ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Stereo-Based Vision and Tactile Sensing for Robust Dual-Arm Robotic Connector Assembly

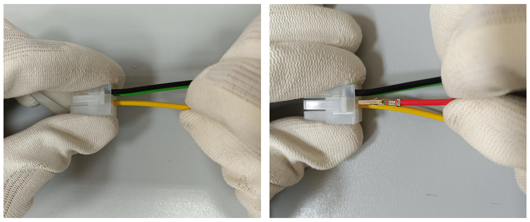
Michele Mirto[1], Alessio Caporali[2], Salvatore Pirozzi[1], Gianluca Palli[2]

***Abstract*— The automation of Deformable Linear Object (DLO) manipulation remains a key challenge in industrial production. While prior works demonstrated reliable wire terminal insertion using vision and tactile sensing, they typically assume a fixed connector pose. This paper presents a dual-arm robotic system for fully autonomous *connector assembly*. A stereo vision system enables robust 6D pose estimation of the wire terminal, while a custom mechatronic gripper with integrated tactile sensing supports accurate insertion monitoring. In parallel, the second arm performs connector grasping. By combining complementary visual and tactile feedback across both manipulators, the system achieves the precision required for tight-tolerance insertion without fixed fixtures.**

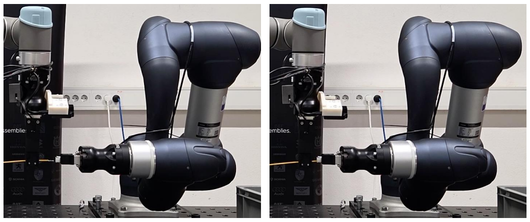
I. INTRODUCTION

The rising demand for industrial automation is driving the development of robotic systems capable of handling deformable objects [1], [2]. Unlike rigid parts, DLOs exhibit complex behaviors, including bending, twisting, and stretching, that make robust robotic manipulation and perception highly challenging [3], [4]. *Connector assembly* is a particularly demanding task within wire harness manufacturing, requiring the high-precision insertion of crimped metal terminals (*pins*) into designated connector housing slots. Traditionally, this task has relied on human operators who use a combination of visual and tactile feedback to ensure correct alignment and axial orientation [5], as shown in Fig. 1a.

In [6], a single-arm methodology to partially automate connector assembly was presented, relying on a custom mechatronic tool with stereo vision and tactile sensors. However, the connector housing was assumed to be fixed at a known position, limiting flexibility. To achieve fully autonomous assembly, this work adopts a dual-arm configuration. One robot handles the *pin-wire*, while the second grasps and positions the *connector*. While standard vision-based methods provide a coarse 6D pose estimate of the connector, their accuracy is insufficient for tight-tolerance insertion. To address this, we propose an in-hand connector pose refinement combined with stereo-based perception and tactile-guided insertion, enabling successful and robust connector assembly in realistic scenarios.



(a) manual insertion by a human operator



(b) autonomous insertion by the proposed dual-arm system

Fig. 1. Comparison between manual and autonomous connector assembly.

Fig. 2. Mechatronic tool for connector assembly comprising a converging stereo-based camera system and sensorized fingers.

II. SYSTEM OVERVIEW AND METHODOLOGY

A. Mechatronic Tool

The proposed methodology relies on a custom mechatronic tool (Fig. 2), built on a parallel-jaw gripper enhanced with a converging stereo vision system and sensorized fingers. The vision system uses two Luxonis OAK-1 cameras (68° orientation, $0.12\,\mathrm{m}$ baseline) to balance depth accuracy and minimize bulk. The fingers feature optoelectronic tactile sensors [7] consisting of a 5×5 matrix of photo-reflectors under a silicone pad, allowing precise detection of contact deformations during the assembly phases.

B. Connector and Pin Pose Estimation

Given the flexibility of the wire and potential misalignments, both the *connector* and the *pin* poses must be esti-

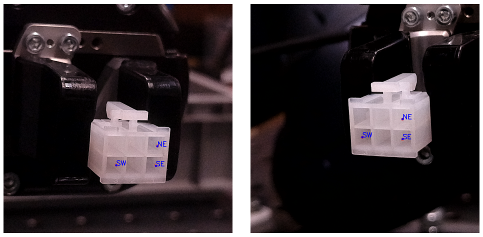
---

[1] Michele Mirto and Salvatore Pirozzi are with Dipartimento di Ingegneria, Università degli Studi della Campania “Luigi Vanvitelli”, Via Roma 29, 81031 Aversa (CE), Italy.

[2] Alessio Caporali and Gianluca Palli are with DEI - Department of Electrical, Electronic and Information Engineering, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy.

This work was supported by the Horizon Europe project *IntelliMan - AI-Powered Manipulation System for Advanced Robotic Service, Manufacturing and Prosthetics* [grant number 101070136]. Alessio Caporali is funded by FSE+ 2021-2027 (REF. PA: 2023-20090/RER - CUP: J19J23000730002).

Corresponding author: michele.mirto@unicampania.it



(a) left view (b) right view

Fig. 3. Detection of connector holes from stereo views using the YOLO-based model, achieving robust and precise in-hand connector pose.

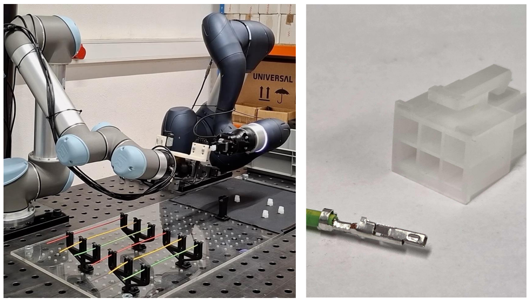


Fig. 4. Experimental setup: overview of the dual-arm robotic platform (left); detailed view of the connector and the pin (right).

mated dynamically. For the *connector*, the right arm's camera initially estimates a coarse 6D pose using DOPE [8] to allow the left arm to grasp it. Once grasped, the left arm presents the connector to the right arm's stereo cameras. A YOLO-based detection method identifies the connector holes in both views. By triangulating the matched detections, the 3D coordinates of the holes are reconstructed, yielding a refined, sub-millimetric 6D pose of the connector. The process is illustrated in Fig. 3.

Simultaneously, the right arm estimates the 6D pose of the *pin-wire*. A multi-head neural network processes the stereo images to predict 2D keypoints (*tip* and *tail*) and the axial rotation angle. The keypoints are triangulated to define the main axis of the pin in 3D space. This geometric fusion avoids the need for large datasets with full 6D pose annotations, ensuring robustness.

C. Tactile-Based Pin Insertion

Once the *pin-wire* is aligned with the connector housing slot, a tactile sensing strategy monitors the insertion. A contact indicator μ_x, representing the centroid displacement of the tactile response along the insertion axis, is analyzed in real-time. The insertion is executed in two stages: a *push phase* and a *pull phase*. A k-Nearest Neighbors (KNN) classifier, trained on tactile data from successful and failed insertions, continuously evaluates μ_x over sliding temporal windows. During the push phase, the robot advances until a *push completed* event is classified. Subsequently, a pull motion verifies locking: if a *pull completed* event is detected, the insertion is successful; otherwise, it is flagged as a failure.

III. PRELIMINARY EXPERIMENTS

The dual-arm robotic strategy is validated using a UR5, i.e. the right arm, equipped with the tool, and a Doosan manipulator, i.e. the left arm, handling the connector, operated within ROS2 (Fig. 4). The evaluated connector features $3.7{\times}3.7$ mm holes, while the pin is $1.6 \times 1.6 \times 14.8$ mm (Fig. 4 right).

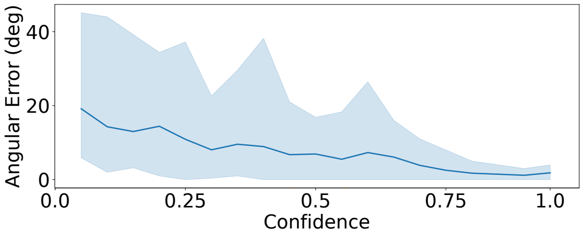


Fig. 5. Relationship between predicted confidence and axial angle estimation error. Higher confidence predictions exhibit lower mean error and smaller variance, indicating an inverse correlation.

A crucial component for successful insertion is the accuracy of the pin's axial angle estimation. The neural network outputs a confidence score alongside the angle prediction. As shown in Fig. 5, there is a strong inverse correlation between this confidence and the angular error. High-confidence predictions consistently yield low errors, making this score a reliable indicator for triggering recovery strategies in highly ambiguous visual conditions (e.g., reflections or low contrast). Overall, preliminary results confirm that the dual-arm configuration, coupled with continuous tactile monitoring, reliably achieves the tight tolerances required for connector assembly.

IV. CONCLUSIONS

This work introduced a dual-arm system for autonomous DLO connector assembly. By combining stereo-vision for refined in-hand pose estimation and tactile feedback for insertion monitoring, the proposed approach eliminates the need for fixed fixtures. Preliminary experiments demonstrate the system's precision and flexibility, marking a significant step forward compared to single-arm solutions.

REFERENCES

[1] H. Yin, A. Varava, and D. Kragic, "Modeling, learning, perception, and control methods for deformable object manipulation," *Sci. Robot.*, 2021.

[2] J. Trommnau, J. Kühnle, J. Siegert, R. Inderka, and T. Bauernhansl, "Overview of the state of the art in the production process of automotive wire harnesses, current research and future trends," *Proc. CIRP*, 2019.

[3] A. Caporali and G. Palli, "Robotic manipulation of deformable linear objects via multi-view model-based visual tracking," *IEEE/ASME T-MECH*, 2025.

[4] J. Guo, J. Zhang, D. Wu, Y. Gai, and K. Chen, "An algorithm based on bidirectional searching and geometric constrained sampling for automatic manipulation planning in aircraft cable assembly," *Journ. of Manufacturing Systems*, 2020.

[5] H. G. Nguyen, R. Habiboglu, and J. Franke, "Enabling deep learning using synthetic data: A case study for the automotive wiring harness manufacturing," *Proc. CIRP*, 2022.

[6] A. Caporali, M. Mirto, S. Pirozzi, and G. Palli, "Vision and tactile sensing for dlo manipulation and pin insertion in robotic connector assembly," *IEEE/ASME Transactions on Mechatronics*, 2026.

[7] A. Cirillo, G. Laudante, and S. Pirozzi, "Tactile sensor data interpretation for estimation of wire features," *Electronics*, 2021.

[8] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," in *Conference on Robot Learning (CoRL)*, 2018. [Online]. Available: https://arxiv.org/abs/1809.10790