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

Performance Comparison of Teleoperation Interfaces for Ultra-Lightweight Anthropomorphic Arms

Filip Zorić[+1], Alejandro Suarez[+2], Goran Vasiljević[1], Matko Orsag[1], Zdenko Kovačić[1], Anibal Ollero[2]

***Abstract*— This paper presents a comparative performance evaluation of three different teleoperation interfaces for very low weight (<3 kg) anthropomorphic dual arms intended to conduct complex manipulation tasks involving a certain level of dexterity, accuracy and agility, either in ground service or in aerial manipulation applications. A visual human pose estimation system is developed to obtain the Cartesian and joint values of the user, which are mapped to the corresponding pose of the dual arm manipulator exploiting the equivalent human-robot kinematics. A leader-follower scheme is also presented, using a reduced scale dual arm that can replicate directly the joint positions of the leader arms to the follower arms. A 6-DOF (degrees of freedom) joystick is proposed to generate linear motions more accurately. A total of 60 ground tests were conducted involving 10 participants to determine the accuracy and time performance in two benchmarks (box edges and S contour tracking). Finally, the visual and leader-follower interfaces were evaluated with the dual arm aerial manipulator on flight tests, reporting several findings derived from the system evaluation.**

***Index Terms*— teleoperation interfaces; human pose estimation; anthropomorphic robotic arms;**

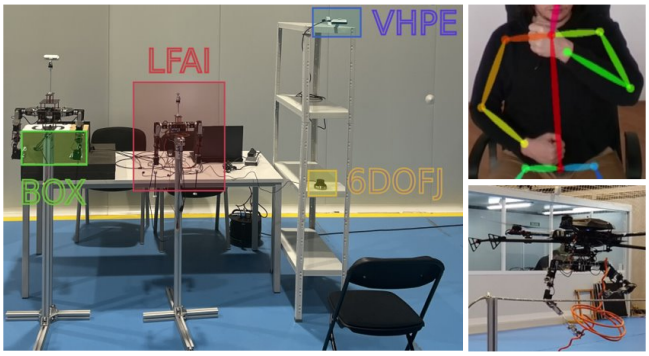

Fig. 1: Experimental setup for evaluating teleoperation interfaces 6-DOF Joystick (6DOFJ), Leader-Follower Arms Interface (LFAI), and Visual Human Pose Estimation (VHPE) (*left*). Torso of the operator during bird diverter setup procedure (*upper right*). Aerial manipulator replicating human operator movement (*bottom right*).

I. Introduction

It is desirable and expected that dexterous manipulation robots are capable to conduct complex manipulation tasks in diverse application domains such as inspection and maintenance, assembly and manufacturing, logistics, or health care and home service, either in ground [1], aerial [2] or space [3] environments. However, despite the significant advances in robot autonomy and manipulation performance, humans are still considered nowadays as the best general purpose manipulator due to the high level of integrated cognition, sensing, perception, and dexterous manipulation skills. In this sense, the adoption of anthropomorphic robotic arms contributes to facilitate the transferability of skills from humans to robots [4], resulting in more natural and intuitive interactions for non expert users intended to use this kind of robots in different tasks. Not only that, but human-like and human-size robotic arms can also benefit from the vast examples of manipulation tasks contained in online videos and learn by themselves to conduct certain operations [5].

In a more practical sense, and focusing on human-like robotic arms, there are three main reasons for exploring and evaluating the possibilities of teleoperation interfaces. First, in some application domains, such as in space [6] or aerial [7] robotic manipulation, it is not possible, feasible or safe for a human operator to conduct a task in the workspace. Second, when the operation cannot be conducted autonomously by the robot due to its complexity, expert human workers can be introduced for replicating the operations or for providing the robot some sample trajectories that serve as reference for further improvement with some machine learning scheme [8]. Third, given the significant differences between human users, it is convenient to experimentally evaluate different interfaces to determine the suitability to different tasks [9].

Although the literature in teleoperation systems for robotic arms is quite extended, most implementations rely on relatively heavy manipulators (the so called "lightweight" industrial manipulator KUKA LBR iiwa weights 24 kg). In order to extend the adoption of dexterous robotic arms in a wider range of applications out of research laboratories, it is necessary to overcome two practical isues: cost and weight. This has motivated the development of lightweight and compliant anthropomorphic dual arm systems (LiCAS), derived from the aerial robotic manipulation field [10], [11], and whose features in terms of very low weight (<3 kg) and mechanical joint compliance result particularly suitable for other ground service operations involving the manipulation of light loads (<1 kg weight). The application of dexterous aerial manipulators in complex and challenging maintenance operations like installation of bird flight diverters in power

+Authors contributed equally

[1]Authors are members of the Laboratory for Robotics and Intelligent Control Systems (LARICS) at the Faculty of Electrical Engineering and Computing, University of Zagreb, Unska ulica 3, 10000 Zagreb, Croatia, e-mail: filip.zoric@fer.hr

[2]Authors are members of the GRVC Robotics Lab from the University of Seville, C. San Fernando 4, 41004 Sevilla, Spain, e-mail: asuarezfm@us.es, aollero@us.es

2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-9190-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IROS55552.2023.10342484

lines [12] is a representative example where intuitive and easily deployable teleoperation interfaces are required [13].

The main contribution of this paper is the development and comparative evaluation of three different teleoperation interfaces for fast deployment, natural and intuitive replication of human movements using very low weight (<3 kg) anthropomorphic dual arm systems capable of conducting dexterous manipulation tasks in ground or while flying when integrated in multi-rotors, as depicted in Fig. 1. These interfaces are Visual Human Pose Estimation (VHPE), Leader-Follower Arms Interface (LFAI), and 6-DOF Joystick (6DOFJ). Rationale for choosing each of the interfaces is following: 6DOFJ is affordable computer peripheral, mainly used in 3D modelling which can be reprogrammed to provide control inputs to the robot arms, therefore, availability, affordability and simplicity are main factors for choosing such interface. LFAI provides easy and intuitive way for the end users to control anthropomorphic arms with the clear idea how should commanded arms configuration look alike. VHPE is choosen because we believe that with easy to deploy interface, humans with little or no training - can use their arms to intuitively command anthropomorphic robot arms. Each of the interfaces is evaluated in two trajectory tracking benchmarks (box and S contours) to compare the accuracy, time performance, and users performance according to the NASA Task Load Index (TLX) indicators [14]. The VHPE and LFAI are evaluated with a dual arm aerial manipulation robot in a bimanual operation conducted on flight.

The rest of the paper is organized as follows. Section II revises the literature in human interfaces for teleoperation. Section III presents the system modeling and design, whereas Section IV focuses on the visual human pose estimation. Experimental results are reported in Section V, providing the conclusions in Section VI.

II. Related work

The idea of replicating human motion with the robot is not new. One of the oldest work of replicating human motion on the humanoid robot can be found in [15]. Zhao et al. presented kinematics mapping and similarity of the humanoid and the robot using motion capture system (MCS), consisting in six cameras, and 38 reverberation markers on actors body. Production of the affordable depth cameras and development of the Human Pose Estimation (HPE)Visual Human-Pose Estimation (VHPE), Leader-Follower algorithms, made such MCS or Body-Machine Interfaces (BoMI) redundant. Ou et al. [16] developed a system imitating human motion captured with kinect with NAO humanoid robot. Human motion mimicking is achieved with the help of normalized vector representation of the human body and the optimization of the error function that compared normalized vector representations of the robot and the human. Paper also introduced collision avoidance and balance maintenance to satisfy constraints imposed by the robot design. Alibeigi et al. [17] presented similar system that maps upper-limb human motion obtained with kinect sensor to the NAO robot.

Lin et al. [18] presented a system that mapped discrete body gestures obtained by Kinect sensor to the 5DOF robot manipulator. Syakir et al. [19] used kinect depth sensor to obtain human pose which was used to directly command 4DOF robot manipulator. Angles determined from the human pose were directly mapped onto the robot manipulator. Bujalance et al. [20] presented real-time gestural control of the 6DOF robot manipulator with human gestures inferred from OpenPose [21] and HMR [22] as basis for human-robot mapping. Only direct geometrical mapping of the human arm was achieved in real-time. Inverse kinematics mapping based on the end-effector position was not possible. Luo et al. [23] presented a system for bimanual teleoperation of 6DOF robot manipulators based on human pose obtained from a Kinect sensor, achieving robot arm motion with on-line trajectory generator paired with Cartesian impedance control.

VHPE interface is an improved version of the human pose control (HPC) interface used to control UAV in a maze scenario as presented in [24]. Similar approach of using human pose estimation to control UAV was presented by Marinov et. al. [25]. Several software modules have been developed that use different gestures to discretly control UAV. Our work differs from presented as it uses human pose estimation to continuously command lightweight manipulators. It also serves as first systematical comparison of the different teleoperation interfaces for the lightweight manipulators.

III. System Modeling and Design

A. System Model

The teleoperation system developed in this work consists of three main elements, as depicted in Fig. 1 and Fig. 2: the human operator, the teleoperation interfaces (VHPE, LFAI, 6DOFJ), and the anthropomorphic dual arm (aerial) manipulator. The following notation will be employed for the human and the robotic arms, as illustrated in Fig. 2:

- $s = \{H, L, F\}$: superscript denoting the human, leader dual arm, or follower dual arm, respectively.
- L_1^s, L_2^s: upper arm and forearm links lengths.
- D^s: separation between left and right arms.
- $i = \{1, 2\}$, $j = \{1, 2, 3, 4\}$: arm (left, right) and joint indices, respectively.
- $\mathbf{q}_i^s = \{q_{ij}^s\} \in \mathbb{R}^4$: joint position vector of the i-th arm.
- $\mathbf{r}_i^s = [x_i^s, y_i^s, z_i^s]^\top$: wrist position of the i-th arm.

The arms provide 4-DOF for end effector positioning in the usual anthropomorphic kinematic configuration: shoulder flexion/extension (q_{i1}), shoulder adduction/abduction (q_{i2}), upper arm lateral/medial rotation (q_{i3}), and elbow flexion/extension (q_{i4}). The forward and inverse kinematics of the anthropomorphic arms, detailed in [11], is denoted as:

$$\mathbf{r}_i^s = \mathbf{FK}(\mathbf{q}_i^s, L_1^s, L_2^s) \;\; ; \;\; \mathbf{q}_i^s = \mathbf{IK}(\mathbf{r}_i^s, L_1^s, L_2^s, \phi_i) \quad (1)$$

Here $\phi_i = q_2^i$ corresponds to the redundant shoulder joint, taken as parameter. This kinematic model is applicable to either the human user, the leader or the follower arms. The

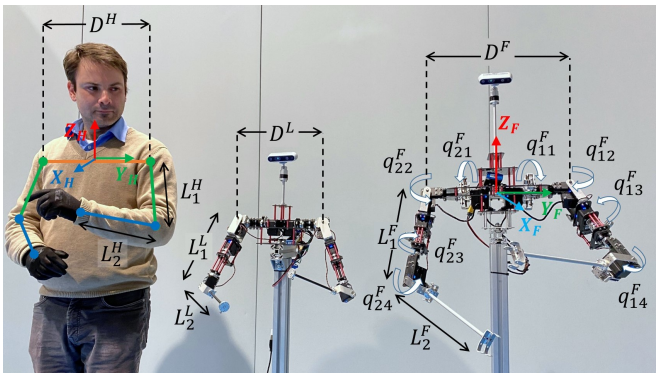

Fig. 2: Links, joint variables and reference frames for the human, leader, and follower anthropomorphic arms.

differences between the human and robotic arms in terms of links lengths and arms separation requires a calibration and scaling procedure, described in Section III-C. A $\{\mathbf{XYZ}_s\}$ reference frame is attached to the shoulder structure of the arms, with the X-axis pointing forwards, the Z-axis pointing upwards, and the Y-axis parallel to the shoulder baseline.

B. System Design

The components and architecture of the teleoperation system developed in this work are depicted in Fig. 3. The three interfaces are connected to the Ground Control Station (GCS) where the two main software modules (the human pose estimation, detailed in next Section, along with the dual arm control program) are executed. We employ an Intel RealSense D415 camera for the VHPE, and a 3DConnexion Space Mouse as 6DOFJ. The dual arm program, developed in C/C++, comprises the low-level arm controller and servo interface, the kinematics methods, and the task manager class where the functionalities are implemented [11]. Data flows from the different interfaces are handled by threads, using a wireless link for sending the references to the follower dual arm that executes in the on-board computer the same software architecture. Two models of lightweight anthropomorphic arms are employed in this system, as shown in Fig. 2. The leader dual arm handled by the user is the LiCAS AC1 model (1.2 kg weight) built with Herkulex DRS-0201 servos, whereas the follower dual arm is the LiCAS A1[1] (2.5 kg) built with DRS-0402/0602 servos.

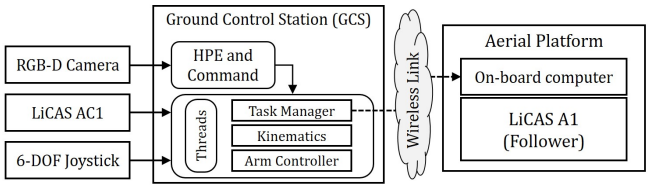

Fig. 3: Architecture of the anthropomorphic dual arm teleoperation system with the three interfaces and aerial platform.

[1]LiCAS Robotic Arms webpage: https://licas-robotic-arms.com/

C. Lightweight Anthropomorphic Arms control

The follower dual arm is built as a chain of smart servo actuators that implement an embedded position/velocity controller with trapezoidal velocity profile generation, as detailed in [10]. In practice, a data packet is sent to each of the arms servos specifying the reference angular position $q^F_{ij,ref} \forall i,j$.

Now, two control modes are applied depending on each type of interface. On the one hand, the LFAI applies directly a joint-to-joint mapping such that $q^F_{ij,ref} = q^L_{ij} \forall i,j$, that is, the joint position in the leader dual arm, read by the servo encoders, is directly set as position reference for the corresponding joint in the follower. Note that in this mode, the torque control of the leader arm servos is disabled so the user can move easily the joints, experiencing only the small friction of the gearbox and control or communication delays (under 100 ms).

On the other hand, the VHPE and 6DOFJ provide motion commands in Cartesian space which are mapped into joint references through the inverse kinematic model given by Eq. (1), that is, $\mathbf{q}^F_{i,ref} = \mathbf{IK}(\mathbf{r}^F_{i,ref}, L^F_1, L^F_2, \phi_i)$.

The 6DOFJ interface provides normalized translational and rotational velocity references in the range -1 to $+1$, denoted as $\mathbf{u}_{6D} = [\mathbf{u}_T^\top \ \mathbf{u}_R^\top]^\top$, where $\mathbf{u}_T = [u_x,\ u_y,\ u_z]^\top$ and $\mathbf{u}_R = [u_\phi,\ u_\theta,\ u_\psi]^\top$ are the XYZ and roll-pitch-yaw commands, respectively. The 6DOFJ is used to command simultaneously the wrist point of the two arms in Cartesian space. Denoting by T_s to the control period (0.02 s), and v_{max} to the maximum speed (0.2 m/s), the Cartesian position reference of the follower arms is given by:

$$\mathbf{r}^F_{i,ref} = \mathbf{r}^F_i + v_{max} T_s [\mathbf{u}_T + (-1)^i \mathbf{R}_{6D} \mathbf{u}_R] \tag{2}$$

where $\mathbf{R}_{6D} \in \mathbb{R}^{3\times3}$ is the joystick rotation map matrix that determines the way the roll-pitch-yaw commands are applied to the Cartesian position of the arms. The $(-1)^i$ term imposes asymmetric motions of the left and right arms due to the action of the rotation component of the joystick. For this interface, the redundant joint (shoulder adduction/abduction) is set to a constant value ($q^F_{i2} = \pm 10°$), considering only the yaw input for increasing/decreasing the Y_F-axis position of the arms (this will be used in the box contour benchmark).

The VHPE interface described in next section provides two references for the follower arms: the Cartesian position of the human wrist point, $\mathbf{r}^H_i$, and the shoulder adduction/abduction angle, q^H_{i2}. As mentioned before, taking into account the possible differences in forearm, upper arm links lenghs and arms separation between human user and robotic arms, the adopted solution consists of performing a calibration process so the motion commands of the arms are relative to an initial nominal L-pose ($\mathbf{q}^{H,0}_i = \mathbf{q}^{F,0}_i = \{0, 0, 0, -\pi/2\}$), applying a scaling factor $\mathbf{\Delta} = diag(\delta_x, \delta_y, \delta_z) \in \mathbb{R}^{3\times3}$ to adapt the displacement of the human arms according to the robot size:

$$\begin{aligned} \mathbf{r}^F_{i,ref} &= \mathbf{r}^{F,0}_i + \mathbf{\Delta}(\mathbf{r}^H_i - \mathbf{r}^{H,0}_i) \\ \mathbf{q}^F_{i,ref} &= \mathbf{IK}(\mathbf{r}^F_{i,ref}, L^F_1, L^F_2, q^H_{i2}) \end{aligned} \tag{3}$$

where $\mathbf{r}^{s,0}_i = \mathbf{FK}(\mathbf{q}^{s,0}_i, L^s_1, L^s_2)$ is the initial calibration pose of the human/robot arms in Cartesian space.

Although one can argue that comparing interfaces that command robot arms in different operational spaces (LFAI in joint space, 6DOFJ and VHPE in cartesian space) can render comparison incomplete, in this instance this is not the case. Mainly due to the system design, small size difference between leader and follower arms doesn't affect operators perception of the difference in between commanded and achieved end effector pose, which in terms doesn't affect operator performance. Especially bearing in mind the human's capability to use visual feedback to correct position of the follower arms as needed.

IV. Visual Human Pose Estimation (VHPE) Teleoperation

The VHPE teleoperation method generates Cartesian and joint references for dual robotic arms based on the RGBD camera image of the human user. VHPE can be divided into human pose estimation, command generation, and filtering.

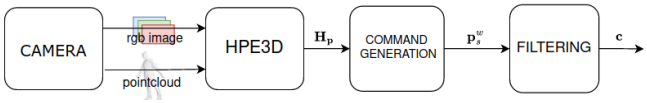

Fig. 4: Scheme of the VHPE teleoperation system

1) Human pose estimation (HPE): OpenPose [21] neural network architecture is used for 2D human pose estimation. Input in the neural network is the RGB image. Output of the HPE is array of detected keypoints $\mathbf{k_p}$, representing the shoulder, elbow, or wrist points:

$$\mathbf{k_p} = [k_{p1} \quad k_{p2} \quad \dots k_{pn-1} \quad k_p]^\top, \quad \mathbf{k_p} \in \mathbb{R}^{n\times 2} \tag{4}$$

which consists of the pixel positions of n detected keypoints in the camera image:

$$k_{pn} = (p_{xn}, p_{yn}) \tag{5}$$

With keypoint pixel positions and its respective measured depth, from the pointcloud we can get $\mathbf{H_p}$ which consists of the keypoint positions in the Cartesian space w.r.t. camera coordinate frame:

$$\mathbf{H_p} = [\mathbf{h_{p1}} \quad \mathbf{h_{p2}} \quad \dots \mathbf{h_{n-1}} \quad \mathbf{h_n}]^\top, \quad \mathbf{H_p} \in \mathbb{R}^{n\times 3} \tag{6}$$

The conversion of the detected keypoints from camera frame to the human operator body frame is done as follows:

$$\hat{\mathbf{H}}_\mathbf{p} = \mathbf{T}_C^H \mathbf{H_p} \tag{7}$$

where $\mathbf{T}_C^H$ represents homogeneous transformation matrix which denotes spatial transformation between camera frame and human trunk frame, and it is determined beforehand.

2) Command generation: Detected shoulder, elbow and wrist keypoints in the Cartesian space w.r.t coordinate frame of the human trunk are used to calculate $\mathbf{p}_s^e$ and $\mathbf{p}_e^w$ which denote vectors from shoulder to elbow and elbow to wrist respectively. We use position vectors for each arm independently. To extract roll angle of the human arm, orthogonal projection of the left and right $\mathbf{p}_s^e$ on the yz plane of the corresponding shoulder coordinate frame is used. Orthogonal projection is calculated as follows:

$${}^s_e\mathbf{p}_{yz} = \mathbf{M_{yz}}(\mathbf{M_{yz}}^\top \mathbf{M_{yz}})^{-1}\mathbf{M_{yz}}^\top \mathbf{p}_s^e. \tag{8}$$

Matrix $\mathbf{M_{yz}} \in \mathbb{R}^{3\times 2}$ is composed of column unit vectors $\mathbf{e}_y = [0,1,0]^\top$ and $\mathbf{e}_z = [0,0,1]^\top$ of the shoulder coordinate frame. Shoulder roll angle is calculated as:

$$\theta_s = sgn({}^e_s p_x) acos\left(\frac{{}^e_s\mathbf{p}_{yz}\cdot \mathbf{e}_z}{\|{}^e_s\mathbf{p}_{yz}\|}\right), \tag{9}$$

and can be thought of as a rotation about shoulder anterior/posterior axis. The end effector position of the follower robotic arms is then emulated as scaled human wrist position w.r.t. shoulder coordinate frame, applying the calibration and scaling procedure given by Eq. (3).

3) Filtering: Command references generated by the VHPE should be filtered to remove outliers due to occlusions and noise from depth camera, implementing for this purpose a Kalman filter that assumes a constant velocity model for the human arms. The state vector comprises the XYZ position and velocity of the wrist point of the human user:

$$\mathbf{x}_i^H = \begin{bmatrix} x_i^H & y_i^H & z_i^H & v_{x,i}^H & v_{y,i}^H & v_{z,i}^H \end{bmatrix}. \tag{10}$$

The measurements vector includes the wrist position given by the VHPE:

$$\mathbf{z}_i^H = \begin{bmatrix} x_i^H & y_i^H & z_i^H \end{bmatrix}. \tag{11}$$

Estimation uncertainty is characterized by process and measurements covariance matrices $\mathbf{Q}$ and $\mathbf{R}$, respectively, assumed to be diagonal. Occlusion of the shoulder or elbow joints caused by the forearm or the hand of the human user may cause incorrect depth reading which results with wrong position vector. To filter wrongly determined position vectors, we measure norm of the $\mathbf{p}_e^w$ and the $\mathbf{p}_s^e$. We divide current norm of the position vectors $\mathbf{p}$ with average of that position vector $\mathbf{a_p}$ obtained during calibration procedure. During calibration procedure, the human user adopts an L-pose with the forearm lifted, so there are no occlusions.

$$f_p = \frac{\|\mathbf{p}\|}{\|\mathbf{a_p}\|}. \tag{12}$$

If f_p is greater than 1.5 the measurement is omitted, otherwise, the measurement is kept. Occlusions present greatest limitation of the system. Norm filtering prevents unwanted reference jumps caused by the occlusions.

V. Experimental validation

In order to compare the three teleoperation interfaces (6DOFJ, LFAI, VHPE) we evaluated how users executed two different benchmarks, taking inspiration on ISO 9283:2003 norm and from our previous work [26]: *1) Following edges of the box, 2) Following letter S contour*. These are depicted in Fig. 6. The LFAI and VHPE interfaces were also tested on flight with the aerial manipulation robot. Experimental parameters are presented in Table I.

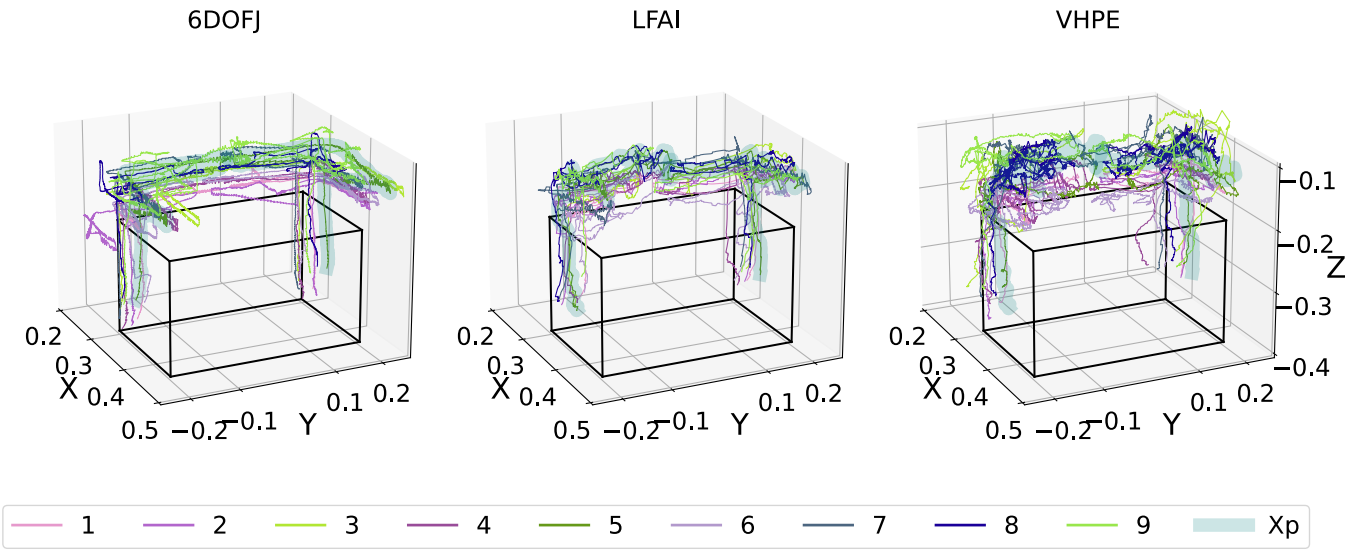

Fig. 5: Comparison of box edge following benchmark with 9 study participants and 3 interfaces: 6DOFJ, LFAI and VHPE.

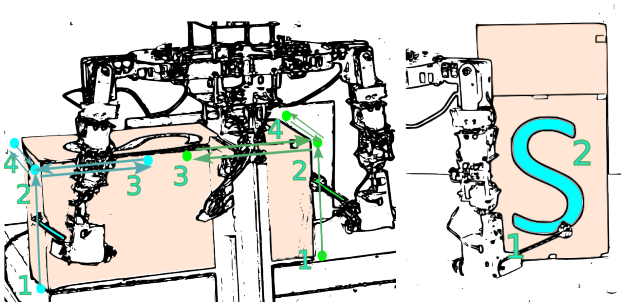

Fig. 6: Controlling both arms to follow edge of the box through points 1-2-3-2-4-2 (*left*), Controlling one arm to follow contour of the letter S from point 1 to point 2 (*right*).

TABLE I: System parameters for the experimental validation

Parameter	Value	Parameter	Value
$\mathbf{R}$	$500\mathbf{I}(3)$	T_s	0.04 s
$\mathbf{Q}_{11}, \mathbf{Q}_{33}, \mathbf{Q}_{55}$	5	z_h^c	2 m
$\mathbf{Q}_{22}, \mathbf{Q}_{44}, \mathbf{Q}_{66}$	0.5	ϕ_h^c, θ_h^c, ρ_h^c	130, 0, 90
L_1^L, L_2^L	0.2 m	L_1^F, L_2^F	0.25 m
D^L	0.25 m	D^F	0.36 m

A. User Study

We evaluted three teleoperation interfaces: 1) 6DOFJ 2) LFAI, 3) VHPE, shown in Fig. 1 with two benchmarks. First task was box edge tracking with both manipulators, and second task was following contour of the letter S with just one manipulator as shown in Fig. 6. Tasks were executed in sequential order. There were 9 participants which were divided into three different groups, and one expert user taken as ground truth. Names of the group corresponds to the order of usage of the teleoperation interfaces. Group 123 firstly used 1) 6DOFJ, then 2) LFAI and lastly, 3) VHPE control. Second group is 321 and third group is 231. Different group ordering was used to average across groups to remove possiblity of influencing results by the order of experimentation.

NASA Task Load Index (TLX) is used to rate operator workload while executing some task. Workload consists of mental demand, physical demand, temporal demand, performance, effort and frustration. Complete NASA TLX consists of two parts. First part is used to determine participant's subjective importance for each category that makes workload. Second part includes rating each workload category for certain task on linear scale from 1-21. NASA raw TLX (RTLX) omits first part due to some concerns that it skews overall measurements. For experimental evaluation of the different control modalities, we employed raw NASA TLX with linear scale range from 1-10 where lower number indicates less workload. Although there is lower resolution overall trend will be the same.

B. Population

Experiments were conducted in collaboration with 9 participants and one expert user whose performance is considered as ground truth. Participants ranged in age 22-39 (mean: 26, std: 5.76) and were male. A limitation of the study arises from small diversity of the study population, which consisted mainly of young male university students. It is not known to what extent age and gender influence the subjective experience of workload. Such discrepancies might reveal interesting differences between age and gender groups rather than invalidate the obtained results.

C. User Study Results

Box following trajectories for each participant and the different teleoperation interfaces are shown in Fig. 5. Fig.

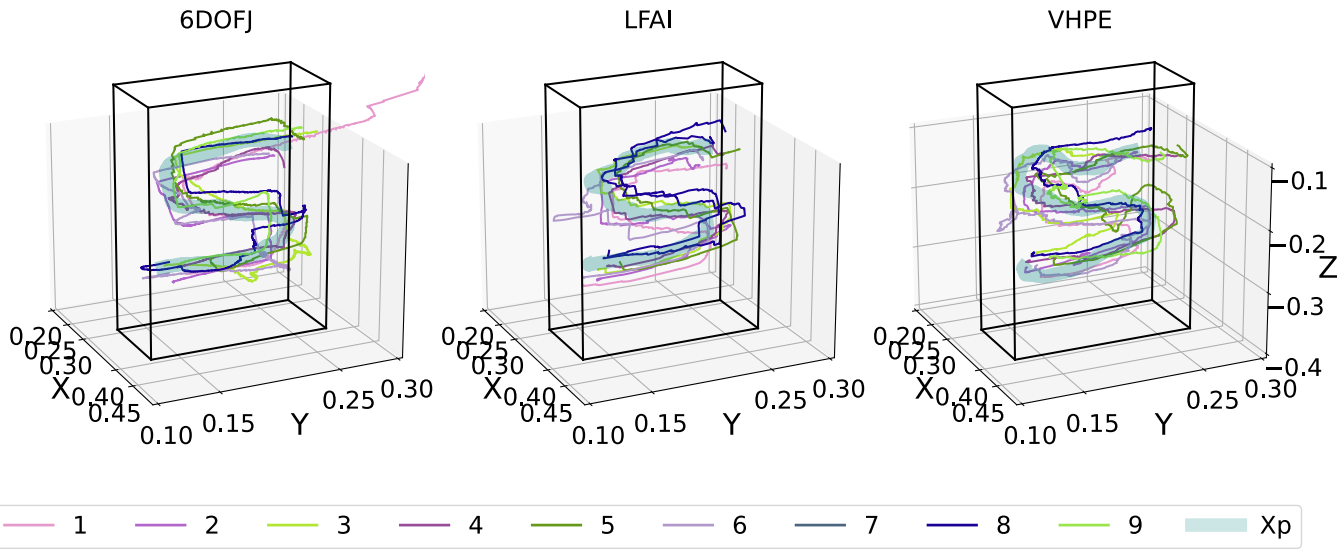

Fig. 7: Comparison of S-letter following benchmark across 9 participants with three interfaces: 6DOFJ, LFAI and VHPE.

.

8 represents the overall NASA RTLX ratings for every participant and control modality (6DOFJ, LFAI and VHPE), for the box following benchmark. Besides that, experiment duration of each participant is shown.

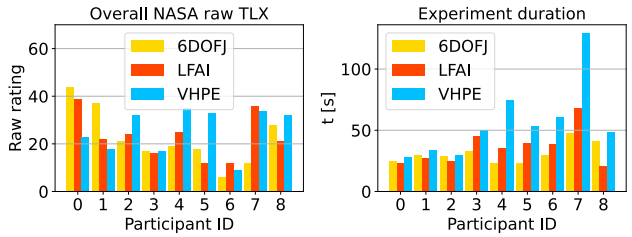

Fig. 8: NASA RTLX rating for each participant (*lower is better*) and duration of the box edges following benchmark (*lower is better)*

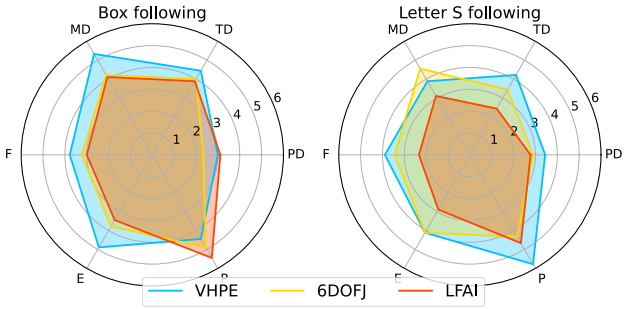

Fig. 9: Comparison of the averaged NASA RTLX across workload categories for different tasks (note: scale ranges from 1-10, for visibility purposes, we plot only to 6, *lower is better*).

Compared to the other teleoperation modalities, VHPE has highest overall RTLX ratings across participants for the edge of the box following task. Average duration of the box following experiment using VHPE is 56.7 seconds, using LFAI 35.8 seconds and with 6DOFJ is 31.2 seconds. It took participants almost twice as long to finish experiment with the VHPE compared to the 6DOFJ. Which is partially in line with NASA RTLX results.

Most of the participants found 6DOFJ the least straining for the box-following task which can be seen in Fig. 9. It is possible to notice how trajectories of the 6DOFJ are most accurate for the box edge following task.

Fig 10 shows NASA RTLX ratings and experiment durations of the each participant for all teleoperation interfaces in the S following task. Average duration of the S following experiment for the VHPE was 24 seconds, for the 6DOFJ 16.8 seconds and for the LFAI 14.8 seconds.

Averaged workload categories shown in Fig. 9 show that LFAI averaged smallest workload on each of the categories compared to the 6DOFJ and the VHPE. It also evidences that using 6DOFJ for the task of following complex curved 3D trajectory is harder in terms of the mental demand, frustration, effort and physical demand compared to the VHPE. Participant trajectories are shown in the Fig. 7. It can be seen that trajectories made with 6DOFJ are slightly different than trajectories made with LFAI or the VHPE. From obtained data, it is possible to conclude that 6DOFJ is by far superior in the edge box following task, taking in consideration experiment duration and overall taskload ratings. However, in the S following task, participants performed better with LFAI and VHPE compared to the 6DOFJ. Averaged NASA

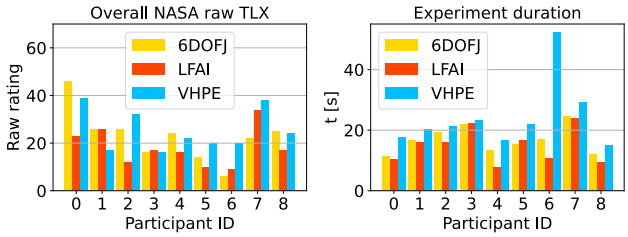

Fig. 10: NASA RTLX rating for each participant (*lower is better*) and duration of the contour S following benchmark (*lower is better)*

RTLX shows that LFAI induces the least workload across participants and tasks.

D. Bimanual Aerial Manipulation of Flexible Object

This experiment is intended to evaluate the leader-follower dual arm interface (LFAI) and the visual human pose estimation (VHPE) in a bimanual manipulation operation with an helical bird flight diverter, a device typically installed on high voltage power lines to prevent the collision of birds. These devices are nowadays installed by skilled human workers in highly risky conditions. Therefore, the goal is to evaluate the suitability of these interfaces on a lightweight and compliant, anthropomorphic dual arm manipulator integrated on a multi-rotor platform. The operation, depicted in Fig. 11, consists of rotating the device from the horizontal to the vertical position to prepare for the insertion of the device on the power line. Three experiments are conducted with the LFAI: 1) motion sequence in testbench without the device grasped, 2) motion sequence in testbench with the device grasped, 3) motion sequence on flight with the device grasped. The evolution of the joints trajectory is depicted in Fig. 12. Note that the weight of the device (0.6 kg) causes a significant deflection on the compliant joints that must be compensated by the user. The performance of the LFAI compared to the VHPE can be appreciated in the video. In this case, the LFAI results are better suited since it allows fast and direct replication of the joint angles generated by the human operator.

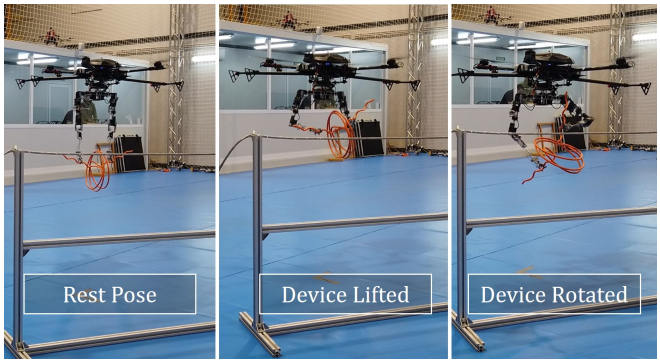

Fig. 11: Anthropomorphic dual arm aerial robot manipulating an helical bird flight diverter in a power line mockup testbed.

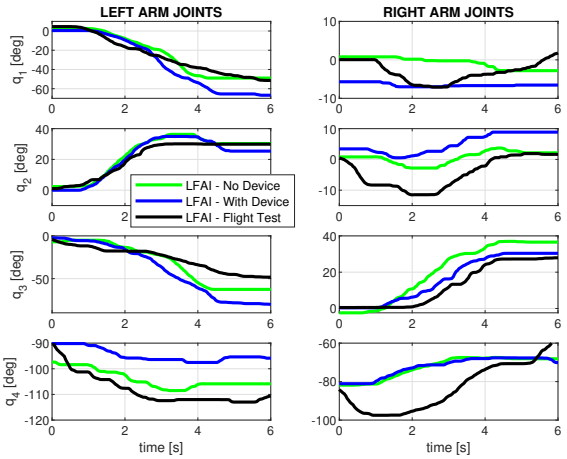

Fig. 12: Evolution of the left and right arms joints during the rotation of the helical bird diverter in three cases: no device grasped, device grasped in testbench, and device grasped on flight (LFAI teleoperation).

VI. CONCLUSION

In this paper we presented three different teleoperation interfaces for the lightweight robotic manipulators. We present 6DOF Joystick (6DOFJ), Leader-follower Arms Interface (LFAI), and Visual Human Pose Estimation interface (VHPE). We extensively tested interfaces with multiple users in the laboratory environment on two different benchmarks. Box following task was designed to evaluate performance of users when following straight paths with both hands, and S contour is used to evaluate following of the curved paths with one hand. For each task, each participant and each control modality, we measured experiment duration, end-effector position and overall workload felt by participant with NASA raw TLX. Based on the obtained data, we can conclude that LFAI is by far easiest to use in terms of induced workload across participants. LFAI and VHPE are evaluated for the bird diverter insertion task, where LFAI proved superior to the VHPE. Informed improvements of the VHPE teleoperation system could result in higher usability and less workload imposed on the operator. Main improvement is related to complete mitigation of the problems caused by the occlusion (wrong depth measurements). Adding one more camera to the system, and using different human pose estimation algorithm could significantly improve the system. Further work will be related to the installation part of the helical bird diverter, forces interaction and improvement of the VHPE system.

ACKNOWLEDGEMENT

This work is supported by the project AERIAL COgnitive Integrated Multi-task Robotic System with Extended Operation Range and Safety (AERIAL-CORE) EU-H2020-ICT (grant agreement No. 871479) and by the European ROBotics and AI Network (euROBIN, Grant agreement ID: 101070596). We want to thank all members of the GRVC

who devoted their time and energy to participate in the experiments.

The research work of Alejandro Suarez is supported by the Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía (Spain) through a post-doctoral research grant. The research work of Filip Zorić is supported by the Croatian Science Foundation under the project "Young Researchers' Career Development Project – Training New Doctoral Students" (DOK-2020-01).

REFERENCES

[1] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[2] A. Ollero, M. Tognon, A. Suarez, D. Lee, and A. Franchi, "Past, present, and future of aerial robotic manipulators," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 626–645, 2021.

[3] E. Papadopoulos, F. Aghili, O. Ma, and R. Lampariello, "Robotic manipulation and capture in space: A survey," *Frontiers in Robotics and AI*, p. 228, 2021.

[4] C.-Y. Weng, Q. Yuan, F. Suarez-Ruiz, and I.-M. Chen, "A telemanipulation-based human–robot collaboration method to teach aerospace masking skills," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3076–3084, 2019.

[5] Y. Yang, Y. Li, C. Fermuller, and Y. Aloimonos, "Robot learning manipulation action plans by" watching" unconstrained videos from the world wide web," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 29, no. 1, 2015.

[6] N. Y.-S. Lii, C. Riecke, D. Leidner, S. Schätzle, P. Schmaus, B. Weber, T. Krueger, M. Stelzer, A. Wedler, and G. Grunwald, "The robot as an avatar or co-worker? an investigation of the different teleoperation modalities through the kontur-2 and meteron supvis justin space telerobotic missions," in *Proceedings of the International Astronautical Congress, IAC*, 2018.

[7] G. A. Yashin, D. Trinitatova, R. T. Agishev, R. Ibrahimov, and D. Tsetserukou, "Aerovr: Virtual reality-based teleoperation with tactile feedback for aerial manipulation," in *2019 19th International Conference on Advanced Robotics (ICAR)*. IEEE, 2019, pp. 767–772.

[8] L. Shao, T. Migimatsu, Q. Zhang, K. Yang, and J. Bohg, "Concept2robot: Learning manipulation concepts from instructions and human demonstrations," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1419–1434, 2021.

[9] A. Grabowski, J. Jankowski, and M. Wodzyński, "Teleoperated mobile robot with two arms: the influence of a human-machine interface, vr training and operator age," *International Journal of Human-Computer Studies*, vol. 156, p. 102707, 2021.

[10] A. Suarez, A. E. Jimenez-Cano, V. M. Vega, G. Heredia, A. Rodriguez-Castaño, and A. Ollero, "Design of a lightweight dual arm system for aerial manipulation," *Mechatronics*, vol. 50, pp. 30–44, 2018.

[11] A. Suarez, G. Heredia, and A. Ollero, "Design of an anthropomorphic, compliant, and lightweight dual arm for aerial manipulation," *IEEE Access*, vol. 6, pp. 29 173–29 189, 2018.

[12] J. Cacace, S. M. Orozco-Soto, A. Suarez, A. Caballero, M. Orsag, S. Bogdan, G. Vasiljevic, E. Ebeid, J. A. A. Rodriguez, and A. Ollero, "Safe local aerial manipulation for the installation of devices on power lines: Aerial-core first year results and designs," *Applied Sciences*, vol. 11, no. 13, p. 6220, 2021.

[13] J. Lee, R. Balachandran, Y. S. Sarkisov, M. De Stefano, A. Coelho, K. Shinde, M. J. Kim, R. Triebel, and K. Kondak, "Visual-inertial telepresence for aerial manipulation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1222–1229.

[14] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006. [Online]. Available: https://doi.org/10.1177/154193120605000909

[15] X. Zhao, Q. Huang, Z. Peng, and K. Li, "Kinematics mapping and similarity evaluation of humanoid motion based on human motion capture," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 1, 2004, pp. 840–845 vol.1.

[16] Y. Ou, J. Hu, Z. Wang, Y. Fu, X. Wu, and X. Li, "A real-time human imitation system using kinect," *International Journal of Social Robotics*, vol. 7, no. 5, pp. 587–600, Nov 2015. [Online]. Available: https://doi.org/10.1007/s12369-015-0296-9

[17] M. Alibeigi, S. Rabiee, and M. N. Ahmadabadi, "Inverse kinematics based human mimicking system using skeletal tracking technology," *Journal of Intelligent & Robotic Systems*, vol. 85, no. 1, pp. 27–45, Jan 2017. [Online]. Available: https://doi.org/10.1007/s10846-016-0384-6

[18] C.-S. Lin, P.-C. Chen, Y.-C. Pan, C.-M. Chang, and K.-L. Huang, "The manipulation of real-time kinect-based robotic arm using double-hand gestures," *Journal of Sensors*, vol. 2020, p. 9270829, Jan 2020. [Online]. Available: https://doi.org/10.1155/2020/9270829

[19] M. Syakir, E. S. Ningrum, and I. Adji Sulistijono, "Teleoperation robot arm using depth sensor," in *2019 International Electronics Symposium (IES)*, 2019, pp. 394–399.

[20] J. B. Martin and F. Moutarde, "Real-time gestural control of robot manipulator through deep learning human-pose inference," in *Computer Vision Systems*, D. Tzovaras, D. Giakoumis, M. Vincze, and A. Argyros, Eds. Cham: Springer International Publishing, 2019, pp. 565–572.

[21] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[22] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-end recovery of human shape and pose," in *Computer Vision and Pattern Regognition (CVPR)*, 2018.

[23] R. C. Luo, B.-H. Shih, and T.-W. Lin, "Real time human motion imitation of anthropomorphic dual arm robot based on cartesian impedance control," in *2013 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, 2013, pp. 25–30.

[24] F. Zorić, G. Vasiljević, M. Orsag, and Z. Kovačić, "Towards intuitive hmi for uav control," pp. 325–332, 2022.

[25] Z. Marinov, S. Vasileva, Q. Wang, C. Seibold, J. Zhang, and R. Stiefelhagen, "Pose2drone: A skeleton-pose-based framework for human-drone interaction," *CoRR*, vol. abs/2105.13204, 2021. [Online]. Available: https://arxiv.org/abs/2105.13204

[26] 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.