Dual-arm robotic manipulation using visual guidance

Pol Ramon-Canyameres
Inst. of Industrial and Control Eng.
Universitat Politècnica de Catalunya
Barcelona, Spain
pol.ramon@upc.edu

Leopold Palomo-Avellaneda
Inst. of Industrial and Control Eng.
Universitat Politècnica de Catalunya
Barcelona, Spain
leopold.palomo@upc.edu

Isiah Zaplana
Inst. of Industrial and Control Eng.
Universitat Politècnica de Catalunya
Barcelona, Spain
isiah.zaplana@upc.edu

Jan Rosell
Inst. of Industrial and Control Eng.
Universitat Politècnica de Catalunya
Barcelona, Spain
jan.rosell@upc.edu

***Abstract*—Dual-arm manipulation poses many challenges in order to actually perform manipulation tasks with enough efficiency. To cope with this challenges, this work-in-progress describes a system for motion coordination of a mobile anthropomorphic dual-arm robot that focuses on the capability for performing manipulations actions with precision by means of a visual guidance provided by a camera mounted on an articulated head. The proposal presents different control modes, evaluated with real experiments, and discusses the main future research lines.**

I. Introduction

Both in industrial and daily life scenarios, many robotic manipulation tasks require the coordination between two robotic arms. For a person with full physical and intellectual capabilities, the coordination of his body to perform several tasks is considered innate, but in the field of robotics it can become quite a big challenge [1], [2]. For instance, simple domestic tasks in a kitchen environment, like putting cereals into a bowl or holding a pot with both hands, involve non-trivial coordination of the two arms. To perform these tasks with a robot, the robot must be able to plan and execute the arm motions in a coordinated way, as well as be able to adapt these motions based on the inputs it perceives, being aware of the environment and reacting to unexpected (and then non pre-programmed) situations. For this, this work-in-progress paper focuses on both the implementation of motion coordination modes for the robotic arms, and the development of a visual guidance framework, with the final goal of manipulating objects in a coordinated manner to bring capabilities that are fundamental for robots to become full coworkers.

II. Problem statement and solution overview

This work copes with the control challenge of motion coordination for dual-arm manipulation tasks not involving bi-manual grasping. Some of the tasks that the proposed developments will allow are, for instance, tasks where one arm holds an open container while the other arm places items inside or removes items from it, or where each arm manipulates or reorients a part to perform an assembly or a packaging task.

The problem of coordinating two arms require, on the one hand, the capability to plan and execute simultaneously the motions of both arms, i.e. a *joint-motion schema*. Here, the availability of a dual-arm motion planner is assumed, able to plan collision-free trajectories for both arms. A control mode will be implemented in order to guarantee the execution of these trajectories.

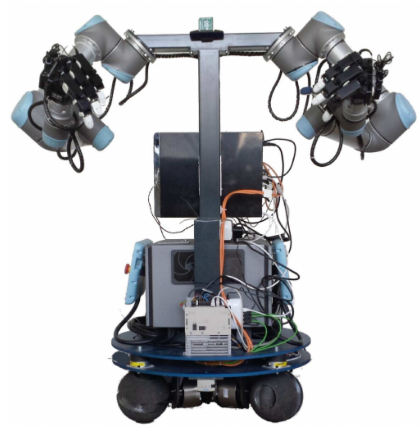

Fig. 1. The MADAR robot [3].

On the other hand, the motion coordination may require a *leader-follower schema*, i.e. one arm performs some pre-computed motions and the other complies to them following a task-dependant predefined reference frame. A control mode will be implemented to allow an arm to act as follower, given a reference frame as input. Also, a perception module with tracking capabilities will be developed to track the frame to be followed.

The paper is structured as follows: Sec. III introduces the MADAR robot used for the validation of the proposals; Sec. IV and Sec. V develop, respectively, the arm control modes and the head perception module; finally, Sec. VI and Sec. VII present the vision guided system and its experimental validation, and Sec. VIII the conclusions and future developments.

III. The MADAR robot

The MADAR robot, shown in Fig. 1, emulates the human morphology to facilitate the robot to work in human environments performing grasping and manipulation tasks [3]. It is composed of several subsystems:

- **Platform**: An omnidirectional mobile base with three traction groups, composed each one of a wheel and a motor, disposed in radial directions with an angle of 120 degrees between them.

- **Arms**: Two UR5 from Universal Robots are used as arms. These arms have 6 DOFs and have collision-detection capabilities that allow a safe human-robot interaction. Both arms are attached to a central metal T-shaped structure, on top of the mobile base.
- **Hands**: Two Allegro Hands V4 (AH) from Wonik Robotics, as the end-effectors of the arms. Each hand has four fingers with four independent torque-controlled joints per finger and tactile sensors WTS-FT 0408 from Weiss at the fingertips.
- **Head**: A pan-and-tilt structure (WidowX XM430) with an OAK-D camera mounted on top.

The mounting orientation of the arms and the inter-arm distance have been chosen to obtain a suitable overlapping of the two workspaces of the robotic arms, as well as a good value of the manipulability measure in front of the robot [3], where it is expected to perform most of the bi-manual manipulation tasks.

All the subsystems are connected to and controlled by an on-board central processing unit using C++ and ROS 2. Besides, the robot can also be remotely controlled using a 5 GHz wireless connection. In this paper, the control modes of the arms and head subsystems are discussed.

IV. The arm control modes

Two methods have been implemented to control the arms for coordinated manipulation tasks.

A. Trajectory control mode

The trajectory control mode has been designed to accept and execute a predefined path from a high-level algorithm to coordinately move the different chosen robot subsystems. The leader arm in a leader-follower coordinated task will be controlled in this mode.

This control mode has been implemented as a ROS 2 action server. The key to using ROS 2 actions is that the client can cancel an ongoing trajectory if required, or it can be aborted by the robot itself if an unexpected problem occurs. The action server can be flexibly configured, according to the needs of the task, to react differently to new incoming trajectories during the execution of an action:

- **REJECT**: The incoming action is not accepted and it is rejected, because there is a current action being executed.
- **ACCEPT**: The incoming action is accepted, but it will wait until the current action finishes.
- **EXECUTE**: The incoming action is accepted and executed, as the current one is immediately stopped.

The MADAR robot can use this trajectory controller mode to act on different parts (arms and hands) according to different configurations modes set by the following set of ROS 2 actions exposed:

- /madar/arm_right/move
- /madar/arm_left/move
- /madar/hand_right/move
- /madar/hand_left/move

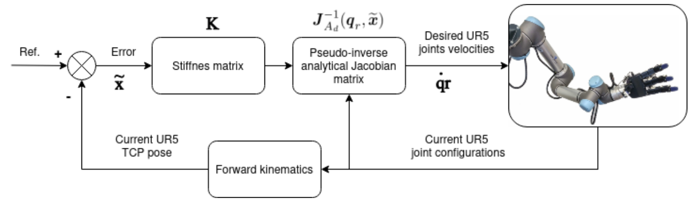

Fig. 2. Block diagram of the Resolved-Velocity Control

- /madar/manipulator_right/move
- /madar/manipulator_left/move
- /madar/dual_arm/move
- /madar/dual_manipulator/move

When several subsystems are involved, e.g. the two arms in the */madar/dual_arm/move* action, the synchronized execution is achieved by using the Time Trajectory Generator algorithm [4] that synchronizes in time the joint-space paths of the different subsystems. This algorithm sets velocities to each joint considering the maximum joint velocities and the required displacement of each joint at each way-point in the path.

B. Follower control mode

The main idea of this control mode is to work online to follow and maintain a desired target alignment, and allow for online dynamic adjustments based on the incoming data from the desired target.

Any reference frame attached to the hand-arm system that works under this control mode can be used as the reference frame to be controlled (i.e., to which a specified set of constraints can be defined). Also, any reference frame (e.g., attached to the other hand-arm system or to the object it carries) can be defined as the reference frame with respect to which the constraints are defined. This frame can also be detected and tracked by a perception system.

This controller uses a *Resolved-Velocity Control* in the operational space (see Fig. 2 for a block diagram representation of the system) [5]. It is based on the computation of joint velocities from the operational space error. In particular, the error between the controlled reference frame and the desired frame is computed and evaluated so that the joint velocities that minimize such an error are calculated and sent to the arm.

The order 6 *stiffness matrix* $\boldsymbol{K}$ influences the transient behaviour, which can be determined with physical sense if the error is defined as shown below, and $\boldsymbol{K}$ is set diagonal, i.e. each term of the diagonal determines how the transient is in each direction (in translation and rotation) of the desired reference frame, e.g. if we are mainly interested in maintaining the x-y plane parallel to the floor, the values of $\boldsymbol{K}$ corresponding to the rotations about x and y will be set with higher values, whereas the others will be set to low or zero values.

The operational space error vector is defined as:

$$\widetilde{\boldsymbol{x}} = -\begin{pmatrix} \boldsymbol{o}_{d,e}^{d} \\ \boldsymbol{\phi}_{d,e} \end{pmatrix} \tag{1}$$

where $\boldsymbol{o}^d_{d,e}$ is the position error between the desired and controlled frames with respect to the desired frame, and $\boldsymbol{\phi}_{d,e}$ is its orientation error represented using the XYZ-Euler angles. The Euler angles have been selected to be in the range $[-\pi, \pi]$ to avoid instabilities when the orientation error is close to zero.

The analytic Jacobian corresponding to the definition of the error (1) in the operational space is [5]:

$$J_{A_d}(\boldsymbol{q}, \widetilde{\boldsymbol{x}}) = T_A^{-1}(\boldsymbol{\phi}_{d,e}) \begin{bmatrix} R_d^T & 0_3 \\ 0_3 & R_d^T \end{bmatrix} J(\mathbf{q}) \tag{2}$$

where $T_A^{-1}(\boldsymbol{\phi}_{d,e})$ is the matrix relating the geometric and analytic Jacobians, R_d is the rotation matrix relating the orientation of the desired pose with respect to the base of the robot, and $J(\mathbf{q})$ is the geometric Jacobian of the arm that, in this case, has a closed formula. Using the XYZ-Euler angles (α, β, γ), $T_A(\phi_d)$ is expressed as follows:

$$T_A(\phi_{d,e}) = \begin{bmatrix} I_3 & 0_3 \\ 0_3 & R \end{bmatrix} \tag{3}$$

where

$$R = \begin{bmatrix} 1 & 0 & \sin(\beta) \\ 0 & \cos(\alpha) & -\sin(\alpha)\cos(\beta) \\ 0 & \sin(\alpha) & \cos(\alpha)\cos(\beta) \end{bmatrix} \tag{4}$$

Finally, the commanded velocities are:

$$\dot{\boldsymbol{q}}_{\boldsymbol{r}} = J_{A_d}^{-1}(\boldsymbol{q}, \widetilde{\boldsymbol{x}}) \boldsymbol{K} \widetilde{\boldsymbol{x}} \tag{5}$$

Regarding the implementation details, these velocities have been saturated at the maximum joint velocities, and, for computational efficiency purposes, the geometric Jacobian and the forward kinematics matrices of the UR5 (MADAR's arms) have been hard-coded into the controller algorithm, where only the error and the joint state are evaluated.

V. The Head Perception Module

MADAR's robotic head module is composed of a camera mounted on a turret structure. Since the vision field of the camera does not span the whole working area of MADAR's robot, the camera's mounting structure must allow it to be moved to reach the full range [6]. Therefore, the head perception module allows to capture the complete environment where the MADAR robot can perform the task.

Moreover, the perception module is equipped with a tracking system capable of following a given reference frame. With the aim to visualize the target frame at the center of the image, the error $(\Delta x, \Delta y, \Delta z)$ between the current position of the optical frame and the target position is first computed and, then, the pan and tilt angles that minimize this error in the y and z components are set as follows:

$$\begin{cases} \Theta^{\text{Pan}}_{\text{Target}} = \Theta^{\text{Pan}}_{\text{Current}} + \min\limits_{\Delta y} \left(\operatorname{atan2}\left(\Delta y, d\right)\right) \\ \Theta^{\text{Tilt}}_{\text{Target}} = \Theta^{\text{Tilt}}_{\text{Current}} - \min\limits_{\Delta z} \left(\operatorname{atan2}\left(\Delta z, d\right)\right) \end{cases} \tag{6}$$

where $d = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ and the minimum is computed analytically.

Unlike the other subsystems in MADAR, the head works independently.

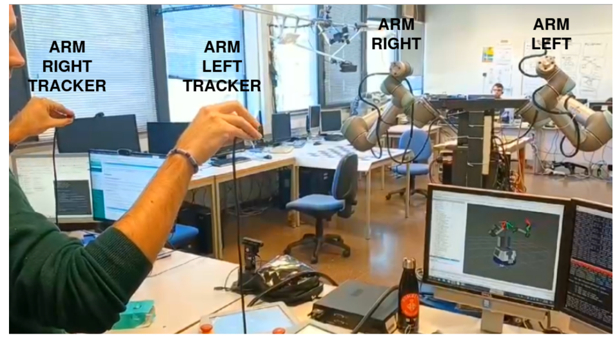

Fig. 3. Follower control mode demo: Each arm follows the reference frame of a tracker manipulated by the user.

VI. Vision Guided Manipulation

Vision guided robotic manipulation is achieved by coordinating the control modes of the arms with the sensing capabilities of the head perception module. Two different working modes have been designed:

- **Joint-motion schema**:The perception module is used to capture the environment and determine which are the desired goal poses for the arms, then the dual-arm motion planner plans the motions that are finally executed in a coordinated way using the *trajectory control mode* introduced in Sec. IV-A.
- **Leader-follower schema**: The perception module is responsible to track a reference frame attached to the leader arm, e.g. the reference frame of the object it may be helding, and this frame is sent to the follower arm, that, using the *follower control mode* introduced in Sec. IV-B, maintains a given relative pose.

A high-level manipulation framework [7], will be the responsible of controlling the system by, for instance, determining the type of schema to use, the frame to be followed or the relative pose to be maintained.

VII. Exeprimental Validation

Independent experiments were designed and conducted to validate the motion coordination with visual guidance, and finally, a global experiment was carried out to test the integration. The videos of all the experiments are available at https://sir.upc.edu/projects/madar/videos/etfa24:

- **Trajectory mode experiment**: This experiment consists in executing a predefined joint-space path provided by the planner for both arms as a time-coordinated trajectory. The resulting motions are executed in a coordinated way, being the total duration of the trajectory a parameter that can be configured.
- **Follower mode experiment**: This experiment consists in moving both arms with the *following control mode* by following the reference frame of a tracker device being moved by a user. The resulting arm motions actually follow the user motions and the desired target alignment is maintained (Fig. 3).

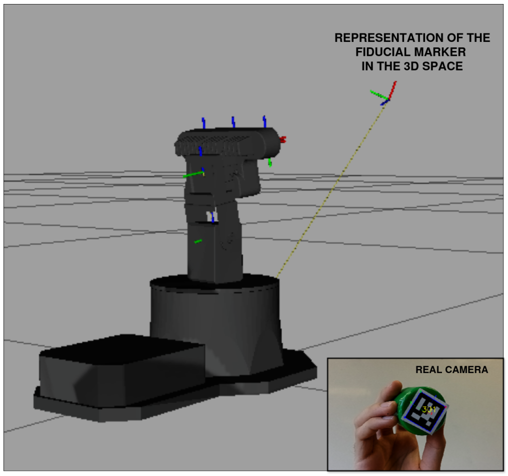

Fig. 4. Head-tracking demo: The pan and tilt angles keep changing so as to maintain the reference frame of the fiducial marker moved by the user at the center of the image.

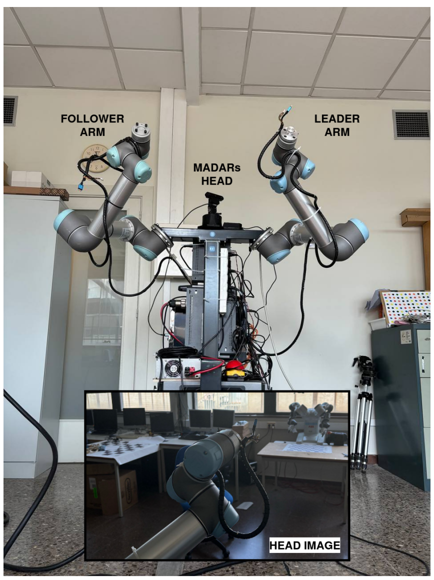

Fig. 5. Visual guidance demo: The user manipulates the leader arm and the visual guidance system tracks its end-effector (maintaining it at the center of the image) sending its frame to the other arm, that follows it with a given predefined transformation.

- **Head tracking mode experiment**: This experiment consists in tracking a frame, i.e. moving the pan and tilt joints so as to have the target frame always at the center of the image. The experiment has bee conducted using the reference frame of a fiducial marker (Fig. 4).
- **Visual guidance experiment**: This experiment consists in moving one arm and making the other arm to follow the first one with a given predefined transformation between the end-effectors, as well as making the head to follow the leader arm end-effector (Fig. 5).

VIII. Conclusions and Future Work

A proposal for the motion coordination of a mobile anthropomorphic dual-arm robot using visual guidance has been presented. The proposal has been implemented on the MADAR robot, by integrating advanced control modes and a head tracking module to enable precise and adaptive manipulation tasks. Through some preliminary experiments, it has been checked that the trajectory and follower control modes can execute coordinated movements aided by visual feedback. Therefore, the bases have been set for the MADAR to be able to execute a high-level coordination tasks in the future.

The validation results confirmed that the trajectory control mode is effective for predefined path execution, while the follower control mode excels in real-time adjustments and maintaining target alignment. The visual guidance system, combining head tracking with arm control, proved essential for continuous monitoring and precise task execution.

Future work will focus on enhancing some capabilities:

- Integration with the high-level reasoning module to define the working modes and the constraints to be set for the task [7].
- Inclusion of a smart perception module to detect the pose of the objects being manipulated and their main features [8].
- Inclusion of collision-check capabilities to avoid the risk of collision in the follower arm [9].

Acknowledgements

This work was supported by the European Commission's Horizon Europe Framework Programme with the project IntelliMan (AI-Powered Manipulation System for Advanced Robotic Service, Manufacturing and Prosthetics) under Grant Agreement 101070136.

References

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

[2] H. A. Park and C. G. Lee, "Dual-arm coordinated-motion task specification and performance evaluation," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 929–936.

[3] R. Suárez, L. Palomo-Avellaneda, J. Martínez, D. Clos, and N. García, "Manipulador móvil, bibrazo y diestro con nuevas ruedas omnidireccionales," *Revista Iberoamericana de Automática e Informática industrial*, vol. 17, no. 1, pp. 10–21, 2020.

[4] T. Kunz and M. Stilman, "Time-Optimal Trajectory Generation for Path Following with Bounded Acceleration and Velocity," in *Robotics: Science and Systems VIII*. The MIT Press, 07 2013.

[5] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[6] M. Aguilar, D. Ronquillo, J. Rosell, L. Palomo-Avellaneda, and R. Suárez, "A simple yet smart head module for mobile manipulators," in *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2023, pp. 1–4.

[7] O. Ruiz-Celada, A. Dalmases, R. Suárez, and J. Rosell, "BE-AWARE: an ontology-based adaptive robotic manipulation framework," in *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2023, pp. 1–4.

[8] O. Ruiz-Celada, A. Dalmases, I. Zaplana, and J. Rosell, "Smart Perception for Situation Awareness in Robotic Manipulation Tasks," *IEEE Access*, vol. 12, pp. 53 974–53 985, 2024.

[9] J. Rosell, A. Pérez, A. Aliakbar, Muhayyuddin, L. Palomo, and N. García, "The Kautham project: A teaching and research tool for robot motion planning," in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, 2014, pp. 1–8.