Towards Coordinated Dual-arm Snap-fit Assembly Skill for Delicate Applications

Shreyas Kumar[1,*], Barat S[1], Debojit Das[1], Siddhi Jain[2], Rajesh Kumar[2], and Harish J. Palanthandalam-Madapusi[1]

***Abstract*— Delicate snap-fit assemblies, such as those in precision fits like inserting a lens into an eyewear frame or in electronics, demand timely engagement detection and rapid force attenuation to prevent overshoot-induced component damage or assembly failure. In this work, we introduce a bimanual manipulation framework that integrates both capabilities into a unified skill for snap-fit assembly. The system relies solely on joint-level proprioception: a learned model detects engagement from joint-velocity transients and subsequently triggers a task-aware stiffness modulation along the insertion axis. The bimanual policy is structured around a coupled dynamical system (DS) that coordinates synchronized transport motion with selective decoupling during insertion. We evaluate the framework across varied geometries and robot platforms, demonstrating its applicability to real-world snap-fit tasks. Project video: `https://shr-eyas.github.io/SNAP/`**

I. INTRODUCTION

Snap-fit mechanisms are widely used in industrial assembly due to their simplicity, speed, and cost-effectiveness. However, automating these insertions with bimanual robots for delicate parts like electronics or eyewear poses significant challenges, such as tight tolerances, brief but forceful engagement events, and the risk of overshoot-induced part damage or assembly failure.

A human performs such insertions with coordinated motion, natural recognition of engagement events, and intuitive modulation of compliance. We propose a unified framework that captures these characteristics, is deployable on bimanual robots, and scalable to humanoid systems in industrial settings. The key components are: (i) snap detection via joint-level proprioception, (ii) multi-arm coordination via coupled DS, and (iii) event-triggered compliance modulation to mitigate impact forces.

II. METHODOLOGY

Our framework is designed to support coordinated, compliant snap-fit insertions on bimanual platforms. We now describe the three key modules in the proposed method.

A. Snap Detection via Proprioception

Snap-fit insertions produce brief but repeatable transients in joint velocity at the moment of engagement. To detect these events reliably, we introduce SnapNet, a CNN-GRU-attention model trained on sequences of joint velocities from the inserting arm. The input $\dot{\mathbf{q}}_{1:t} \in \mathbb{R}^{t \times n}$ is a temporal window of joint velocities across n joints, that the model maps to a binary prediction of snap-fit engagement $\hat{y} \in \{0, 1\}$.

The authors gratefully acknowledge support for this research from Addverb Technologies Pvt. Ltd.

[1] IITGN Robotics Laboratory, Department of Mechanical Engineering, Indian Institute of Technology Gandhinagar, Gujarat 382355, India. [2] Addverb Technologies Private Limited, Noida, Uttar Pradesh 201305, India. *Corresponding author: shreyas.kumar@icloud.com.

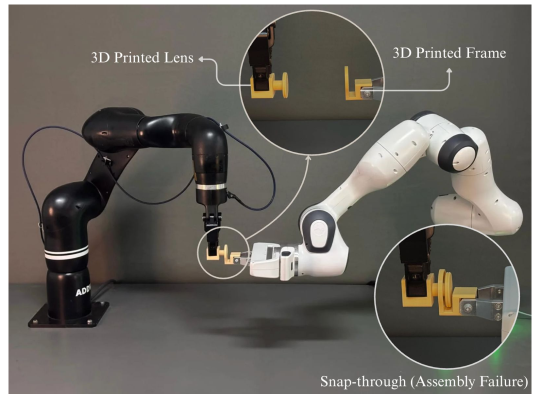

Fig. 1. Bimanual snap-fit task performed under a unified framework using Franka FR3 and Addverb Heal Cobot. The proposed method is directly applicable to snap-fit assembly skills for bimanual and humanoid robots.

B. Coupled Bimanual Coordination

The framework executes bimanual transport and insertion via a DS-based approach inspired by prior work on coupled, synchronous and asynchronous motion generation [1], [2]. Each end-effector trajectory evolves according to

$$\dot{z}_i = g_i(z_i) + \kappa(\theta)\, g_c(z_i), \qquad (1)$$

where $g_i(z_i)$ drives convergence toward the target, and $g_c(z_i)$ enforces coupling with the partner arm. These are defined as

$$g_i(z_i) = -k_1 \frac{z_i - z_i^d}{\left(|z_i - z_i^d| + \epsilon\right)^{1-\eta}}$$

and

$$g_c(z_i) = -k_2 \frac{z_i - z^c}{\left(|z_i - z^c| + \epsilon\right)^{1-\eta}}.$$

Here, z_i^d is the desired progress, z^c is the symmetric attractor for coordination, and k_1, k_2, η, ϵ are scalar gains and smoothing parameters. The coupling gain $\kappa(\theta) \in \{0, 1\}$ is binary: it is set to 1 for transport and alignment (synchronized motion), and switched to 0 during insertion (decoupled motion).

This formulation ensures smooth transitions and global convergence to the attractors, as the DS components are constructed using Lyapunov-stable vector fields.

C. Event-Triggered Compliance

Upon detecting the snap event, the system triggers an exponential stiffness ramp-down along the insertion axis. This compliance mechanism allows the robot to absorb post-engagement impact forces, which is essential for preventing overshoot in tight-tolerance or fragile assemblies. The stiffness adaptation is governed by a Variable Impedance Controller (VIC), which modulates the Cartesian impedance after successful engagement is detected.

We base the VIC on the Cartesian impedance law [3]

$$\mathbf{F}_{\text{cmd}}(t) = \mathbf{K}(t)\big(\mathbf{x}_{\text{ref}}(t)-\mathbf{x}(t)\big)+\mathbf{D}(t)\big(\dot{\mathbf{x}}_{\text{ref}}(t)-\dot{\mathbf{x}}(t)\big), \quad (2)$$

where $\mathbf{K}(t), \mathbf{D}(t) \in \mathbb{R}^{3\times3}$ are diagonal, positive-definite gain matrices. Let t_s denote the time of first snap detection, we then schedule

$$K(t) = \begin{cases} K_0, & t < t_s, \\ K_0 + \big(K_{\text{f}} - K_0\big)\, e^{-\lambda(t-t_s)}, & t \geq t_s, \end{cases} \quad (3)$$

where K_0 and K_{f} are the pre- and post-snap stiffness levels, and $\lambda > 0$ is the decay rate. This design ensures that the robot behaves rigidly during alignment and only softens its response once snap engagement is confirmed.

III. EXPERIMENTAL RESULTS

A. SnapNet: Proprioceptive Engagement Detection

Experiments were conducted on two manipulators. All SnapNet training was performed on a Franka Emika FR3, a 7-DoF collaborative arm whose joint-velocity measurements, sampled at 100 Hz, forming the basis of our model. Training data was collected from ~ 500 manually labeled insertion trials across 7 objects, each chosen to represent diverse snap-fit mechanics. We deployed the same trained network on an Addverb Heal Cobot, an industrial 6 DoF arm with a different kinematic structure. Results are summarized in Table I.

TABLE I

SNAPNET DETECTION SUCCESSES (15 TRIALS PER CONDITION)

Snap-Fit Part	**FR3**	**Heal**
Marker Pen Cap	14/15(93.3%)	13/15(86.7%)
Highlighter Cap	15/15(100.0%)	14/15(93.3%)
Bottle Lid Clip	15/15(100.0%)	—
USB-C/Lightning Connector	13/15(86.7%)	8/15(53.3%)
3D-printed Lens & Frame	15/15(100.0%)	12/15(80.0%)
Emergency-Stop Button	—	15/15(100.0%)
Custom 3D-printed Part	—	10/15(66.7%)
Overall Recall	**96.0%**	**80.0%**

B. Bimanual Snap-Fit Assembly

We evaluate our framework for a snap-fit task on 3D-printed test parts simulating the insertion of a lens into an eyewear frame, demanding sub-millimeter alignment and highly succeptible to assembly failure. The task is executed under the previously described bimanual DS-based coordination policy: the FR3 stabilizes the frame while the Heal cobot performs the insertion. We first compare two baseline strategies: position control and fixed-gain Cartesian impedance. As shown in Table II, position control achieved only 40% success due to excessive rigidity, often leading to jamming and breakage. Fixed impedance improved success to 73% by introducing compliance, yet remained vulnerable to snap-through and breakage due to unmitigated release of stored elastic energy. We then deploy the proposed event-triggered variable impedance controller (VIC), which initiates an exponential decay of Cartesian stiffness along the insertion axis upon engagement detection by SnapNet. This reduces peak forces while maintaining lateral rigidity. VIC achieves success in $15/15$ trials and reduces contact forces by 30% compared to fixed impedance.

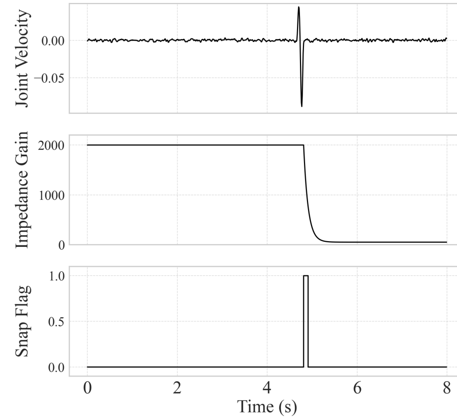

Fig. 2. Top: Joint velocity transient indicating snap event. Middle: Exponential decay of impedance gain triggered post-engagement. Bottom: Binary snap detection flag: initiates compliance modulation.

TABLE II

ASSEMBLY SUCCESS AND PEAK FORCES (LENS-FRAME TASK)

Control Mode	**Success Rate**	$\mathbf{F}_{\max}$ (N)
Position Control	6/15 (40%)	25.1 ± 3.8
Fixed Impedance	11/15 (73%)	22.4 ± 4.2
Event-Triggered VIC (Proposed)	15/15 (100%)	16.3 ± 3.1

IV. CONCLUSIONS

We introduced a unified snap-fit assembly framework for bimanual robots, integrating proprioceptive detection, dynamic coordination, and event-triggered compliance into a single skill. Our approach requires no external sensing and is designed to be adaptable across hardware platforms. While demonstrated on a test assembly simulating a lens-in-frame snap-fit, the framework is applicable to a range of sensitive snap-fit tasks in industrial settings. This work represents a step toward equipping humanoids with robust assembly capabilities for complex, contact-rich tasks.

REFERENCES

[1] F. Khadivar and A. Billard, "Adaptive fingers coordination for robust grasp and in-hand manipulation under disturbances and unknown dynamics," *IEEE Transactions on Robotics*, vol. 39, no. 5, 2023.

[2] S. S. M. Salehian, N. Figueroa, and A. Billard, "A unified framework for coordinated multi-arm motion planning," *The International Journal of Robotics Research*, vol. 37, no. 10, 2018.

[3] N. Hogan, "Impedance control: An approach to manipulation," in *1984 American Control Conference*, 1984.