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

# Low-Dimensional Tactile Glove for Visuo-Tactile Robot Hand Control: A Preliminary Study

Gi-Gwang Baek and Dongwook Kim†
Department of PIRoL, DGIST, Daegu, Republic of Korea
{gigwang.baek, dw_kim}@dgist.ac.kr

***Abstract*—Dexterous control of multi-joint manipulators such as humanoid robot hands using vision alone faces fundamental limitations. Cameras cannot directly observe contact forces, slip, and deformation arising from robot-environment interactions, and are vulnerable to occlusion, making them insufficient for contact-rich manipulation tasks. Tactile sensing is therefore considered an essential component for dexterous manipulation. However, existing tactile-based approaches rely primarily on high-performance, high-cost sensors, imposing significant cost burdens in collecting human demonstration datasets and deploying tactile sensors on robot hands, which limits the scalability of tactile information in robotic systems.**

**In this work, we present a low-dimensional wearable tactile glove as a scalable platform for visuo-tactile robot hand control, and propose a two-level learning framework built upon it. The tactile glove hardware consists of 20 FSR400 piezoresistive sensors, achieving stable 300 Hz data acquisition via dual multiplexers, WiFi TCP communication, and clock synchronization. Hardware validation confirms 300.03 Hz sampling, 20.1 $\mu$s jitter, and noise-free signal quality, demonstrating that the system meets the requirements for tactile-based learning and control.**

**The Level 1 framework investigates whether binary tactile signals alone can recover meaningful force distributions, aiming to show that low-cost, low-dimensional sensing can approximate high-dimensional information through uncertainty-aware learning. In Level 2, 300 Hz tactile signals are combined with RGB camera input to pretrain a shared visuo-tactile representation via contrastive learning. Furthermore, by integrating the Level 1 framework into Level 2, we aim to explore whether high-quality shared visuo-tactile representations can be pretrained using low-dimensional tactile inputs through transformer-based models. Subsequent work will validate the pretrained representations through imitation learning and reinforcement learning, and apply the full system to real-time physical robot hand control.**

***Index Terms*—Tactile Sensing, Visuo-Tactile Learning, Dexterous Manipulation, Wearable Sensors, Robot Hand Control.**

## I. System Design

### A. Hardware

Fig. 1 illustrates the overall system architecture. Table I lists the hardware components used in the system.

The tactile glove integrates 20 FSR400 pressure sensors mounted on a standard fabric glove, arranged to mirror the contact regions of a robot hand for direct layout transfer. Two 16-channel analog multiplexers (MUX1: FSR 1–10, MUX2: FSR 11–20) share selection lines (S0–S3) driven by an ESP32 microcontroller, enabling parallel scanning of all 20 channels. Each FSR is paired with an independent 10 k$\Omega$ pull-down resistor, sampled via the ESP32's 12-bit ADC at 0–3.3 V input range with a 20 $\mu$s stabilization delay per channel.

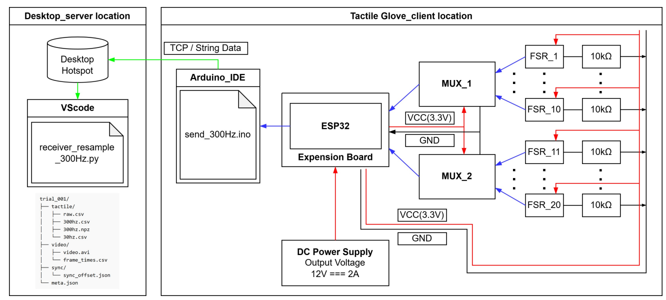


Fig. 1. System architecture of the tactile glove platform. The glove (client) streams 300 Hz tactile data over WiFi TCP to the desktop server for real-time processing and dataset storage.

TABLE I
Hardware Components

| Component | Specification | Quantity |
|---|---|---|
| ESP32 | Microcontroller | 1 |
| Expansion Board | GPIO Expansion | 1 |
| Multiplexer | Analog MUX | 2 |
| FSR Sensor | FSR 400 | 20 |
| Resistor | 10 k$\Omega$ | 20 |
| Breadboard | 5.5 $\times$ 16.5 cm | 1 |
| Power Supply | 12 V / 2 A | 1 |

### B. Software Pipeline

The ESP32 streams sensor data wirelessly over WiFi TCP at 300 Hz. Each packet contains a sequence number, an ESP32 hardware timer timestamp ($\mu$s resolution), and 20 ADC values. On the host PC, an NTP-style exchange protocol estimates the clock offset between the ESP32 and PC, enabling precise temporal alignment with external sensors such as RGB-D cameras. Incoming data are stored as timestamped datasets in both CSV and NPZ formats.

## II. Proposed Framework

### A. Overview

Our approach aims to establish a scalable visuo-tactile learning pipeline that bridges human tactile sensing and robot hand control. The overall system consists of three components: (1) a low-dimensional wearable tactile glove for data acquisition, (2) a two-level learning framework for tactile and visuo-tactile representation learning, and (3) an embodiment-aligned transfer mechanism enabling deployment on robot hands. We first collect synchronized tactile and visual data from human demonstrations using the wearable glove. Based on this data, we progressively learn representations through a two-stage framework, starting from low-dimensional tactile signals and scaling toward cross-modal visuo-tactile representations.

### B. Level 1: Binary Tactile → Force Inference

At the first level, we investigate whether minimal tactile signals can recover meaningful force distributions. The input consists of binary tactile signals indicating contact states across the 20 sensors. Despite their simplicity, these signals implicitly encode spatial and temporal patterns of interaction. We formulate this as a learning problem that maps sparse binary inputs to continuous force representations. To account for ambiguity and sensor sparsity, uncertainty-aware learning is incorporated, allowing the model to infer plausible force distributions. This stage aims to demonstrate that low-dimensional tactile sensing can approximate high-dimensional contact information through learning.

### C. Level 2: Visuo-Tactile Pretraining

In the second level, we extend the framework to cross-modal representation learning by combining high-frequency tactile signals with visual observations (RGB or RGB-D). This is achieved through: (1) contrastive alignment between visual and tactile features, and (2) masked cross-modal prediction to capture inter-modal dependencies. To further enhance scalability, the Level 1 model can be integrated into this stage, allowing the system to leverage force representations inferred by Level 1 instead of raw binary signals. The learned visuo-tactile representations can be used for downstream manipulation via imitation learning or reinforcement learning.

### D. Embodiment Alignment

A key design principle of our system is embodiment alignment. The tactile sensor layout of the glove mirrors the structure of a robot hand, enabling direct correspondence between human tactile signals and robot contact locations. This allows learned representations to be transferred without additional hardware adaptation, making the framework more scalable across different robotic platforms.

## III. Preliminary Results

We conducted three hardware validation experiments on the assembled tactile glove. Results are summarized in Table II.

**Acquisition timing.** The system achieved a stable sampling rate of 300.03 Hz with an inter-frame interval jitter of 20.1 $\mu$s (std), corresponding to 0.6% of the 3333 $\mu$s target period, confirming timing stability sufficient for deep learning input pipelines.

TABLE II
HARDWARE VALIDATION RESULTS

| Metric | Target | Result | Status |
|---|---|---|---|
| Sampling rate | 300 Hz | 300.03 Hz | Nominal |
| Interval jitter | $<$500 $\mu$s | 20.1 $\mu$s | Nominal |
| Static noise std | $<$15 ADC | 0.00 ADC | Nominal |
| Channel response | 20/20 | 20/20 | Nominal |
| Press amplitude | $>$200 ADC | +996$\sim$+1509 ADC | Nominal |
| Crosstalk | $<$30 ADC | None detected | Nominal |

**Static noise.** With the glove at rest and no pressure applied, all 20 channels recorded a mean value of 0.00 ADC (std = 0.00), indicating a clean and noise-free baseline signal.

**Channel response.** All 20 channels responded normally following replacement of a defective multiplexer unit. Press amplitude ranged from +996 to +1509 ADC across channels, demonstrating consistent and distinguishable sensor response. No crosstalk was detected on any tested channel.

These results validate the glove as a reliable data collection platform for subsequent visuo-tactile learning experiments.

## IV. Conclusion & Future Work

We present a low-cost wearable tactile glove achieving stable 300 Hz acquisition as a foundation for visuo-tactile pretraining toward dexterous robot hand control. The sensor layout is designed to mirror the robot hand, enabling direct representation transfer across embodiments. Future work will focus on collecting large-scale visuo-tactile datasets from human demonstrations and leveraging them for visuo-tactile pretraining to learn robust cross-modal representations. We aim to enable effective representation transfer from human sensing to robot hands through embodiment alignment, and to integrate these representations into downstream manipulation tasks via imitation learning and reinforcement learning. Ultimately, this framework will be extended toward real-time closed-loop robot hand control, while maintaining scalability to higher-density and vision-based tactile sensing modalities.

## References

[1] Q. Ye et al., “Visual-tactile pretraining and online multitask learning for humanlike manipulation dexterity,” *Science Robotics*, vol. 11, eady2869, 2026.

[2] Q. Liu, Q. Ye, Z. Sun, Y. Cui, G. Li and J. Chen, “Masked Visual-Tactile Pre-training for Robot Manipulation,” in *Proc. IEEE ICRA*, Yokohama, Japan, 2024, pp. 13859–13875.

[3] A. George, S. Gano, P. Katragadda and A. B. Farimani, “VITaL Pretraining: Visuo-Tactile Pretraining for Tactile and Non-Tactile Manipulation Policies,” in *Proc. IEEE ICRA*, Atlanta, GA, USA, 2025, pp. 258–264.

[4] Y. Yang et al., “Touch and Go: Learning from Human-Collected Vision and Touch,” in *Proc. NeurIPS*, 2022.

[5] M. Lambeta et al., “DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor with Application to In-Hand Manipulation,” *IEEE RA-L*, vol. 5, no. 3, pp. 3838–3845, 2020.