Gesture-Based Cartesian Control with Real-Time Feedback for Object Positioning and Screw Insertion

M A Hafiz
Friedrich-Alexander-Universität Erlangen–Nürnberg (FAU)
Email: mahafizsourav@gmail.com

***Abstract*—Most gesture-based robot control interfaces map to joint commands or lack precise task-space feedback. We present a gesture-based Cartesian control framework that integrates EMG and IMU signals with an SVM-based classifier (15 s per gesture training). Users control end-effector translation (X, Y, Z) and rotation (roll, pitch, yaw) through a hierarchical interface that combines IMU-based menu navigation with gesture-based axis selection and positive/negative motion commands in task space. A real-time Cartesian error feedback mechanism provides positional differences with respect to the target, enabling iterative refinement during task execution. The system is evaluated on a Franka Emika FR3 robot with three users performing object positioning and screw insertion tasks. The SVM classifier achieves high recognition accuracy (97–98%) across users. Positioning errors decrease from approximately 200 mm to 3–12 mm within five trials, while screw insertion achieves 4–18 mm accuracy with rotational error below 0.15°. These results suggest that gesture-based Cartesian control with real-time feedback can support intuitive control, rapid user adaptation, and preliminary performance in contact-rich manipulation tasks.**

***Index Terms*—Gesture-based control, Cartesian control, EMG, IMU, human–robot interaction, teleoperation, real-time feedback, robot manipulation**

I. Introduction

Intuitive human–robot interaction is increasingly important as robots move toward collaborative environments. However, most existing control interfaces rely on low-level joint commands or predefined motion programs, which require training and do not naturally express high-level human intent.

Wearable gesture-based control using surface electromyography (EMG) provides an intuitive alternative [1]. However, many gesture-based approaches use recognized gestures as discrete commands or low-level robot actions [2], [3], which can limit their effectiveness in task-space manipulation requiring precise end-effector alignment.

In this work, we propose a gesture-based Cartesian control framework in which EMG and IMU signals are integrated within a hierarchical interface to control end-effector translation (X, Y, Z) and rotation (roll, pitch, yaw) in task space through axis selection and incremental motion commands. Unlike prior approaches, the proposed system incorporates real-time Cartesian error feedback that provides continuous information about the positional difference with respect to a target, enabling human-in-the-loop refinement during task execution.

In this work, semantic reasoning is considered at the level of human–robot command abstraction rather than scene understanding. The proposed interface abstracts low-level EMG and IMU biosignals into meaningful robot action primitives, such as axis selection, direction selection, locking, unlocking, return, and screw insertion. This provides a semantic action-selection layer that bridges human intent and low-level Cartesian robot execution.

The system consists of three components: (1) multimodal gesture recognition using an SVM classifier, (2) a gesture-to-Cartesian control mapping, and (3) a real-time feedback mechanism. We validate the approach on screw placement and insertion tasks using a Franka Emika FR3 robot [4] with three users. Experimental results show rapid user adaptation, with positioning errors decreasing from approximately 200 mm to 3–12 mm within five trials.

Contributions:

- A gesture-based Cartesian control framework using multimodal EMG and IMU signals
- A lightweight user-specific training approach requiring only 15 seconds per gesture
- A real-time Cartesian error feedback mechanism enabling precise human-in-the-loop control in contact-rich manipulation tasks

II. Related Work

Gesture recognition using EMG and IMU signals has been widely studied for prosthetic control, assistive robotics, and human–robot interaction [1], [3], [2]. Surface EMG can capture muscle activation patterns related to intended hand or arm motions [1], [3], while IMU sensing provides complementary information about limb orientation and movement [2]. These modalities are therefore well suited for wearable robot-control interfaces.

Many gesture-based robot control approaches use recognized gestures as discrete commands or map them to low-level robot actions [2]. While such mappings are useful for simple interaction, they can become less intuitive for manipulation tasks that require precise end-effector positioning. Cartesian control provides a more task-oriented alternative because users can reason directly in terms of end-effector translation and rotation rather than individual robot joints [5].

Teleoperation and human-in-the-loop robot control studies have shown that feedback is important for improving task awareness, user adaptation, and control performance [6]. However, gesture-based control systems often rely mainly on visual observation of the robot motion, requiring users to correct errors through trial and error [2]. Providing explicit task-space error feedback can support more precise iterative correction during manipulation [6], [5].

Contact-rich manipulation requires stable physical interaction between the robot and the environment. Impedance and compliant control methods are commonly used to support such interaction [7]. However, these methods do not directly address how non-expert users should provide intuitive task-space commands for alignment and insertion. In contrast, the proposed framework combines EMG and IMU-based gesture recognition, semantic action selection, Cartesian control, and real-time positional feedback for human-in-the-loop object placement and screw insertion.

III. System Overview

The proposed system enables gesture-based Cartesian control through four stages: signal acquisition, gesture classification, command mapping, and feedback (Fig. 1).

Sensing: A Muovi 32-channel EMG bracelet captures muscle activation patterns, while an integrated IMU provides orientation information used for interface navigation.

Gesture Recognition and Interface: An SVM classifier recognizes six discrete gestures: rest, positive action, negative action, lock axis, unlock axis, and return to the main menu. The classifier is adapted from an existing pipeline and retrained with user-specific calibration data (15 s per gesture), achieving high recognition accuracy.

Instead of directly mapping each gesture to a separate Cartesian command, a hierarchical menu-based interface is used. This design reduces the number of gestures that users must learn and classify, while still allowing access to multiple robot actions. With a limited gesture set, users can select between task options such as object placement, Cartesian axes, rotational axes, screw, and unscrew. IMU orientation is used to scroll through menu options, while EMG gestures are used for confirmation and action commands. Once an option is selected, users apply lock/unlock gestures to confirm or leave the selected mode. Positive and negative gestures are then used to generate motion along the selected axis or execute the selected task-level primitive.

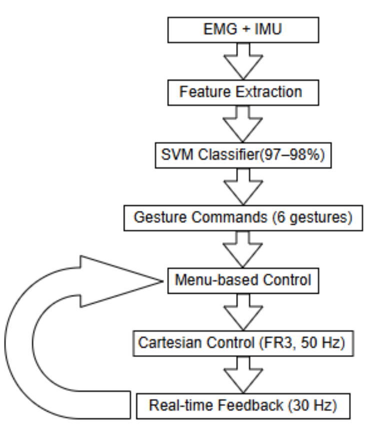

Fig. 1. Gesture-based Cartesian control pipeline.

Fig. 2. Gesture set used in the interface: rest, return, lock, unlock, positive action, and negative action.

TABLE I
Context-dependent gesture-to-command mapping.

Locked mode	Positive gesture	Negative gesture
Pose capture at lock	$p_{\text{ref}} = [x_{\text{ref}}, y_{\text{ref}}, z_{\text{ref}}]$, $q_{\text{ref}} = [q_x, q_y, q_z, q_w]$	
X	$x_{\text{new}} = x_{\text{ref}} + \Delta x$	$x_{\text{new}} = x_{\text{ref}} - \Delta x$
Y	$y_{\text{new}} = y_{\text{ref}} + \Delta y$	$y_{\text{new}} = y_{\text{ref}} - \Delta y$
Z	$z_{\text{new}} = z_{\text{ref}} + \Delta z$	$z_{\text{new}} = z_{\text{ref}} - \Delta z$
Roll/Pitch/Yaw	$q_{\text{ref}} \rightarrow$ Euler $\rightarrow \alpha + \Delta\alpha \rightarrow q_{\text{new}}$	$q_{\text{ref}} \rightarrow$ Euler $\rightarrow \alpha - \Delta\alpha \rightarrow q_{\text{new}}$
Screw mode	yaw rotation + Z-down	reverse yaw rotation + Z-up
Output command	$[x, y, z, q_x, q_y, q_z, q_w] \rightarrow$ robot controller	

For rotational commands, α denotes the selected Euler component, i.e., roll ϕ, pitch θ, or yaw ψ.

This hybrid interface combines continuous IMU-based navigation with discrete EMG-based control, enabling intuitive interaction with a limited gesture set. Fig. 2 shows the gesture set, while Table I summarizes how the same positive and negative gestures are interpreted depending on the locked menu option.

Cartesian Control: Recognized gestures are translated into task-space motion commands through the menu interface. For object placement, users select one translational axis (X, Y, or Z) and generate incremental positive or negative end-effector motion along the selected axis. Rotational axes can also be selected for orientation refinement when required.

The gesture-to-motion mapping is implemented as an incremental pose update. When a user selects and locks an axis or task mode, the system stores the current robot pose as the reference pose. The pose is represented as a 7D Cartesian target pose $[x, y, z, q_x, q_y, q_z, q_w]$, where position is given in Cartesian coordinates and orientation is represented as a quaternion. Once an axis is locked, each positive action gesture increments the selected variable by one fixed step, while each negative action gesture decrements it. All non-selected pose components remain unchanged. The updated target pose is then streamed to the robot controller.

For translational control, the selected Cartesian component is updated directly. For example, if the Z axis is locked, one positive action gesture updates the target as

$$z_{\text{new}} = z_{\text{ref}} + \Delta z,$$

while x, y, and the orientation remain unchanged. In the implemented system, each action gesture corresponds to a translational step of $\Delta p = 0.0008$ m along the selected Cartesian axis. For rotational refinement, each action gesture applies an angular increment of 0.004 rad ($\approx 0.23°$), and the updated orientation is represented as a quaternion before being sent to the robot.

For contact-rich screw insertion, the interface provides task-level screw and unscrew options. The screw command generates a coupled motion primitive consisting of yaw rotation and downward Z-axis motion, while the unscrew command generates the reverse primitive consisting of opposite yaw rotation and upward Z-axis motion. This avoids requiring users to manually alternate between separate yaw and Z-axis commands during insertion.

Commands are executed through the Franka Emika FR3 Cartesian control interface at 50 Hz, with velocity filtering applied to ensure smooth motion. Object placement uses position-based incremental updates for precise alignment, while screw insertion uses velocity-based coupled yaw and Z motion to enable continuous rotational and translational interaction during contact.

Real-time Feedback: Positional errors along the X, Y, and Z axes with respect to the target are displayed on a monitor at 30 Hz. This feedback allows users to iteratively refine alignment during task execution.

Although the Franka Emika FR3 robot has 7 degrees of freedom, control is performed in 6-DOF Cartesian space, with redundancy handled by the robot controller. The FR3 robot provides torque-controlled actuation, enabling compliant interaction during contact-rich tasks such as screw insertion. This closed-loop human-in-the-loop framework allows precise manipulation without requiring joint-level programming.

IV. Experiments

The proposed system was evaluated with three participants (P01–P03) performing screw placement and insertion tasks on a Franka Emika FR3 robot (Fig. 3). Each participant completed five trials per task (30 trials total) following a short calibration phase of approximately 15 seconds per gesture using an SVM classifier. The SVM classifier achieved high accuracy across users, averaging 97–98% during the calibration phase. All experiments were conducted in a controlled laboratory environment, with participants standing in front of a monitor displaying real-time feedback.

All participants were informed about the experimental procedure and provided consent before participation. The study involved non-invasive wearable sensing and supervised interaction with a robot system. No personally identifying participant information is reported.

Screw Placement: The task requires positioning the end-effector above a target, aligning the screw with the hole, and placing it accurately using translational (X, Y, Z) and rotational gestures. This task primarily evaluates fine positioning accuracy and alignment in Cartesian space.

Fig. 3. Real-world screw insertion with gesture-based Cartesian control.

Screw Insertion: For the contact-rich task, the interface provides dedicated screw and unscrew options rather than requiring the user to manually alternate between yaw and Z-axis commands. In screw mode, the command generates a coupled motion primitive consisting of yaw rotation and downward Z-axis motion. In unscrew mode, the command generates the reverse primitive, combining opposite yaw rotation with upward Z-axis motion. This evaluates performance in contact-rich manipulation requiring coordinated rotational and translational motion under torque-controlled interaction, with contact forces in the range of 1–20 N.

Real-time Feedback: Positional errors along the X, Y, and Z axes with respect to the target are displayed at 30 Hz, enabling users to iteratively refine alignment. The displayed feedback is computed from the difference between the current end-effector position and the predefined target position, and is presented numerically and visually to support correction during task execution.

Metrics: Performance was evaluated using final positional error, rotational error, and consistency across repeated trials. Final positional error was defined as the Euclidean distance between the final measured end-effector position and the target position:

$$e_p = \|p_{\text{final}} - p_{\text{target}}\|_2.$$

In addition, axis-wise errors along X, Y, and Z were displayed to the user in real time to support iterative correction. For the screw insertion task, rotational error was computed as the absolute angular deviation between the desired screw yaw orientation and the measured final yaw angle. Repeated trials were used to analyze user adaptation and learning behavior. Task completion time was not used as a primary metric in this preliminary study because the focus was on accuracy, controllability, and human-in-the-loop refinement rather than speed.

V. Results and Discussion

Three users completed 30 trials across screw placement and insertion tasks, showing an overall reduction in positioning error across repeated trials (Figs. 4-5).

Quantitative Results:

Initial positioning errors decreased from ∼200 mm to 3–12 mm within five trials (Fig. 4), demonstrating rapid user adaptation. For screw insertion, users achieved consistent accuracy of 4–18 mm with rotational error below 0.15° (Fig. 5).

The real-time feedback (30 Hz display of X, Y, Z positional error) enabled iterative correction, while the FR3 torque control ensured stable contact interaction within 1–20 N. The short calibration time (15 s per gesture) allowed quick adaptation across users.

Discussion: The results indicate that combining gesture-based Cartesian control with real-time feedback can support intuitive manipulation in contact-rich tasks. Overall, the reduction in final positioning error across repeated trials suggests rapid user adaptation. However, participant-level variation was observed. In particular, P02 showed increased error in later trials, which may be attributed to inconsistent gesture execution, overcorrection while using the real-time feedback display, or differences in how the participant interpreted the incremental motion commands.

This variation highlights the need for larger user studies and more robust interaction mechanisms, such as adaptive gesture thresholds, confidence-based gesture rejection, or additional smoothing of commanded motion. Limitations of the current study include the small sample size (3 users), the absence of a baseline comparison such as joystick or keyboard-based Cartesian control, and the lack of task completion time analysis.

TABLE II
Performance metrics across 30 trials (mean $\pm$ std).

Metric	Screw Placement	Screw Insertion
Final Pos. Error (mm)	7.5 ± 4.2	11.2 ± 5.1
Rotational Error (°)	–	0.12 ± 0.04
Contact Force (N)	–	1–20

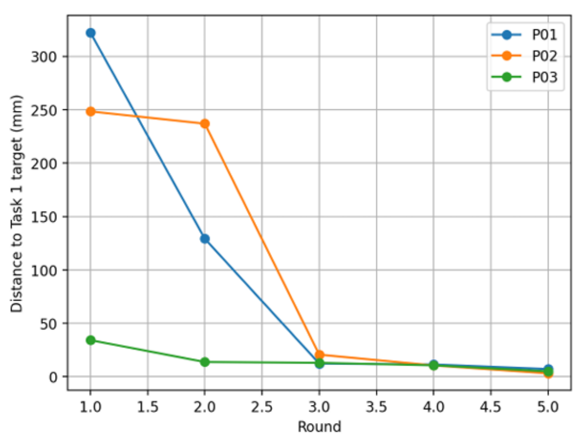

Fig. 4. Reach task: Positional error decreases from ∼200 mm to 3–12 mm over 5 trials.

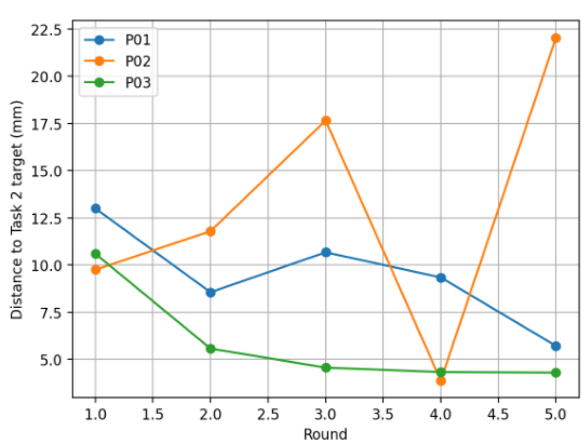

Fig. 5. Screw insertion: Consistent 4–15 mm accuracy across trials.

VI. Conclusion

We presented an EMG+IMU-based gesture-driven Cartesian control framework with real-time feedback for human-in-the-loop robot manipulation. The system maps a limited set of wearable gestures to semantic task-space action primitives through a hierarchical menu interface, enabling Cartesian positioning and screw/unscrew control on a Franka Emika FR3 robot.

Preliminary experiments with three users suggest that the proposed interface can support rapid user adaptation and accurate positioning in contact-rich tasks, with final positioning errors reaching 3–12 mm across repeated trials. However, the study is limited by the small number of participants and the absence of a baseline comparison. Future work will focus on larger-scale user studies, comparison against joystick or keyboard-based Cartesian control, and integration of vision-based target perception.

References

[1] D. Farina, R. Merletti, and R. M. Enoka, "The extraction of neural strategies from the surface EMG," *J. Appl. Physiol.*, vol. 96, no. 4, pp. 1486–1495, 2004.

[2] Z. Wang, Y. Chen, and X. Liu, "Hand and Arm Gesture-Based Human–Robot Interaction: A Review," arXiv:2209.08229, 2022.

[3] T. Song, Z. Yan, S. Guo, Y. Li, X. Li, and F. Xi, "Review of sEMG for robot control: techniques and applications," *Appl. Sci.*, vol. 13, no. 17, Art. no. 9546, 2023.

[4] Franka Emika GmbH, "FR3 Robot Documentation," 2023. [Online]. Available: https://www.franka.de

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

[6] T. B. Sheridan, "Human–robot interaction: status and challenges," *Hum. Factors*, vol. 58, no. 4, pp. 525–532, 2016.

[7] N. Hogan, "Impedance control: An approach to manipulation," *J. Dyn. Syst., Meas., Control*, vol. 107, no. 1, pp. 1–24, 1985.