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

Assembling Solar Panels by Dual Robot Arms Towards Full Autonomous Lunar Base Construction

Luca Nunziante*[1,2], Kentaro Uno[2], Gustavo H. Diaz[2],
Shreya Santra[2], Alessandro De Luca[1] and Kazuya Yoshida[2]

***Abstract*— Since the successful Apollo program, humanity is once again aiming to return to the Moon for scientific discovery, resource mining, and inhabitation. Upcoming decades focus on building a lunar outpost, with robotic systems playing a crucial role to safely and efficiently establish essential infrastructure such as solar power generating towers. Similar to the construction of the International Space Station (ISS), shipping necessary components via modules and assembling them in situ should be a practical scenario. In this context, this paper focuses on the integration of vision, control, and hardware systems within an autonomous sequence for a dual-arm robot system. We explore a perception and control pipeline specifically designed for assembling solar panel modules, one of the benchmark tasks. Ad hoc hardware was designed and tested in real-world experiments. A mock-up of modular solar panels and active-passive connectors are employed, with the control of this grappling fixture integrated into the proposed pipeline. The successful implementation of our method demonstrates that the two robot manipulators can effectively connect arbitrarily placed panels, highlighting the seamless integration of vision, control, and hardware systems in complex space applications.**

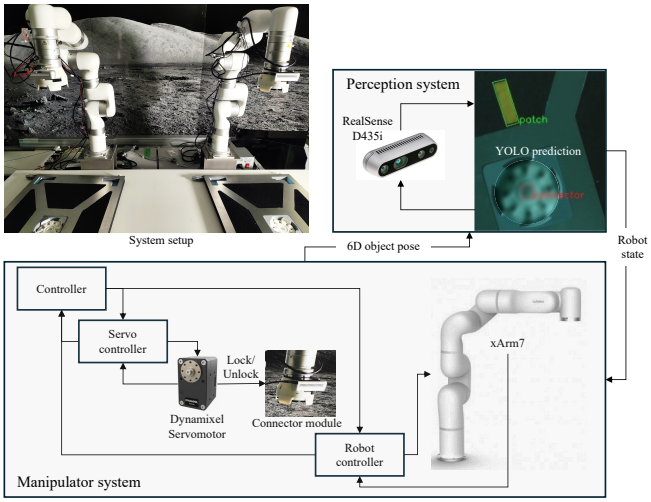


Fig. 1: Overall hardware setup (top) and functional scheme to illustrate the integration and interaction of the perception and control modules (bottom).

I. Introduction

In the past years, the interest in lunar exploration has been growing worldwide [1]. Up-to-date successful missions include the Indian Space Research Organization (ISRO)'s Chandrayaan-3, Japan Aerospace Exploration Agency (JAXA)'s extremely precise soft landing on the moon with the SLIM (smart lander for investigating moon) mission, and a USA based startup, Intuitive Machines's IM-1 lunar landing [2]. From navigating the challenging lunar terrain to conducting complex manipulation tasks, autonomous robotic systems offer a unique set of capabilities that have the potential not only to boost scientific progress but also to mitigate risks associated with human missions. To ensure self-sustenance and to support human presence on the Moon, robots capable of autonomous manipulation tasks are particularly relevant. During the early stages of lunar base development, these robots must create vital infrastructure such as solar power units and communication stations with Earth. A practical approach is to ship all required components in modules and have the robots assemble them on-site. In this scenario, the robots should also be modular.

In this work we present a perception and control pipeline to perform autonomous assembly tasks in a multi-robot setup; two robot arms have the task to localize two solar panels, pick them up with dedicated connectors, safely lift them avoiding collision with the table they are placed on and ultimately assemble them. The perception module is a YOLOv8.1 [3] (You Only Look Once) model, an extension of YOLOv8 trained to predict Oriented Bounding Boxes. The control module integrates data coming from the vision system with classical control methods like impedance control, Nonlinear Model Predictive Control (NMPC) and force control [4], showcasing a comprehensive system integration. The major contributions of this paper are highlighted as follows:

- Fully autonomous pipeline of the robotic assembly of the solar panels, which is regarded as the essential milestone task, is designed and implemented.
- Hardware and software systems are optimally integrated, resulting in successful real-world task demonstration.

The main goal of this work is to present a fully autonomous pipeline that integrates a state-of-the-art perception module with the control of novel hardware in a multi-robot cooperative scenario.

The remaining part of this paper is structured as follows. In Section II we review the state-of-the-art perception and

[1]L. Nunziante and A. De Luca are with the Department of Computer, Control and Management Engineering, Sapienza University of Rome, Italy. This work was performed while L. Nunziante was visiting the Tohoku University. Email: nunziante.2015361@studenti.uniroma1.it, deluca@diag.uniroma1.it

[2]L. Nunziante, K. Uno, G.H. Diaz, S. Santra, and K. Yoshida are with the Space Robotics Lab. (SRL) in Department of Aerospace Engineering, Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan.

*Corresponding author is Luca Nunziante.

control schemes as standalone components. In Section III we design the aggregation of the modules and the overall pipeline for the panels assembling. Section IV details the experimental setup to conduct the robotic demonstration considering the actual task scenario. Section V finally summarizes the results and discusses future work.

II. Related Work

In this work, object detection, impedance control, force control, and Nonlinear Model Predictive Control (NMPC) are exploited to design a pipeline to achieve full autonomous assembly of structure modules. The relevant background on these techniques is recalled hereafter.

A. Perception

Object detection is a computer vision task aimed at localizing and identifying objects within an image, assigning a label to each identified object. Thanks to the seminal work [5] that introduced Deep Convolutional Neural Networks (DCNNs) into the computer vision field, and to the technological advancements of the last decades, DCNNs undoubtably became the standard for computer vision tasks at large.

In the evolution of DCNNs for object detection, two branches can be identified: two-stage and one-stage detectors [6]. The models of the former category separate the object location task and the object classification task, while models of the latter generate the class probabilities and location coordinates of an object in a single stage. Separating the two tasks yields accurate models with slower inference speed [7] which may not be suitable for real-time applications, while one-shot models trade off higher speed for lower accuracy.

Through the years, the one-shot YOLO series stood out as a standard for real-time object detection thanks to the high speed and accuracy achieved. Upon its release, YOLO was the first one-stage model for object detection, boasting impressive inference speed but suffering in terms of localization accuracy [8]. Over the subsequent versions much improvement was made, until the latest release which is YOLOv9 [9]. YOLO models are extremely versatile and have been applied to a wide array of different tasks, including but not limited to construction: building classification [10], healthcare: fracture detection [11] and agriculture: fruit ripeness identification [12].

B. Control

Impedance control is a control method that makes the robotic system react in a desired way to external force, imposing a desired dynamic behavior to the interaction between the robot's end-effector and the environment. The impedance model imposed is

$$\boldsymbol{M}_{\mathrm{m}}(\ddot{\boldsymbol{r}} - \ddot{\boldsymbol{r}}_{\mathrm{d}}) + \boldsymbol{D}_{\mathrm{m}}(\dot{\boldsymbol{r}} - \dot{\boldsymbol{r}}_{\mathrm{d}}) + \boldsymbol{K}_{\mathrm{m}}(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{d}}) = \boldsymbol{f}_{\mathrm{e}} \quad (1)$$

where $\boldsymbol{r} \in \mathbb{R}^r$ is a representation for the relevant components of the end-effector pose (position and orientation), r is the dimension of the robot task — e.g., three for planar tasks, six for spatial — and $\boldsymbol{r}_{\mathrm{d}}$ is a desired pose. $\boldsymbol{M}_{\mathrm{m}} > 0, \boldsymbol{D}_{\mathrm{m}} \geq 0, \boldsymbol{K}_{\mathrm{m}} > 0 \in \mathbb{R}^{r \times r}$ are respectively the desired mass, damping and stiffness imposed by control, and $\boldsymbol{f}_{\mathrm{e}} \in \mathbb{R}^r$ are the generalized external forces applied on the end-effector. With this scheme, it is possible to indirectly control the contact forces, e.g., to prevent damaging impacts in case of environment uncertainties or to assign a soft environment interaction. Thanks to such versatility, this control scheme is widely used not only in mechanical manipulation but also in human-machine interaction and motion of robotic devices with adjustable compliance [13].

Unlike impedance control, force control schemes are aimed at precisely regulating a desired contact force at the end-effector level where contact is expected. Thus, differently from impedance control where a desired end-effector pose $\boldsymbol{r}_{\mathrm{d}}$ is used, in this scheme a desired force reference is required. To reduce the dynamics of the manipulator to a free-floating mass, a Cartesian feedback linearization law can be employed first [14]. It is then possible to use linear, lumped parameter models of different order to analyze the interaction between the robot arm, the force sensor at its end-effector, and the environment [15]. In this way, simple linear controllers (e.g., PID or one of its subsets) can be implemented to regulate efficiently the contact force.

Model Predictive Control (MPC) is a control strategy that operates by repeatedly solving a constrained Optimal Control Problem (OCP) over a finite time horizon, using a model of the plant to predict future behavior. At each control sampling instant an optimal control sequence that maximizes a performance measure, which usually takes the form of a cost to be minimized, is found and the first element of the control sequence is issued. At subsequent control instants, the process is repeated using the latest available information on the system state. This introduces feedback into the control scheme and robustness to uncertainties or unknown variations in the model [16]. In general, one distinguishes between linear and nonlinear model predictive control (NMPC). In the former case, linear models are used for the system dynamics, states and inputs are subject to linear constraints, and a quadratic cost function is used; instead, NMPC refers to MPC schemes that are based on nonlinear models, and/or consider non-quadratic cost functionals and/or impose general nonlinear constraints on the states and inputs [17]. Due to its robustness and ability to handle constraints and general nonlinear systems, MPC has been widely used in process control in industries both in its linear [18] and nonlinear formulation [19].

III. Method

The proposed method integrates the previously outlined perception and control methodologies into an assembly pipeline. Fig. 2 illustrates a flowchart delineating the various phases of the assembly process. In this section, we elaborate on each element within this scheme, whose structure is reminiscent of the Finite State Machine developed in [20].

A. Approaching and Grasping

The perception module comprises a YOLOv8.1 model and a deprojection operation. As the YOLO model performs

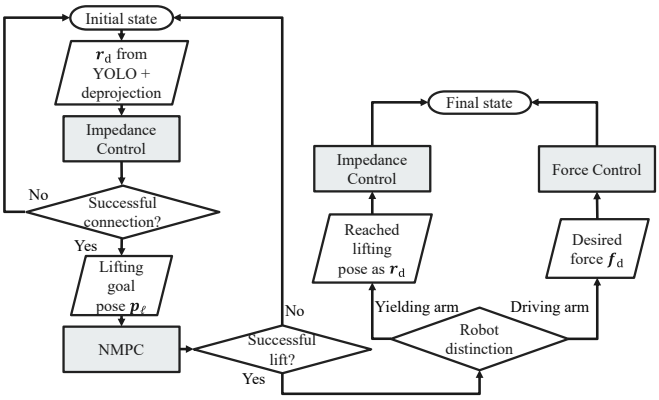


Fig. 2: Flowchart of the assembly pipeline. Impedance control is utilized to adjust to the pushed force when grasping the panels as well as inserting one panel into the other. NMPC allows controlling the panel picking up trajectory avoiding the collision with the surrounding environment.

Fig. 3: Result of the inference made by the YOLO model in challenging lighting conditions: dark assumed in Lunar environment. From the patch detection we retrieve the panel orientation, while the connector detection provides the desired position for the end-effector.

inference only once in the initial state, high inference speed is not a requirement and the largest model with 69.5M parameters can be used. The network is pretrained on the DOTA-v1.0 [21] dataset and fine-tuned on a custom dataset made of 309 training images and two classes. Inference results in challenging lighting conditions are shown in Fig. 3.

From the detection of the patch we retrieve the orientation of the panel on the table, and from the pixel coordinates of the connector, using the depth information from the stereo camera mounted on the end-effector, we deproject the point first from the pixel space to the 3D camera frame, and finally to the robot base frame using the robot state. This procedure is analogous to the one in [22]. The deprojected point together with the panel orientation retrieved from the patch detection constitute the desired pose $\boldsymbol{r}_\mathrm{d} \in \mathbb{R}^6$ that is used in (1), where we can further specify the approach velocity and acceleration $\dot{\boldsymbol{r}}_\mathrm{d}, \ddot{\boldsymbol{r}}_\mathrm{d}$. This impedance-based approach will allow the correct grasp of the panel even in the presence of small detection errors that may arise, especially in challenging lighting conditions typical of lunar scenarios (see Fig. 3).

Once the active side of the grappling fixture is properly in place (see Fig. 4), we command the servo motor to lock the mechanism, thereby attaching it to the passive side and securing the grasp.

B. Collision-free lifting

After the successful connection, the robot enters in velocity control mode to bring the end-effector to a desired pose where the panels are lifted. Since the payload is now attached to the end-effector, it is crucial to avoid collisions between the payload, i.e., a panel and obstacles in the environment, such as the table on which it is placed. The choice of NMPC, differently from other control choices, allows to explicitly consider nonlinear collision avoidance constraints, offering also the additional flexibility of dynamically changing constraints in non-static environments.

In this framework, we consider the end-effector position to be directly controllable via velocity commands, while for the orientation we adopt an ingenious approach. Given the current orientation represented by the rotation matrix $\boldsymbol{R}_\mathrm{A} \in SO(3)$ and the desired one represented by $\boldsymbol{R}_\mathrm{B} \in SO(3)$, where $SO(3)$ is the special orthogonal group of order 3, we determine a unit axis $\boldsymbol{a} \in \mathbb{R}^3$ and an angle $\theta_\mathrm{AB} \in \mathbb{R}$ such that

$$\boldsymbol{R}(\boldsymbol{a}, \theta_\mathrm{AB}) = \boldsymbol{R}_\mathrm{A}^\top \boldsymbol{R}_\mathrm{B} \ . \tag{2}$$

Using the axis-angle representation with $(\boldsymbol{a}, \theta_\mathrm{AB})$ has two main advantages:

- Constraining the change of orientation from $\boldsymbol{R}_\mathrm{A}$ to $\boldsymbol{R}_\mathrm{B}$ to occur as a rotation around a predefined axis offers more predictable intermediate orientations than considering the elements of an Euler representation independently. This is not a secondary aspect when dealing with large payloads, as in our application.
- Within the NMPC scheme, a single scalar variable θ is used to represent the orientation. The value of θ will be 0 at the initial instant and has to reach to the desired value of θ_AB.

Thus, the desired end-effector pose $\boldsymbol{r}_\mathrm{d} \in \mathbb{R}^6$ is transformed to a lifting goal $\boldsymbol{p}_\ell = (x_d, y_d, z_d, \theta_\mathrm{AB}) \in \mathbb{R}^4$ where (x_d, y_d, z_d) is the desired end-effector position and θ_AB comes from (2).

Initially, we assume to be able to control the value of θ directly via a velocity input u_θ. Therefore, the state of the system used in the NMPC scheme is $\boldsymbol{x} = \begin{pmatrix} x & y & z & \theta \end{pmatrix} \in \mathbb{R}^4$, the state space representation is

$$\dot{x} = u_x, \quad \dot{y} = u_y, \quad \dot{z} = u_z, \quad \dot{\theta} = u_\theta \tag{3}$$

and we formulate the OCP to solve at each control instant as

$$\begin{aligned} * \min_{\boldsymbol{x}(\cdot), \boldsymbol{u}(\cdot)} & \int_0^T \|\boldsymbol{x}(\tau) - \boldsymbol{p}_\ell\|^2_{\boldsymbol{Q}} + \|\boldsymbol{u}(\tau)\|^2_{\boldsymbol{R}} \mathrm{d}\tau \\ & + \|\boldsymbol{x}(T) - \boldsymbol{p}_\ell\|^2_{\boldsymbol{W}^e} \end{aligned} \tag{5}$$

subject to:

$$\boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{6}$$

$$\boldsymbol{f}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{u}(t)) = \boldsymbol{0} \qquad t \in [0, T) \tag{7}$$

$$\underline{\boldsymbol{h}} \leq \boldsymbol{h}(\boldsymbol{x}(t), \boldsymbol{s}) \leq \overline{\boldsymbol{h}} \qquad t \in [0, T] \tag{8}$$

$$\underline{\boldsymbol{u}} \leq \boldsymbol{u}(t) \leq \overline{\boldsymbol{u}} \qquad t \in [0, T) \tag{9}$$

The cost function has a quadratic cost to go and terminal cost to bring the state to the desired value, while also

minimizing the control effort. Moreover, $T \in \mathbb{R}^+$ is the prediction horizon, $\boldsymbol{Q} \geq 0, \boldsymbol{R} > 0, \boldsymbol{W}^e \geq 0$ are weighting matrices of appropriate size, $\boldsymbol{f} : \mathbb{R}^4 \times \mathbb{R}^4 \times \mathbb{R}^4 \mapsto \mathbb{R}^4$ is the implicit linear system dynamics from (3), $\boldsymbol{x}_0$ is the initial system state, $\boldsymbol{s} \in \mathbb{R}^{n_\mathrm{s}}$ is a vector of system parameters, $\boldsymbol{h} : \mathbb{R}^4 \times \mathbb{R}^{n_\mathrm{s}} \mapsto \mathbb{R}^{n_\mathrm{c}}$ is the nonlinear constraint function imposing a number of n_c constraints for collision avoidance, $\underline{\boldsymbol{h}}, \overline{\boldsymbol{h}} \in \mathbb{R}^{n_\mathrm{c}}$ are lower and upper bounds for the vector function $\boldsymbol{h}$, while $\underline{\boldsymbol{u}}, \overline{\boldsymbol{u}} \in \mathbb{R}^4$ are input bounds. The only source of nonlinearity in this NMPC scheme comes from the collision avoidance constraints imposed in (8).

Functional for collision avoidance is the set of parameters $\boldsymbol{s}$ that includes a representation of the initial orientation of the end-effector $\boldsymbol{R}_\mathrm{A}$, the neutral axis $\boldsymbol{a}$ from (2), and information about the payload geometry. To ensure a collision-free motion, we use the four corners of the payload as control points. Using the information stored in $\boldsymbol{s}$, together with the current state of the system, it is possible to retrieve their coordinates in the robot base frame (see Fig. 4). Observing that at any time instant t the actual end-effector (EE) orientation is given by

$$\boldsymbol{R}_\mathrm{EE}(t) = \boldsymbol{R}_\mathrm{A} \boldsymbol{R}\left(\boldsymbol{a}, \theta(t)\right) \tag{10}$$

and using knowledge about the payload geometry, it is finally possible to expand the constraints (8) as

$$v_{x,i} \geq b_\mathrm{collision}, \quad v_{y,i} \geq y_\mathrm{wall}, \quad v_{z,i} \geq z_\mathrm{min}, \tag{11}$$

for $i = 1, \ldots, 4$, where $\boldsymbol{v}_i = \begin{pmatrix} v_{x,i} & v_{y,i} & v_{z,i} \end{pmatrix}$ is the i-th control point expressed in the robot base frame. The numbering convention begins with $\boldsymbol{v}_1$ representing the top right corner of the panel from a top view (see bottom right image of Fig. 4), and subsequent numbering progresses counterclockwise. Moreover, $b_\mathrm{collision} > 0$ is a constant to avoid collisions of the panel with the base of the robot, y_wall erects a virtual wall separating the two robot arms to prevent collision between the payloads during the lift, and z_min is the minimum height for not colliding with the table. Differently from the previous two, this last value is not predetermined, but obtained by recording the end-effector z-coordinate when the panel is grasped. Finally, note that the nonlinearity of the constraints (11) with respect to the NMPC state $\boldsymbol{x}$ lies in the transformation needed to obtain $\boldsymbol{v}_i$ in the robot arm base frame Σ_B.

C. Assembly

After the robot end-effector reaches the desired pose avoiding collisions, the two arms need to collaborate to successfully assemble the solar panels. To achieve this they are controlled differently: one arm (Yielding arm in Fig. 2) enters again into impedance control mode having a desired pose $\boldsymbol{r}_\mathrm{d}$ that is the one reached at the end of the lifting phase, while the other arm (Driving arm in Fig. 2) enters force control mode applying a desired wrench (force and moment) $\boldsymbol{f}_\mathrm{d} \in \mathbb{R}^6$ at the end-effector level. Thus, the first arm will be stationary and compliant to external forces according to (1), with these external forces being explicitly applied by the Driving arm. This force interaction leads to a successful assembly, compensating for small uncertainties and pose errors.

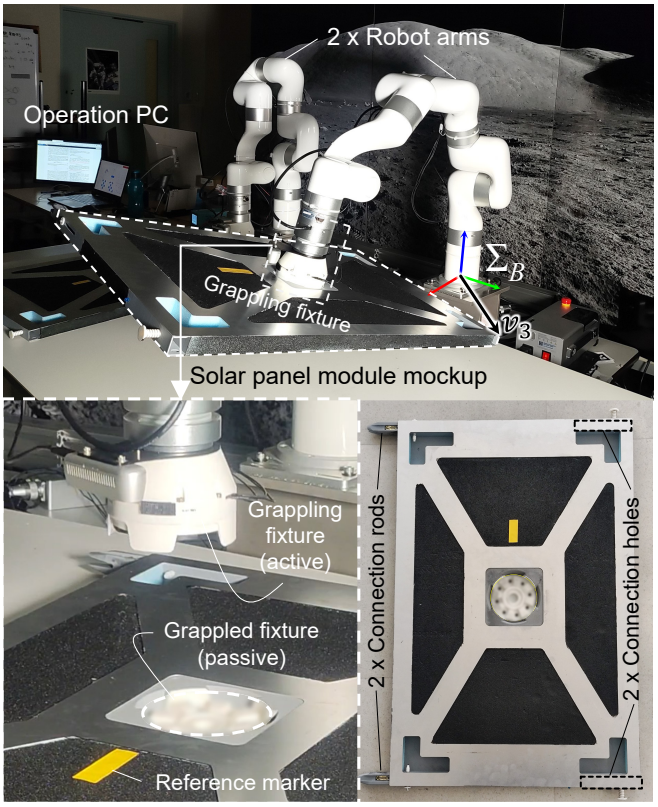


Fig. 4: Experimental setup for solar panels assembly demonstration (top), showing the representation of the robot base reference frame and of the control point $\boldsymbol{v}_3$: left bottom corner of the panel. The end-effector of the robot has the dedicated grappling fixture to grasp the adapter located at the center of the backside of the panel (bottom left). Each panel has two connection rods at two corners which are inserted into the holes prepared on the other edge for assembly (bottom right).

IV. Real World Demonstration

A. Hardware Setup

The setup consists of the following elements (also see Fig. 4). Two solar panels and two fixed 7-DoF robotic arm manipulators, xArm7 from UFactory, whose end-effector is equipped with an Intel Realsense D435i RGBD camera, a 6-axis force/torque (F/T) sensor, and the active side of the novel grappling modules we present.

The grappling module is made up of an active fixture that is mounted after the F/T sensor on each arm, paired with a passive frame embedded in the center of the panel's backside. The active fixture features a servo actuator that locks the connection upon contact with the passive fixture. Both modules are designed with trapezoidal-shaped guides around the outer frame to help compensate for misalignments between the adapters to a certain extent.

Finally, each panel has two connection rods at the upper corners, which are inserted in the holes designed on the other side. The panels are kept lightweight to be withing the payload capacity of the robot arms, weighing $1\,\mathrm{kg}$ each.

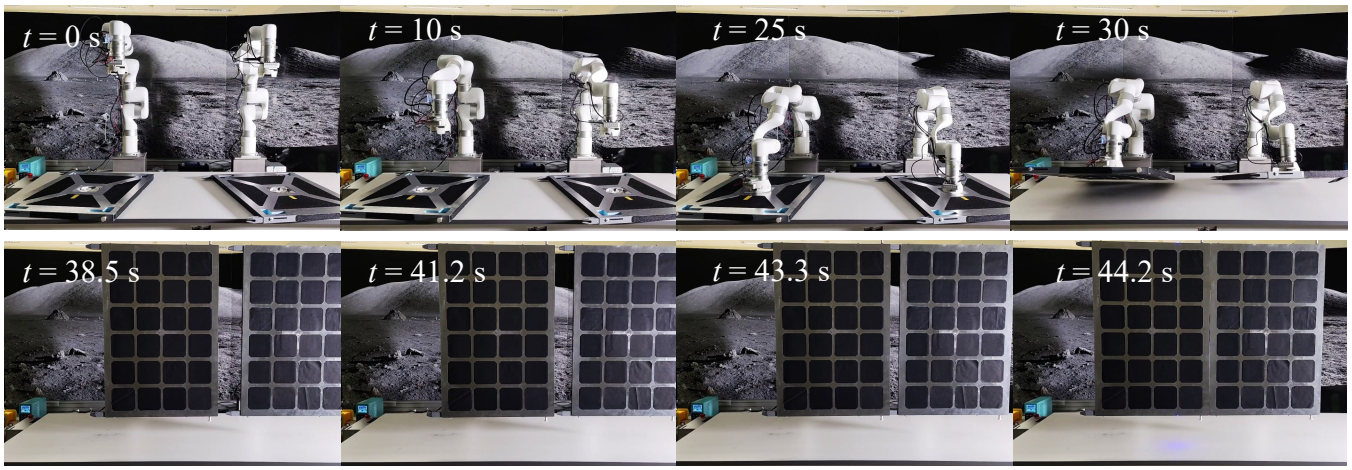
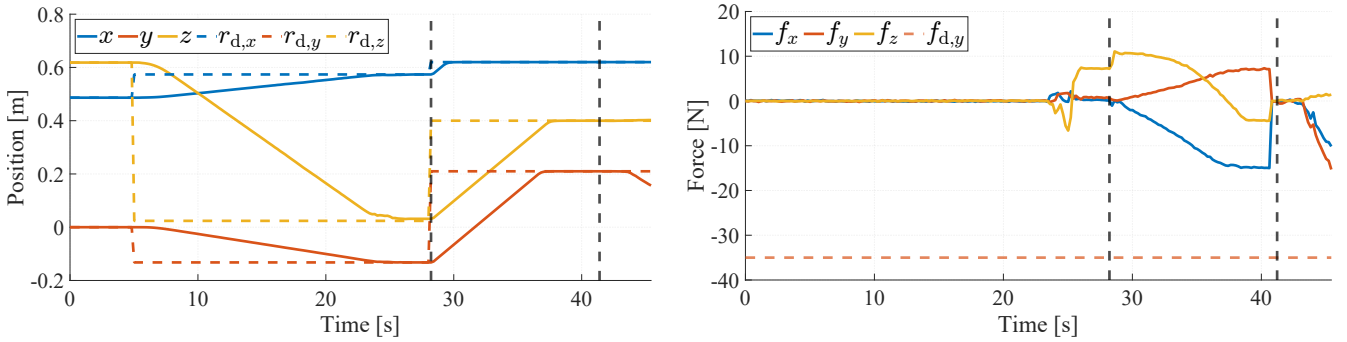


Fig. 5: Snapshots of the two panels assembling demonstration by dual robot manipulators in real world. End-effectors are approaching with visual feedback control the detected grappled fixture ($t = 10$ s), grasping them ($t = 25$ s), picking the panels up avoiding collision with NMPC ($t = 30$ s). Then, after achieving the lift-up positions ($t = 38.5$ s), the right panel is pushed into the other one with force control ($t = 41.2$ s), and thanks to impedance control the panel is compliant to the applied force ($t = 43.3$ s), eventually succeeding in the two panels connected ($t = 44.2$ s). The implemented autonomous sequence was successful irrespectively of the panels' initial pose.



(a) Position profile of the Yielding arm's end-effector expressed in the base frame Σ_B

(b) Force measured by the F/T sensor at the tip of the Driving arm expressed in the sensor frame Σ_S.

Fig. 6: The vertical black dashed line represents the phase distinction: at $t = 28.2$ s the lifting phase starts, while at $t = 41.2$ s the driving arm starts applying a desired $\boldsymbol{f}_{\text{ref}}$ at the end-effector level. At $t = 41.2$ s we reset the zero of the sensor so that we get rid of the effect of the panel. Contact between the two panels happens at around $t = 43.3$ s.

B. Results and Analysis

Until now, the input to the system was assumed as the one from (3). However, while the desired linear end-effector velocity can be directly issued to the xArm7, u_θ is not an available input. Since the available input to affect the end-effector orientation is the rate of change of the euler angles, we apply the following mapping

$$\boldsymbol{u}_\Phi = \boldsymbol{R}_{\text{A}} \boldsymbol{a} u_\theta \tag{12}$$

and the new input $\boldsymbol{u} = \begin{pmatrix} u_x & u_y & u_z & \boldsymbol{u}_\Phi \end{pmatrix} \in \mathbb{R}^6$ is issued to the robot.

A representative demonstration of the task of two panels assembly is showcased in Fig. 5. Analyzing the force profile measured by the F/T sensor mounted on the Driving arm reported in Fig. 6 (b), it is possible to recognize the contact between the arm and the panel at $t = 24$ s due to forces arising particularly along the z axis of Σ_S. These forces depend on how far off the connector is from the right position, but adjusting the parameters $\boldsymbol{M}_{\text{m}}, \boldsymbol{D}_{\text{m}}, \boldsymbol{K}_{\text{m}}$ in (1) also affects their magnitude. At $t = 41.2$ s the Driving arm starts applying the wrench $\boldsymbol{f}_{\text{d}}$ at the end-effector that, expressed in Σ_B, is $\boldsymbol{f}_{\text{d}} = \begin{pmatrix} * & -35 & 0 & * & * & * \end{pmatrix}$ where $*$ means that the robot will act rigidly in response to stimuli in those directions. Simultaneously, the Yielding arm will be compliant to external wrenches except along the x axis and around the z axis of Σ_B, which is equivalent to choosing the corresponding entries in $\boldsymbol{K}_{\text{m}}$ as $+\infty$. The two panels come in contact when the y coordinate of the Yielding arm's end effector starts changing in Fig. 6 (a) at $t = 43.3$ s, and the successful assembly is reached when the connecting rods slip into the connecting holes at $t = 44.2$ s, recognizable in Fig. 6 (b) as a temporary decrease in the measured f_y.

To measure the effectiveness of the proposed method quantitatively, in Table I we report the performance over 40 trials. The main factor that affects these results is a noisy depth estimation, leading to a failed grasp. Improving just this aspect, e.g., by improving the depth estimation from the Intel camera [23] or by using a more accurate 6D perception system [24], will bring a major increase in the pipeline success rate.

TABLE I: Performance of the proposed pipeline over 40 trials.

Success rate	Failure rate	Failure modality	Failure rate per modality
0.61	0.39	Failed grasp	0.66
		Failed panels insertion	0.34

C. Discussions

Together with depth estimation, the other key element for the performance of this pipeline concerns different choices of compliance and rigidity. Beside those discussed in Sec. IV-B, also other combinations have been tested: performance improved while allowing compliance along more axes. Furthermore, during the final assembly phase, we experimented with alternative controller pairs, with some pairs being more favourable. A more sistematic approach to find the right stiffness and damping values in the impedance controller, as well as the force value in the force controller, might improve the success of the overall pipeline.

Having both arms in impedance control required to craft a final goal position for the Yielding arm which would lead to contact between the panels: this is not in the nature of the task as the final assembly position varies depending on the exchanged forces. Having both arms in force control, the ability to control the nature of the interaction between the panels is lost, differently from what it is possible with impedance control, although the task of autonomous assembly can be done with comparable efficacy. Finally, having the Yielding arm in position control compromises the safety of the entire system, as small uncertainties and errors during insertion are met with complete rigidity on one side.

V. Conclusions

In this work, we developed a fully autonomous pipeline for solar panel assembly. The proposed method employs the YOLO visual perception algorithm together with depth information to extract the six dimensional pose of the detected object; then, a nonlinear Model Predictive Control scheme with a minimalist state representation is used for collision avoidance during motion; finally, to accommodate for uncertainties and errors, a combination of impedance and force control makes up the insertion phase. This holistic integration of vision, control, and specialised hardware demonstrates a robust and effective approach to complex assembly tasks in a multi-robot system.

The future scope of this work includes a learning-based approach to evolve between different states and control modes, and implement a method to address the noisy depth estimation limitation.

Acknowledgment

This work was supported by JST Moonshot R&D Program, Grant Number JPMJMS223B. Financial support to A. De Luca and L. Nunziante from PNRR MUR project PE0000013-FAIR is also acknowledged. The authors would like to thank Hamano Products Co., Ltd. and Prof. Fumitoshi Matsuno's research group in Moonshot R&D Program for their invaluable support in the development of the hardware platform, and are grateful to Pascal Pama for the assistance in conducting the experiments.

References

[1] Creech *et al.*, "Artemis: An overview of NASA's activities to return humans to the moon," in *Proc. IEEE Aerospace Conf.*, 2022, pp. 1–7.
[2] "U.S. and Japan spearhead new era of private moon landings," https://asia.nikkei.com/Business/Aerospace-Defense-Industries/U.S.-and-Japan-spearhead-new-era-of-private-moon-landings, 2024.
[3] G. Jocher *et al.*, "Ultralytics YOLO," Jan 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[4] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*. London: Springer, 2010.
[5] A. Krizhevsky *et al.*, "ImageNet classification with deep convolutional neural networks," *Communication of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
[6] Y. Xiao *et al.*, "A review of object detection based on deep learning," *Multimedia Tools and Applications*, vol. 79, no. 33, pp. 23 729–23 791, Sep 2020.
[7] T.-Y. Lin *et al.*, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–947.
[8] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.
[9] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," *arXiv:2402.13616*, 2024.
[10] A. Khatua *et al.*, "Developing approaches in building classification and extraction with synergy of YOLOV8 and SAM models," *Spatial Information Research*, 2024.
[11] C.-T. Chien, R.-Y. Ju, K.-Y. Chou, and J.-S. Chiang, "YOLOv9 for fracture detection in pediatric wrist trauma X-ray images," *arXiv:2403.11249*, 2024.
[12] B. Xiao *et al.*, "Fruit ripeness identification using YOLOv8 model," *Multimedia Tools and Applications*, vol. 83, no. 9, pp. 28 039–56, 2024.
[13] P. Song *et al.*, "A tutorial survey and comparison of impedance control on robotic manipulation," *Robotica*, vol. 37, no. 5, pp. 801–836, 2019.
[14] A. Isidori, *Nonlinear Control Systems*. London: Springer, 1995.
[15] S. Eppinger and W. Seering, "On dynamic models of robot force control," in *Proc. IEEE Int. Conf. Robot. Automat.*, 1986, pp. 29–34.
[16] B. Kouvaritakis and M. Cannon, *Model Predictive Control: Classical, Robust and Stochastic*. Cham: Springer, 2016.
[17] R. Findeisen *et al.*, "State and output feedback nonlinear model predictive control: An overview," *European Journal of Control*, vol. 9, no. 2, pp. 190–206, 2003.
[18] J. Qin and T. Badgwell, "An overview of industrial model predictive control technology," *AIChE Symp. Ser.*, vol. 93(1), pp. 232–256, 1997.
[19] R. Findeisen and F. Allgöwer, "An introduction to nonlinear model predictive control," in *Proc. 21st Benelux Meeting on Systems and Control*, 2002, pp. 119–141.
[20] G. H. Diaz *et al.*, "Toward autonomous assembly of modular robots and structures using real-time object detection and imitation learning for lunar missions," in *Proc. 29th Int. Symp. Artif. Life Robot. (AROB)*, 2024, pp. 1389–1393.
[21] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3974–3983.
[22] C. Boucher *et al.*, "Integration of vision-based object detection and grasping for articulated manipulator in lunar conditions," in *Proc. IEEE/SICE Int. Symp. Syst. Integrat. (SII)*, 2024, pp. 484–489.
[23] S. Rijal, S. Pokhrel, M. Om, and V. P. Ojha, "Comparing depth estimation of azure kinect and realsense d435i cameras," *Available at SSRN 4597442*, 2023.
[24] M. Jin, J. Li, and L. Zhang, "Dope++: 6d pose estimation algorithm for weakly textured objects based on deep neural networks," *PloS one*, vol. 17, no. 6, p. e0269175, 2022.