robot, i.e., *without human intervention*. The only exception was the opening of the parcel with a cutter, for which the robot was assisted by a human due to the complexity of the task.

The last stage of the delivery operation started localizing the parcel. To do so, several state-of-the-art algorithms were integrated into the vision system of ArmarX on the robot, thus allowing the robot to perceive and locate the different objects present in a scene at any time. Specifically, we used GroundingDINO , an open-set object detection model to detect known objects and then tracked their masks using Segment and Track Anything (SAM-Track). Then we leveraged UniMatch to estimate depth maps from the stereo camera of the robot, which are further utilized to obtain the segmented point cloud of each object. The UniMatch model delivers fine-grained depth estimation of thin and reflective handles, allowing the downstream opening drawer tasks. The Dense Object Net (DON) estimates dense correspondences between categorical object instances. We leveraged its image features to detect the corners of the parcels and the handles. The corners of the parcel determine a region on the RGB image, from where we sampled pixels as prompts to query masks of unknown objects using SAM-Track and obtained their segmented point clouds using UniMatch. The DONs were trained on each known object category (e.g., parcel and handle) individually using a dataset generated by a set of Instant Neural Graphic Primitives, which were also trained on the real-world objects.

After localizing the parcel, its retrieval by ARMAR-6 was achieved by leveraging a navigation system to reach the desired location, as well as a learning-from-demonstration strategy and a bimanual impedance controller to lift the parcel. To determine its global pose, the robot uses a graph-based localization method to localize itself in a previously recorded map using 2D laser scanners. The navigation system then takes the surrounding obstacles into account and enables the robot to reach the desired location in a collision-free manner. The trajectory for reaching and holding bimanually the parcel was learned from human demonstration by kinesthetic teaching and represented as a task space via-point movement primitive (VMP) [28] for each hand. The goals of the VMPs were adapted based on the current pose and dimension of the parcel, obtained by fitting a bounding box in the segmented point cloud of the vision system. We used a bimanual impedance controller to execute the learned hand trajectories and leveraged the aforementioned navigation system to bring the parcel to the next location. After placing the parcel on the table, the robot asked a human for assistance using the natural language dialogue system. The human opened the parcel with a cutter and gave the robot a speech cue, to which the robot responded that it will unload the parcel.

To unload the parcel, the affordance-based mobile manipulation framework *MAkEable* [29] was used. Grasp hypotheses for unknown objects were autonomously generated based on object-oriented bounding boxes for a segmented point cloud. Each hypothesis was checked for feasibility — is it reachable, or would the hand collide with the package while executing the grasp? — and then ranked based on multiple heuristics, e.g., the height of the hypothesis or the distance of the fingers to other objects. The best grasp candidate was then selected for execution on the robot.

To store the successfully grasped objects, we leveraged an integrated mobile manipulation system [30], allowing the robot to navigate to the kitchen, detect drawers and door handles using the aforementioned vision pipeline, and open drawers and doors. The latter was achieved with an impedance controller that takes the detected handle and given kinematic model of the kitchen into account. First, the robot established contact with the drawer using a force-based strategy and slid its end-effector upwards until it reached the handle. After grasping the handle, an end-effector trajectory was derived and compliantly executed to reach the desired state of the drawer. After opening the drawer with one arm, the robot used the knowledge of the location of the drawer to generate placement hypotheses inside the drawer. It followed the same procedure as for the grasp hypotheses, including the execution, thus completing the task of unboxing the parcel.

I. Inter-Robot Communication

Collaborative multi-agent tasks require communication — this applies to both humans and robots. To coordinate the ordered execution of the joint experiment, an instance of the episodic memory [19] of the ArmarX cognitive architecture was provided by KIT and served as a communication interface for all project partners. To allow for a programming-language-agnostic and easy-to-use interface to the memory system, a RESTful API was implemented, and documentation and code examples were provided to the partners. Additionally, individual authentication tokens were handed to the partners, allowing them to query the general task execution state, and communicate their own state. With this infrastructure, each robot was able to wait for a trigger to start its own part of the task and to communicate the status of their part to others. The initial trigger was given by a human.

V. Joint Experiment

The execution of the joint experiment involving the eight robots can be followed in the image sequence shown in Figure 8 and in the published video[9], attached also with this paper. The total duration of the experiment is around 18 minutes, as detailed in Table III, comprising the phases described in the following paragraphs.

TABLE III
EXECUTION TIME OF THE TASKS INVOLVED IN THE EXPERIMENT.

#	Robot	Task	Time [s]
1	Franka Emika	Parcel preparation	300
2	Centauro	Approach and grasp parcel	70
3	Centauro	Carry parcel to flight area	70
4	Aerial Manip.	Take-off and grasp parcel	50
5	Aerial Manip.	Drop parcel on quadruped	50
6	W/L Quadruped	Carry parcel to Rollin' Justin	30
7	Rollin' Justin	Deliver parcel to TIAGo-1	130
8	TIAGo-1	Carry parcel and ring doorbell	40
9	TIAGo-2	Open door from inside house	50
10	TIAGo-1	Pass through door, turn back	40
11	ARMAR-6	Take parcel from TIAGo-1	60
12	ARMAR-6	Put parcel on kitchen table	40
13	ARMAR-6	Retrieve/put objects in drawer	130

A. Parcel Preparation with Industrial Robotic Arm

The automatically generated reach-and-grasp trajectories were successfully transferred into the real world during the realization of the joint tests without requiring any retrial. The pick-and-place pipeline allowed the industrial manipulator to load the three objects into the parcel and close its flaps, making it ready to be collected by the Centauro. Roughly speaking, the success rate in the object grasp and drop task was higher than 90 % for about 50 trials, closing successfully the flaps of the box around 85 % of the trials in approximately 30 tests. Failures occurred due to the variance of flaps angle, and their interaction with the gripper's fingers. It is worth noting that the speed of deployment is not informative here, considering that the used method consists of open-loop reach-and-grasp trajectories which are adapted based on the 6-DoF pose of the targeted object [21]. In practice, the grasps can be completed as fast as allowed by the limits of the manipulator.

[9]https://www.youtube.com/watch?v=vrRwY7f0g8I

B. Parcel Grasping and Transportation

After being remotely operated to reach the vicinity of the supply point, Centauro identifies the parcel by detecting an Aruco marker attached to the visible side of the box. Then, it autonomously plans a whole-body trajectory to safely grasp the package using both hands, following the methodology described before. Once the parcel is grasped with the use of force control, the Centauro reverts to its nominal configuration and moves to the unloading area where the aerial manipulation robot will grasp the box directly from the robot's hands. To facilitate the drone's grip on the package, it adjusts the hook-like gripper upwards, and the grip is released once the transfer of the package has been successfully completed.

The aerial manipulation robot, controlled in position using the Opti-Track system, takes off and approaches the Centauro robot until the handle of the parcel is within the reach of the hook-like grippers. The human operator guides both arms towards the handle using the leader-follower teleoperation system with direct visual feedback. During a few seconds, both the Centauro and the aerial manipulator are handling the same object, relying on the mechanical joint compliance of the aerial robot to overcome the forces exerted due to the small position deviations of the aerial platform. When the Centauro releases the parcel, this is held by the aerial robot, which approaches then to the ANYMal on Wheels quadruped to drop the parcel on the carrier box, relying again on the coordination between the human pilot and the arms operator.

C. Parcel Delivery to User's Home

The quadruped is then guided towards Rollin' Justin through the simulated outdoor area combining the wheeled and legged locomotion to overcome the soft floor that makes the navigation slightly difficult for conventional wheeled platforms. In order to facilitate the grasping of the parcel carried at its back, the quadruped is rotated in yaw to give its back to Rollin' Justin at the borderline between the simulated outdoor area and the hallway (see Figure 1). The humanoid robot approaches then to grasp the parcel with its right hand, combining the motion from the arm and torso to carefully place the load at the back of TIAGo-1. This follows a task plan implemented as a state machine and executed autonomously. The events signaling that the parcel was placed on and picked from TIAGo-1 are received through the ArmarX RESTful API that enables multi-robot coordination. The door open or closed status is perceived through the LiDAR on the TIAGo-1 base.

The TIAGo-2 robot, initially idled inside the mockup home environment, opens the door for the delivery robot when TIAGo-1 pressed the doorbell button. The robot was teleoperated with low-latency visual feedback to navigate towards the closed door, where it grasped, rotated, and pulled the door's handle to partially open it. Due to limited arm workspace, the operator repositioned the robot's base and used the gripper and forearm to fully open the door from pushing on the other side of the door.

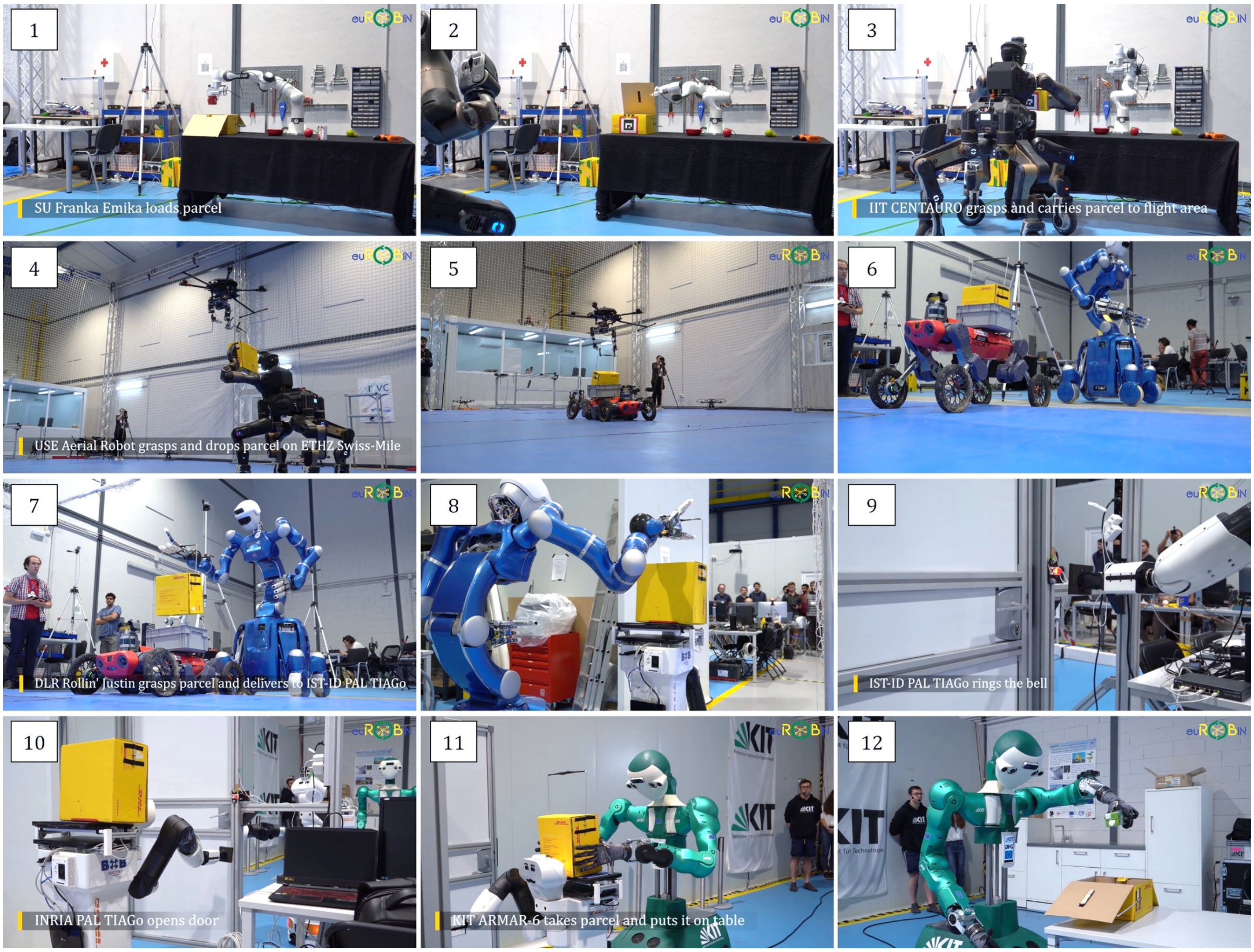

Fig. 8. Sequence of images of the complete experiment.

D. Parcel Unboxing

Once the door was perceived as opened, TIAGo-1 entered inside the mock-up house. The event was signaled through the ArmarX RESTful API, triggering ARMAR-6 to approach TIAGo-1 and to localize the parcel. ARMAR-6 then grasped the parcel using the approach described in Section IV-H (see tile 11 of Figure 8), transported it, and placed it onto a table. The robot then asked a human for assistance to open the parcel using the natural dialog system. The human opened the parcel with a cutter and gave the robot a speech clue to signal the completion of the task. ARMAR-6 then unloaded the parcel using grasp hypotheses for unknown objects, as described in Section IV-H, and ordered them in a kitchen drawer using the navigation system, impedance controller, and placement hypotheses described in Section IV-H too. Tile 12 of Figure 8 shows ARMAR-6 having successfully grasped an unknown object from the parcel.

VI. Lessons Learned

The following paragraphs summarize the lessons learned from each of the teams participating in the hackathon.

A. Lessons Learned from Franka Robot Team (ISIR-CNRS)

The software module employed in the hackathon for loading the parcel relies on the assumption that the 3D models of the objects are known and that the object-gripper interaction can be simulated [20]. This method provides adaptation capabilities for grasping objects within industrial scenarios, in which object variability is limited. However, it is less suitable in scenarios or applications with a wider diversity of objects (e.g., e-commerce, recycling). Rigid and semi-rigid objects were considered, as learning-purpose simulators model well the dynamics of the physical interaction.

Beyond those particular limitations, the proposed approach also accounts for other constraints in this robotic collaborative scenario. The box size had to be graspable by the Centauro. Objects were selected to match gripping constraints: small enough to be graspable by the FR3 gripper, but large enough to be graspable by the ARMAR-6 hand.

B. Lessons Learned from Centauro Robot Team (IIT)

Hybrid locomotion, combining wheels and legs, presents numerous challenges for a robot with the size of Centauro, even within a controlled indoor environment. The robot has been teleoperated acting on the wheel steering and rolling joints to follow a reference velocity commanded by the pilot while the rest of the body is kept fixed. The experimental setup took place in an aerial laboratory, which differed significantly from a conventional robotics lab in terms of its characteristics and requirements. The flying area was equipped with a thick foam surface to cushion potential falls of drones and prevent damage. Transitioning from the solid concrete flooring to the foam carpet of the flying area resulted in a significant increase in friction for Centauro's rubber wheels. This led to internal stresses at the leg joints with a consequent potentially dangerous increase of joint's driver temperature, particularly pronounced due to the robot's long legs being more susceptible to disturbances originating from the wheels. In view of experiments in realistic outdoor environments, we need to take into account this issue by designing a whole-body control strategy able to minimize internal stresses coming from a too stiff interaction with the environment.

C. Lessons Learned from Aerial Manipulator Team (USE)

The aerodynamic downwash effect exerted by multi-rotor propellers over the cardboard parcel had a significant impact on the realization of the parcel grasping and drop-on while flying. The first three preliminary trials for grasping the parcel located in a table failed because the airflow generated by the propellers made it blow away when the multi-rotor was at less than 1 m distance. This effect is worse as the parcel is lighter and its surface area larger. The contact force control implemented by the Centauro robot during the bimanual grasping of the parcel was particularly useful to avoid this problem during the aerial grasping phase. The parcel drop operation failed in another test because the parcel was empty (a 0.3 kg load was used before), so the airflow caused excessive oscillations due to the pendulum effect when it was held by the arms at its handle.

The LiCAS dual arm teleoperation interface was an effective solution to implement quickly the grasping task without requiring onboard image processing, reducing consequently the system complexity and involved payload. However, it would have been convenient to incorporate a first-person view (FPV) camera, typically used in racing drones, to make the task easier for the human operator, since for a visual line of sight above 5 m distance it becomes difficult to distinguish the end effectors of the robot and the handle of the parcel. The video of the experiment (see minute 1' 12") shows that the right arm does not reach correctly the handle on the first try. No risky situation occurred in any flight test, despite the close physical interaction with the Centauro and ANYmal on Wheels robots during the parcel load and drop-on. The accurate positioning provided by the OptiTrack system in the indoor testbed, along with the reliability of the multi-rotor platform and the mechanical joint compliance provided by the LiCAS dual arm were essential in this sense.

D. Lessons learned from ANYmal on Wheels Robot Team (ETHZ)

The evaluation of the energy consumption during the multi-robot delivery hackathon confirmed higher speed and lower cost of transport compared to legged robots, while the machine could still overcome all relevant obstacles [25]. The pre-mapping of the outdoor area and the creation of a digital twin simulation played a critical role in preparing the robot's mission. To reduce the preparatory work in future sites and scenarios, we will work on the integration of existing maps and robotically-assisted generation of digital twins. Employing a robust reinforcement learning-based controller enabled safe locomotion and navigation alongside other robots, highlighting the significance of intelligent decision-making in dynamic environments.

E. Lessons learned from Rollin' Justin Robot Team (DLR)

An essential takeaway from the robotics hackathon is the critical role of safety in multi-institutional, heterogeneous robotic interactions. Each developer and researcher inherently aim to safeguard their robots, often achieved by maintaining a safe distance between the systems. However, this approach becomes untenable when robots are required to interact closely. Consequently, it necessitates the implementation of mechanisms that foster trust among developers, ensuring that one robot will not inadvertently damage another. Furthermore, direct communication between the robots is needed to allow safe motion generation and task execution while being aware of the nearby robots. During the hackathon, a strategy of sequential robot actions was adopted to ensure safety. In this approach, only one robot assumed an active role during an interaction, while the other remained passive. This method, while effective in maintaining safety and reducing the complexity of the hackathon scenario, limits the potential for truly parallel action execution and future scalability. As a lesson learned, we recognize the need for further work in inter-robot information exchange to facilitate simultaneous actions while preserving safety.

F. Lessons learned from TIAGo-1 Robot Team (IST)

When multiple robots collaborate autonomously toward a common goal, they must exchange information and maintain their internal states synchronized. Currently, a standardized approach to knowledge representation is lacking, resulting in each robot employing its unique representation. Consequently, additional engineering efforts are required to enable information exchange between robots. This hackathon underscored the need for standardized knowledge representation, alongside emphasizing the significance of enhancing code reusability and transferability to ensure faster progress within this field. Nevertheless, the implementation of knowledge exchange between robots must not strip away redundant functionalities and perception capabilities of collaborating robots. It was evident that removing such redundancies could introduce vulnerabilities by creating single points of failure during cooperative actions. For instance, when transferring the parcel between robots, mutual confirmation of successful execution by both parties is essential to enhance execution robustness. The results obtained from multiple runs showed no misclassification in the doorbell recognition. Moreover, the autonomous navigation pipeline proved to be robust enough to handle the Hackathon's highly dynamic environment. Despite these achievements, the task still failed 20% of the time because of inadequate pose estimation of the doorbell.

G. Lessons Learned from TIAGo-2 Robot Team (INRIA)

Regarding teleoperation, human control still outperforms autonomous methods in adaptability, performance, and speed. Especially, humans leverage their understanding of the physical world dynamics and contact geometry to perform tasks more effectively, such as using the robot forearm to push and open the door. However, this method inherits classical teleoperation disadvantages and difficulties. Direct line of sight, though simple for local operations, requires mental adaptation from the operator, often hindered by occlusions and limited visibility of the robot's end effector. Remote teleoperation depends on high-quality visual feedback from a robot-mounted camera, but this is often hindered by video transmission issues over busy WiFi signals. Balancing stream compression to reduce latency and stream bandwidth to avoid network congestion is challenging. Even with tools like Gstreamer and NVIDIA Jetson embedded computer, setting up a hardware-accelerated video streaming pipeline remains complex and not user-friendly. Teleoperation performance also depends greatly on the input device. When using VR controllers to teleoperate the INRIA TIAGo, unreliable tracking at workspace edges resulted in potentially dangerous discontinuous commands when tracking was lost.

H. Lessons Learned from ARMAR Robot Team (KIT)

Concerning the last stage of the delivery operation and parcel unboxing, the overall perception and mobile manipulation frameworks implemented on ARMAR-6 generally allowed the robot to autonomously handle the sequence of tasks at hand. In particular, the vision pipeline proved to be highly accurate in detecting and localizing the objects of interest in the scene. Most failures observed during the hackathon occurred while grasping objects in the parcel. These were mostly since several objects were cluttered together in the box or located close to the border of the parcel, resulting in difficulties in grasping them individually from the top. This highlights the need to develop more complex manipulation strategies to grasp cluttered objects in boxes. Such strategies could involve, e.g., separating the objects before grasping them or leveraging different types of grasping beyond power grasps.

Overall, the RESTful API developed as a robot-to-robot communication interface during the hackathon proved to be an effective and efficient way to exchange knowledge between robots. Key advantages of this API are that it can easily be integrated into almost any system and is easy to use. Although the RESTful API was only used to communicate the current status of the task across robots, it may also be leveraged to transfer additional knowledge in the future. For instance, robots could exchange knowledge on their environments or on the objects to be interacted with, among others.

I. General Lessons Learned

Although the hackathon involved three robots intended for outdoor operation (the aerial manipulator, the Centauro, and the wheeled-legged quadruped), the celebration of the event in an indoor facility avoided inconveniences related to weather conditions (rain, wind, sunlight) and promoted the close collaboration between the different teams. However, this required a facility with sufficient space for all participants and robots.

Some robots suffered mechanical damage in their joints due to the transportation by truck or due to the soft floor of the flight arena that caused overloads to the wheeled bases. However, the modular design of the actuators made it relatively easy and fast to replace the damaged units. Battery charging was also carefully scheduled between all teams to avoid electrical overload.

The results collected in Table III relative to the execution time of the different tasks carried out by the robots evidence the necessity to improve the performance time in general, particularly in those tasks involving manipulation. Note however that the definition of the tasks was done during the hackathon, so there was no previous preparation, but the teams had to provide their solutions during the celebration of the event.

Also, as it typically occurs in these events, the interference between the multiple WiFi networks used to connect the robots with the ground control station computers along with the additional interference caused by the mobile data of the participants' phones made highly convenient the use of a common wireless network for the robots with properly assigned channels. During the realization of the final experiment, it was necessary to ask participants and the public to switch their mobiles to flight mode to overcome this problem.

VII. Conclusion and Future Work

This paper reported the platforms, functionalities, and achievements of the robotics hackathon celebrated in Seville on May 2023 in the context of the European Robotics and AI Network (euROBIN), in which eight different robots (one industrial robotic arm, two humanoids, two mobile manipulators, a Centauro robot, a wheeled-legged quadruped, and an aerial manipulator) conducted cooperatively the transportation of a parcel from a supply point to the users' home. The final experiment, taking 18 minutes, was executed without direct human intervention, but using teleoperation in some of the tasks, e.g., the aerial handover, and opening of the door handle.

The work to be done in the next years within the euROBIN project will be focused on exploring the transferability of knowledge and skills between robots relying on the European Robotics Core Repository (EuroCore) that is currently under development. This repository will contain data generated by robots and software modules implementing functionalities that can be adopted by different robotic platforms. In particular, robots should be able to share and reuse previously-generated maps of the environment. This knowledge sharing will be useful for navigating in outdoor scenarios for logistics applications like the one considered in this paper.

Further obvious candidates for transferability and reusability are object detection and localization, world models of the environments, including objects localized by different robots, as well as motion and grasp plans. The next hackathon will focus on promoting and evaluating the degree to which transferred data and knowledge between robots contributed to reducing the programming and execution time of the overall tasks.

REFERENCES

[1] N. Boysen, S. Fedtke, and S. Schwerdfeger, “Last-mile delivery concepts: a survey from an operational research perspective,” *Or Spectrum*, vol. 43, pp. 1–58, 2021.

[2] C. Chen, E. Demir, Y. Huang, and R. Qiu, “The adoption of self-driving delivery robots in last mile logistics,” *Transportation research part E: logistics and transportation review*, vol. 146, p. 102214, 2021.

[3] M. Studley, S. Carter, F. J. Perez-Grau, A. Viguria, G. Ferri, F. Ferreira, D. Nair, S. Schneider, P. G. Plöger, P. U. Lima *et al.*, “European robotics league: Benchmarking through smart city robot competitions,” in *Human-Robot Interaction-Perspectives and Applications*. IntechOpen, 2023.

[4] M. A. Roa, M. R. Dogar, J. Pages, C. Vivas, A. Morales, N. Correll, M. Gorner, J. Rosell, S. Foix, R. Memmesheimer, and F. Ferro, “Mobile manipulation hackathon: Moving into real world applications,” *IEEE Robotics & Automation Magazine*, vol. 28, no. 2, pp. 112–124, 2021.

[5] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, “Robot operating system 2: Design, architecture, and uses in the wild,” *Science Robotics*, vol. 7, no. 66, p. eabm6074, 2022. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.abm6074

[6] N. Kashiri *et al.*, “Centauro: A hybrid locomotion and high power resilient manipulation platform,” *IEEE Robotics and Automation Letters (RAL)*, vol. 4, no. 2, pp. 1595–1602, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8630605

[7] A. Suarez, S. R. Nekoo, and A. Ollero, “Ultra-lightweight anthropomorphic dual-arm rolling robot for dexterous manipulation tasks on linear infrastructures: A self-stabilizing system,” *Mechatronics*, vol. 94, p. 103021, 2023.

[8] A. Suarez, G. Heredia, and A. Ollero, “Physical-virtual impedance control in ultralightweight and compliant dual-arm aerial manipulators,” *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2553–2560, 2018.

[9] M. Bjelonic, R. Grandia, M. Geilinger, O. Harley, V. S. Medeiros, V. Pajovic, E. Jelavic, S. Coros, and M. Hutter, “Offline motion libraries and online mpc for advanced mobility skills,” *The International Journal of Robotics Research*, 2022.

[10] M. Bjelonic, C. D. Bellicoso, Y. de Viragh, D. Sako, F. D. Tresoldi, F. Jenelten, and M. Hutter, “Keep Rollin’ - Whole-Body Motion Control and Planning for Wheeled Quadrupedal Robots,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2116–2123, 2019.

[11] C. Borst, T. Wimbock, F. Schmidt, M. Fuchs, B. Brunner, F. Zacharias, P. R. Giordano, R. Konietschke, W. Sepp, S. Fuchs *et al.*, “Rollin’justin-mobile platform with variable base,” in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 1597–1598.

[12] P. Schmaus, D. Leidner, T. Krüger, R. Bayer, B. Pleintinger, A. Schiele, and N. Y. Lii, “Knowledge driven orbit-to-ground teleoperation of a robot coworker,” *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 143–150, 2019.

[13] D. S. Leidner, *Cognitive reasoning for compliant robot manipulation*. Springer, 2019, vol. 23.

[14] M. Panzirsch, A. Pereira, H. Singh, B. Weber, E. Ferreira, A. Gherghescu, L. Hann, E. den Exter, F. van der Hulst, L. Gerdes *et al.*, “Exploring planet geology through force-feedback telemanipulation from orbit,” *Science Robotics*, vol. 7, no. 65, p. eabl6307, 2022.

[15] N. Y.-S. Lii, P. Schmaus, D. Leidner, T. Krueger, J. Grenouilleau, A. Pereira, A. Giuliano, A. S. Bauer, A. Köpken, F. S. Lay *et al.*, “Introduction to surface avatar: the first heterogeneous robotic team to be commanded with scalable autonomy from the iss,” in *Proceedings of the International Astronautical Congress, IAC*. International Astronautical Federation, IAF, 2022.

[16] A. Sena, Q. Rouxel, E. Ivanova, E. Burdet, and Z. Li, “Haptic bimanual system for teleoperation of time-delayed tasks,” in *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2021, pp. 1234–1239.

[17] T. Asfour, M. Wächter, L. Kaul, S. Rader, P. Weiner, S. Ottenhaus, R. Grimm, Y. Zhou, M. Grotz, and F. Paus, “ARMAR-6: A high-performance humanoid for human-robot collaboration in real world scenarios,” *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 108–121, 2019.

[18] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, “The robot software framework ArmarX,” *Information Technology*, vol. 57, no. 2, pp. 99–111, 2015.

[19] F. Peller-Konrad, R. Kartmann, C. R. G. Dreher, A. Meixner, F. Reister, M. Grotz, and T. Asfour, “A memory system of a robot cognitive architecture and its implementation in armarx,” *Robotics and Autonomous Systems*, vol. 164, pp. 1–20, 2023.

[20] J. Huber, F. Hélénon, M. Coninx, F. Ben Amar, and S. Doncieux, “Quality diversity under sparse reward and sparse interaction: Application to grasping in robotics,” *arXiv preprint arXiv:2308.05483*, 2023.

[21] F. Hélénon, J. Huber, F. Ben Amar, and S. Doncieux, “Toward a plug-and-play vision based graspng module in robotics,” *ICRA’24 MoMa workshop*, 2024.

[22] A. D. Luca, L. Muratore, and N. G. Tsagarakis, “Autonomous navigation with online replanning and recovery behaviors for wheeled-legged robots using behavior trees,” *IEEE Robotics and Automation Letters*, pp. 1–8, 2023.

[23] L. Rossini, P. Ferrari, F. Ruscelli, A. Laurenzi, G. Oriolo, N. G. Tsagarakis, and E. Mingo Hoffman, “Multi-contact planning and control for humanoid robots: Design and validation of a complete framework,” *Robotics and Autonomous Systems*, vol. 166, p. 104448, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889023000878

[24] A. Suarez, V. M. Vega, M. Fernandez, G. Heredia, and A. Ollero, “Benchmarks for aerial manipulation,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2650–2657, 2020.

[25] J. Lee, M. Bjelonic, A. Reske, L. Wellhausen, T. Miki, and M. Hutter, “Learning robust autonomous navigation and locomotion for wheeled-legged robots,” *Science Robotics*, vol. 9, no. 89, p. eadi9641, 2024. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.adi9641

[26] D. Coleman, I. A. Sucan, S. Chitta, and N. Correll, “Reducing the barrier to entry of complex robotic software: a moveit! case study,” *CoRR*, vol. abs/1404.3785, 2014. [Online]. Available: http://arxiv.org/abs/1404.3785

[27] Q. Rouxel, K. Yuan, R. Wen, and Z. Li, “Multicontact motion retargeting using whole-body optimization of full kinematics and sequential force equilibrium,” *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 4188–4198, 2022.

[28] Y. Zhou, J. Gao, and T. Asfour, “Learning via-point movement primitives with inter-and extrapolation capabilities,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4301–4308.

[29] C. Pohl, F. Reister, F. Peller-Konrad, and T. Asfour, “MAkEable: Memory-centered and Affordance-based Task Execution Framework for Transferable Mobile Manipulation Skills,” 2024.

[30] F. Reister, M. Grotz, and T. Asfour, “Combining navigation and manipulation costs for time-efficient robot placement in mobile manipulation tasks,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 9913–9920, 2022.