Eucalyptus: A High-Fidelity Multimodal UMI

Igor Fala, Zubin Kremer Guha, Amar Hajj-Ahmad, Vidullan Surendran, Tim Fofonoff, Ben Thacher, Richard Burnham, Murphy Wonsick, Zhi Ern Teoh, Ciarán T. O'Neill, Peter Whitney, David Watkins

***Abstract*—Portable data-collection interfaces enable the acquisition of real-world manipulation demonstrations without requiring a robot during collection. We present Eucalyptus, an extension of UMI designed for sustained, force-aware data collection. Eucalyptus combines a robustified and ergonomics-oriented mechanical design with a high-fidelity multimodal sensing suite comprising RGB vision, gripper width, interaction forces, and 6-DoF pose. The mechanical redesign eliminates measurable backlash, expands the maximum gripper width, improves ergonomics for a wide range of hand sizes, and supports an expanded sensing stack. Together, these changes improve the experience of data collection for the operator and provide richer measurements of the demonstrator's interaction with the environment. We characterize the system and validate its utility through downstream visuomotor policy learning of several contact-rich tasks.**

I. Introduction

Large-scale robot learning requires collecting demonstrations that capture both diverse environments and complex manipulation behaviors. Teleoperating physical robots can provide high-quality action data, but ties data collection to a particular robot and collection setup [1], [2]. Handheld manipulation interfaces provide an attractive alternative by allowing users to demonstrate tasks directly in the environment while retaining an action representation that can be transferred to robots. [3], [4], [5] The Universal Manipulation Interface (UMI) demonstrated this approach at scale, enabling policies for dynamic, precise, bimanual, and long-horizon manipulation from handheld demonstrations [3].

While UMI demonstrated the capability of this paradigm, sustained and contact-rich data collection places additional demands on both the physical interface and sensing system. The original UMI relies heavily on vision-derived state estimation: gripper width is tracked using fiducial markers, while end-effector pose is reconstructed through monocular visual-inertial SLAM [3]. Moreover, although soft fingers provide passive compliance and allow grasp force to be inferred indirectly through deformation, interaction forces with the environment are not directly measured [3]. Recent work has begun to address this limitation by incorporating tactile and force sensing into handheld demonstration devices for contact-rich manipulation [6], [4], [7]. However, force sensing addresses only one aspect of a practical data-collection system. Sustained use also demands mechanical robustness and operator ergonomics, while reliable collection requires consistent measurement of gripper configuration and motion.

We present **Eucalyptus**, a hardened UMI-style demonstration interface designed for sustained, force-aware manipulation data collection. Eucalyptus redesigns the gripper mechanism and user interface for repeated use and improved ergonomics, replaces vision-based finger-width estimation with direct encoder sensing, integrates two off-the-shelf 3-DoF force sensors at the fingers, and uses a dedicated Xvisio DS80 for 6-DoF pose tracking [8]. A lightweight wide-field-of-view camera preserves the wrist-centric visual observation used by UMI. Together, these changes provide direct multimodal measurements of visual, kinematic, and interaction state while preserving the portability of UMI-style data collection.

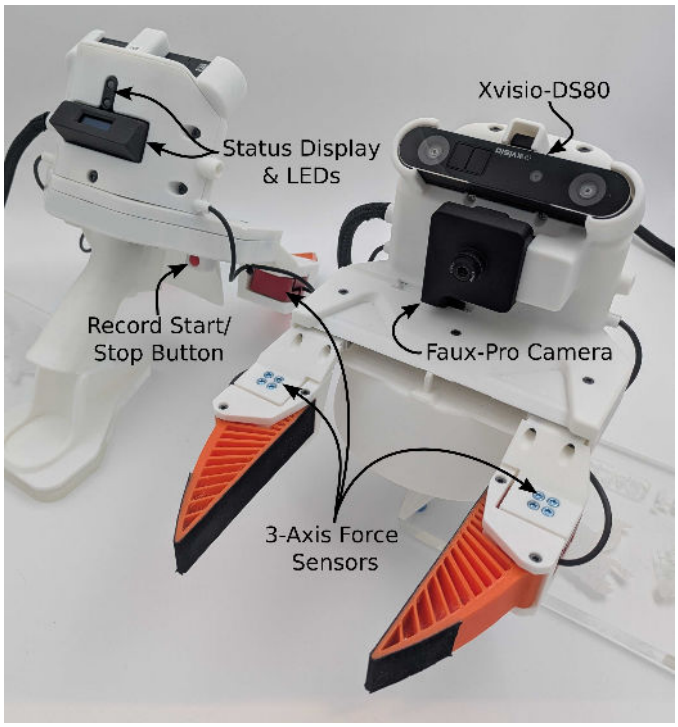


Fig. 1. Overview of the Eucalyptus demonstration interface and its primary sensing and mechanical components.

Our main contributions are:

- A mechanically redesigned UMI-style device with an ergonomics-oriented design intended for repeated, extended data collection.
- A multimodal sensing architecture combining wide-field-of-view vision, direct finger-position sensing, force sensing, and dedicated 6-DoF tracking using commercially available sensors.
- A system characterization and downstream policy evaluation demonstrating that Eucalyptus can collect manipulation data suitable for robot learning.

II. System

Eucalyptus combines a redesigned gripper mechanism, an improved operator interface, direct sensing of finger configuration, interaction forces, and 6-DoF motion. Fig. 1 provides an overview of the complete device.

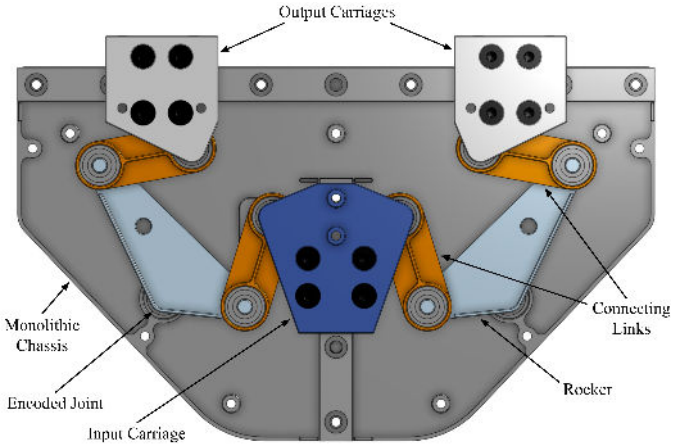


Fig. 2. Eucalyptus six-bar finger actuation mechanism. The linkage couples trigger motion to the gripper fingers while providing a directly measurable mechanical state.

A. Mechanical Design and Ergonomics

Eucalyptus replaces the original UMI finger actuation mechanism with a six-bar linkage that directly couples trigger motion to the two Fin Ray-style fingers. The machined aluminum mechanism was designed for robust, repeatable operation with reduced mechanical play and limited structural deflection under load. It provides a maximum finger stroke of 107 mm, compared with the 80 mm stroke reported for the original UMI gripper [3]. The linkage mechanism, detailed in Fig. 2, is highly repeatable with nearly zero backlash, enabling the gripper width to be accurately calculated from linkage joint angle measurements.

The grip and trigger of the device were fully redesigned to improve operator comfort and for hand sizes from the 10th percentile woman to 95th percentile as defined in [9]. The trigger's stroke was reduced to 30 mm and its height was reduced to support comfortable two-finger actuation. This trigger is fitted into two sizes of contoured pistol-style grip, with foam padding added between the operator's hand and the gripper body to improve comfort and reduce chafing.

B. Sensor System

The more robust and repeatable mechanical design of the Eucalyptus synergizes with an expanded sensing suite to record the gripper state and interaction forces. The sensing system consists of a lightweight wide-field-of-view RGB camera, a magnetic encoder for finger position, two 3-DoF force sensors, and a dedicated 6-DoF tracking camera. Together, these sensors reduce reliance on quantities inferred indirectly from RGB images.

Vision. Eucalyptus uses a lightweight custom RGB camera assembly, termed **Faux-Pro**, comprising an IDS GV-58C1LE Rev.2 board camera equipped with a Sunex DSL218A-670-F2.0 fisheye lens with a 180° maximum field of view [10]. The camera uses a high quantum efficiency 2.12 MP Sony IMX662 sensor with a resolution of 1936×1096 and supports frame rates up to 59.8 Hz [11]. The full camera assembly weighs 55 g, 70% lighter than the 181 g GoPro Hero9 with Max Lens used on UMI.

Faux-Pro was designed to closely match the visual observation geometry and camera placement of the original UMI setup [3]. Its lens and mounting geometry produce a closelymatched wrist-centric view of the gripper and workspace using the same camera mount, as shown in Fig. 4.

Finger state. Eucalyptus replaces the vision-based finger-width estimation used in the original UMI with direct measurement of the linkage configuration using an MA600 magnetic angle sensor [12] at the joint noted in Fig. 2. The sensor provides a configurable 12–15 bit absolute angle measurement, which is mapped to physical gripper width in meters through a per-device calibration, yielding sub-millimeter accuracy.

Force sensing. To capture interaction forces, Eucalyptus mounts a Forsentek F3P triaxial load cell at the base of each finger. These sensors effectively record a 5-axis wrench, where the moment about the axis connecting their two centers is indeterminate. Each sensor measures tension and compression along three orthogonal axes with $\pm 0.3\%$ full-scale nonlinearity, $\pm 0.3\%$ hysteresis, and $\pm 0.1\%$ non-repeatability [13]. Each axis is independently calibrated from raw voltage to force in Newtons using zero-load and known-load measurements.

6-DoF tracking. Eucalyptus uses a dedicated XVisio DS80 visual-inertial tracking camera for 6-DoF pose estimation [8]. The tracker is rigidly mounted facing forward and pitched approximately 45° upward, reducing occlusion of the tracking view by the manipulated object and the operator's hands. The DS80 provides the 6-DoF pose of its tracking frame, which is calibrated relative to the gripper tool-center-point (TCP).

C. Data Acquisition and User Interface

A custom electronics board provides the interface between the on-device sensors, user controls, and host computer. The board acquires data from the magnetic encoder and both 3-axis load cells, records user-control button states, supports calibration of the recorded sensor streams, and streams resultant data to the host computer over ethernet. It is also designed to support additional sensors and sensing modalities, with interfaces for microphones, tactile sensors, and CAN, SPI, or I^2C enabled sensors, simplifying future extensions of Eucalyptus to richer multimodal data collection.

As shown in Fig. 1, easily reachable start and stop buttons are duplicated on both sides of the device for convenient left-handed, right-handed, or bimanual operation. Integrated LEDs and a display provide immediate feedback on recording and system status.

III. Data Collection

Software and recording pipeline. The Eucalyptus collection software is implemented as a distributed ROS system. Each data source runs in an independent producer process and publishes its data on one or more ROS topics. These streams include compressed RGB images, gripper state, force measurements, user-input states, and DS80 pose estimates. A system-health process verifies that each required stream is available and that its publication rate remains within an expected range. RViz provides live visualization of the sensor streams, publication rates, and system status during collection,

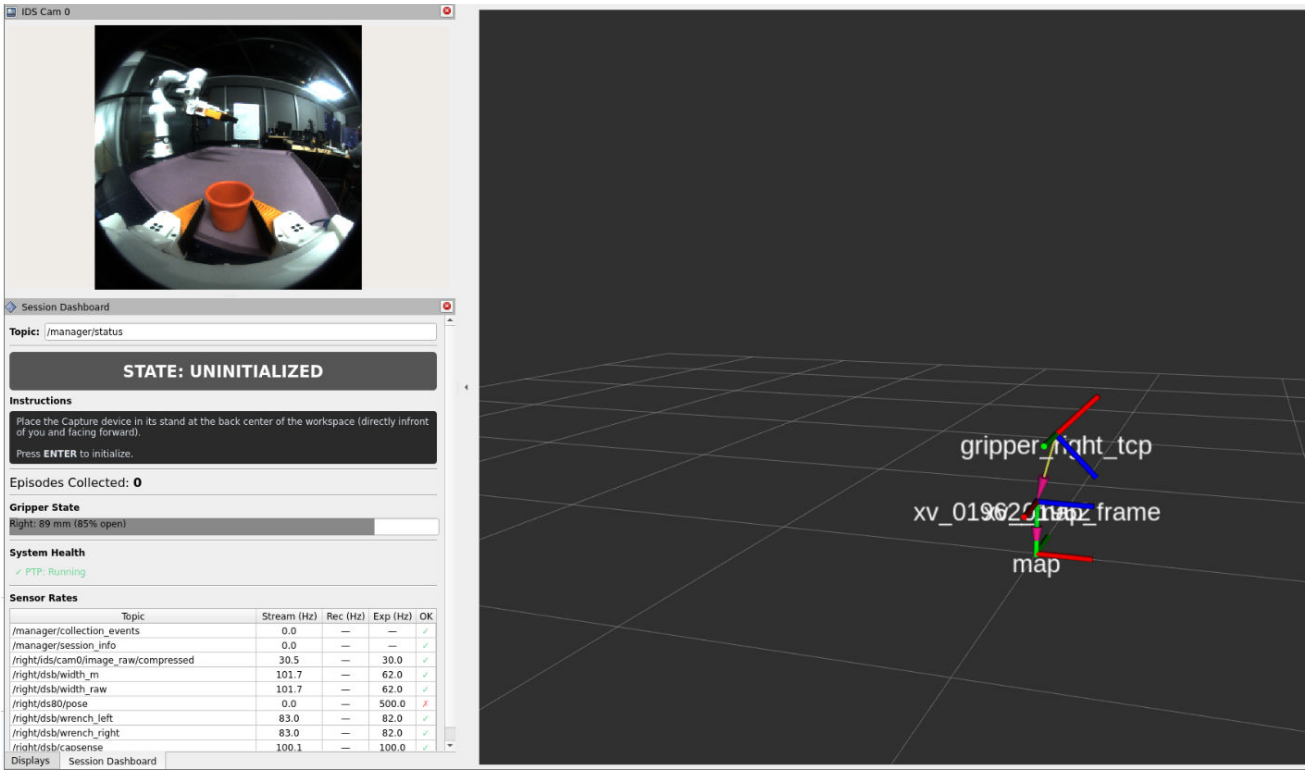


Fig. 3. RViz-based Eucalyptus data-collection interface showing live sensor visualization, acquisition rates, and recording status.

as shown in Fig. 3. A separate recording process subscribes to a configured set of topics and records them in MCAP format.

The raw DS80 tracking-frame pose is transformed using a calibrated tracker-to-TCP transform to obtain the gripper tool-center-point (TCP) pose. Eucalyptus records the TCP pose rather than the raw tracker pose. The same TCP convention is used to command the robot during policy execution, providing a common Cartesian pose representation between the handheld demonstration device and the robot.

Collection procedure. Before collecting demonstrations, the operator prepares the workspace without the movable task objects and places Eucalyptus in its stand at a designated central location on the table, providing a roughly consistent DS80 world-frame initialization across collection sessions. After launching the collection software, the operator initiates the DS80 mapping phase and moves the device throughout the workspace for approximately one minute to capture visual features and construct the SLAM map. Once sufficient coverage has been obtained, the operator ends the mapping phase and the system is ready to collect demonstrations.

The task objects are then placed in the workspace according to the task-specific randomization scheme. The operator starts recording using the controls on Eucalyptus, performs the demonstration, and stops recording once the behavior is complete. The scene is then reset and randomized before the next demonstration. This process is repeated for the duration of the collection session.

After collection, recorded demonstrations are visualized using inspection tools and filtered to remove invalid or low-quality episodes. The retained demonstrations are then post-processed and uploaded to persistent storage for downstream policy training.

IV. Results

A. System Characterization

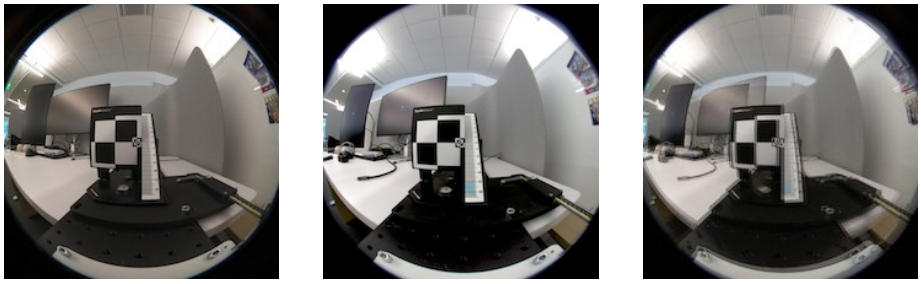
Mechanical characteristics. Measurements on a stock UMI gripper showed position-dependent gripper-width backlash of 7–12 mm over the stroke. In contrast, no measurable backlash was observed in the Eucalyptus six-bar linkage. Eucalyptus weighs 975 g, compared with 780 g reported for the original UMI gripper [3], a 195 g increase associated with the more rigid actuation mechanism, dedicated pose tracking, and finger-level force sensing.

Visual compatibility. To directly compare Faux-Pro with the GoPro configuration used by UMI, the two camera systems were mounted in an identical fixture to reproduce the same viewpoint. Fig. 4 shows observations from both cameras together with a 50/50 alpha blend. The comparison shows that Faux-Pro preserves similar wide-angle, wrist-centric visual coverage despite using different camera and lens hardware.

Pose tracking. The Eucalyptus tracking subsystem was previously characterized against a Vicon motion-capture system, with mean relative translation and rotation errors of 1.1 mm and 0.2°, respectively [14]. For context, UMI reports 6.1 mm/3.5° ATE and 10.1 mm/0.8° RPE under a different evaluation protocol [3].

B. Policy Execution

Compatibility with existing UMI policies. As an end-to-end test of visual compatibility, we deployed the Cup-in-the-Wild UMI policy using images from Faux-Pro without camera-



(a) UMI GoPro (b) Faux-Pro (c) 50/50 overlay

Fig. 4. Visual comparison between the GoPro configuration used by UMI and Faux-Pro. Both cameras were mounted in an identical fixture to reproduce the same viewpoint. The 50/50 alpha blend highlights differences in image geometry and field of view between the two camera systems.

specific retraining or adaptation. The policy successfully executed the task in our initial trials, suggesting that the Faux-Pro observation is sufficiently close to the original UMI camera configuration to support zero-shot policy transfer for this task.

Collection scale and downstream policy use. During Summer 2026, five collectors using four Eucalyptus devices collected more than 18,000 demonstrations across three manipulation tasks. Peak single-collector throughput exceeded 700 episodes per day, providing practical evidence of sustained collection at scale.

In a separate effort, demonstrations collected with Eucalyptus were used to train visuomotor policies for contact-rich robot manipulation, including pulley routing, pipe insertion, and battery insertion, as previously reported in [14].

V. Discussion and Limitations

The richer sensing stack comes with additional operational considerations. Unlike the original UMI, which records primarily through an onboard GoPro, Eucalyptus is tethered to a host computer during collection and requires the collection software to be running. Sensor streams are timestamped by different devices and therefore originate from different clock domains. Their timestamps are first normalized to a common clock domain, after which the streams can be temporally aligned and resampled for downstream policy training.

Several limitations remain. The DS80 provides accurate relative pose estimates in our prior characterization [14], but remains a visual-inertial SLAM system and may degrade under unfavorable visual conditions. The finger-mounted load cells provide 3D interaction forces but do not measure the location or spatial distribution of contact on the compliant fingers. The benefit of force observations for downstream policy performance also remains to be systematically evaluated.

Future work will address this missing contact information by integrating custom low-cost full-finger capacitive tactile sensing into the fingers. The Eucalyptus acquisition electronics were designed to support additional tactile and audio sensing, enabling future versions of the platform to capture richer multimodal demonstrations.

VI. Conclusion

We presented Eucalyptus, a hardened UMI-style interface for sustained, force-aware manipulation data collection. Eucalyptus combines a redesigned mechanical interface with direct gripper-state and force sensing, dedicated 6-DoF tracking, and a wide-angle RGB camera designed for visual compatibility with existing UMI systems. We described the hardware and software pipeline used to collect synchronized multimodal demonstrations and validated its use for downstream robot policy learning. Eucalyptus provides a practical foundation for richer, robot-independent manipulation data collection, with future extensions toward tactile and audio sensing.

References

[1] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," *Robotics: Science and Systems*, 2023.

[2] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, "Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators," *IEEE International Conference on Intelligent Robots and Systems*, pp. 12 156–12 163, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10801581

[3] C. Chi, Z.-J. Xu, C.-E. Pan, E. Cousineau, B. Burchfiel, S.-Y. Feng, R. Tedrake, and S. Song, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," *Robotics: Science and Systems Conference*, 2024.

[4] H. Choi, Y. Hou, C. Pan, S. Hong, A. Patel, X. Xu, M. R. Cutkosky, and S. Song, "In-the-wild compliant manipulation with umi-ft," 1 2026. [Online]. Available: https://arxiv.org/pdf/2601.09988

[5] K. Liu, C. Guan, Z. Jia, Z. Wu, X. Liu, T. Wang, S. Liang, P. Chen, P. Zhang, H. Song, D. Qu, D. Wang, Z. Wang, N. Cao, Y. Ding, B. Zhao, X. Li, S. A. Lab, and S. Jiao, "Fastumi: A scalable and hardware-independent universal manipulation interface with dataset," *Conference on Robot Learning*, pp. 3069–3093, 2025. [Online]. Available: https://arxiv.org/pdf/2409.19499

[6] W. Liu, J. Wang, Y. Wang, W. Wang, and C. Lu, "Forcemimic: Force-centric imitation learning with force-motion capture system for contact-rich manipulation," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 1105–1112, 10 2024. [Online]. Available: https://arxiv.org/pdf/2410.07554

[7] T. Cheng, K. Chen, L. Chen, L. Zhang, Y. Zhang, Y. Ling, M. Hamad, Z. Bing, F. Wu, K. Sharma, and A. Knoll, "Tacumi: A multi-modal universal manipulation interface for contact-rich tasks," 1 2026. [Online]. Available: https://arxiv.org/pdf/2601.14550

[8] "Seersense® ds80 module - xvisio technology." [Online]. Available: https://en.xvisiotech.com/SeerSense/10

[9] S. Rajulu and G. Klute, "Anthropometric survey of the astronaut applicants and astronauts from 1985 to 1991," 1993.

[10] "Dsl218 specification sheet excellence in digital imaging optics." [Online]. Available: https://www.optics-online.com/OOL/DSL/DSL218.PDF

[11] "Gige ueye le rev.2 — single board camera — imx662 — 2 mp sony sensor — webstore." [Online]. Available: https://en.ids-imaging.com/store/gv-58c1le-rev-2.html

[12] "Ma600 — sensors >>angular position sensors." [Online]. Available: https://www.monolithicpower.com/en/ma600.html

[13] "3-axis force sensor 50n 100n 200n 3d load cell." [Online]. Available: http://www.forsentek.com/prodetail_503.html

[14] V. Surendran, N. Peri, and D. Watkins, "Bridging handheld and teleoperated supervision for contact-rich manipulation via state-gated experts," 6 2026. [Online]. Available: https://arxiv.org/pdf/2606.26603