

VisTaR: Visuo-Tactile In-Hand Object Pose Refinement for a Multi-Finger Robot Hand

Supervisor: Rohit Menon

Why this project: it combines RGB-D, tactile sensing, hand proprioception, and Vicon reference data in one robot-hand experiment.

Learning outcome: calibrate and synchronize a multi-modal robot-hand setup, implement $SE(3)$ pose refinement, and evaluate it against Vicon reference poses.

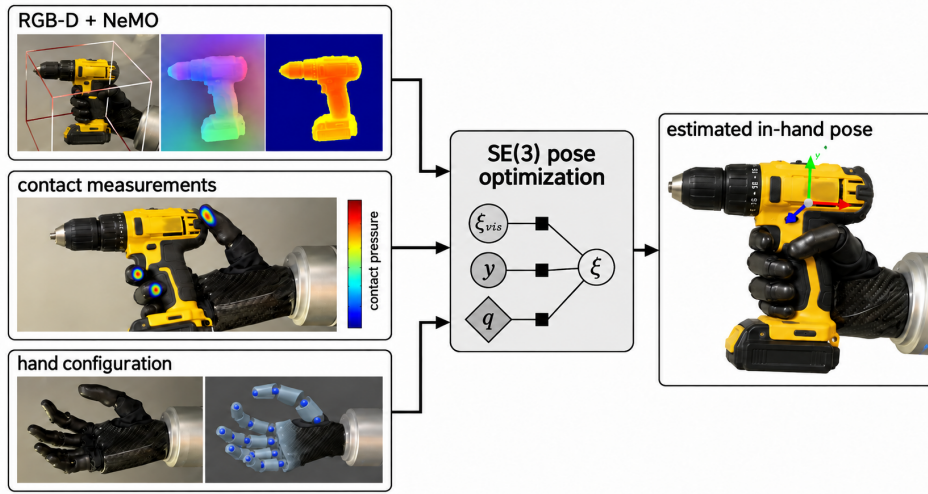


Figure 1: Visuo-tactile pose refinement. NeMO [2] provides the RGB pose hypothesis; depth, contacts, and hand configuration refine it [1].

Motivation

In-hand pose is required for regrasping, insertion, and placement [1]. During closure, fingers occlude the object while tactile measurements become available.

Problem

Refine an RGB pose hypothesis of a grasped object using registered depth, hand joint positions, and fingertip tactile values.

Goal

Develop and evaluate a pose-refinement method on a PSYONIC Ability Hand mounted on a robot arm [4]. NeMO [2] supplies an RGB pose prior; registered depth, hand configuration, and tactile contacts refine it during grasp closure.

Task Setting

The robot closes from a fixed preshape. An RGB-D camera, hand joint positions, and 30 tactile channels are recorded synchronously; Vicon supplies the reference pose [4, 5].

Objects

The YCB validation objects are the yellow mustard bottle, sugar box, and mug/cup [3]. At every outcome level, transfer to two supervisor-supplied test objects without task-specific retraining.

Work Packages

Student B's synchronized setup is the critical path. It must complete calibration and one synchronized trial by mid-December, leaving a two-week buffer before January integration.

Student	Owns	Also contributes
A	Camera calibration, NeMO templates, RGB-D pose prior, and depth-consistency formulation.	Synthetic-noise stress test.
B	Hand/arm interface, tactile and joint logging, Vicon, synchronization and timing-error analysis, and grasp protocol.	Shared final arm trials and evaluation.
C	$SE(3)$ refinement estimator (depth $\rightarrow$ proprioception $\rightarrow$ tactile) and error evaluation.	Shared final arm trials and evaluation.
All	Final integration, transfer to two test objects, final report, and demonstration.	

Project Schedule

Period	Planned Milestone
November	Hardware setup, frame calibration, and recording interface.
December	Implement the three work packages.
January	Integrate and evaluate fixed-hand trials.
Late January	Midterm: calibrated setup and an offline end-to-end refinement on one YCB object.
February	Validate on the YCB objects and transfer to test objects.
March	Robot-arm demonstration and final report.

Outcome Levels

At every outcome level, the stated capability must be transferred and evaluated on the two supervisor-supplied test objects.

Minimum Outcome

- Synchronized RGB-D, hand, tactile, and Vicon recording.
- RGB-D refinement on the three YCB objects, with translation and rotation error against Vicon.

Expected Outcome

- Integrated refinement with hand configuration and tactile contacts.
- Error analysis before contact, after contact, and under hand occlusion.

Desired Outcome

- Pick-and-place using the refined in-hand pose under substantial hand occlusion.
- A stress test with synthetic RGB-D noise and severe occlusion, reporting initial and refined pose errors.

Risk and Contingency

The main risk is hardware and synchronization delays in Student B's work. The November–December plan retains a two-week buffer; NeMO integration can proceed in parallel on recorded data.

Prerequisites

Python, basic ROS2, linear algebra ($SE(3)$ and least squares), and willingness to work with hardware. PyTorch experience is helpful but not required.

Deliverables

A working refinement pipeline, synchronized recordings, a robot-arm demonstration, an evaluation report with pose-error tables, and a final presentation.

References

- [1] L. Mack, F. Grüninger, B. A. Richardson, R. Lendway, K. J. Kuchenbecker, and J. Stueckler, “Visuo-Tactile Object Pose Estimation for a Multi-Finger Robot Hand with Low-Resolution In-Hand Tactile Sensing,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025.
- [2] S. Jung, L. Klüpfel, R. Triebel, and M. Durner, “Finding NeMO: A Geometry-Aware Representation of Template Views for Few-Shot Perception,” in *Proceedings of the International Conference on 3D Vision (3DV)*, 2026, pp. 342–352. [Online]. Available: <https://github.com/DLR-RM/nemo>
- [3] B. Calli, S. Srinivasa, A. Singh, P. Abbeel, A. Walsman, and A. M. Dollar, “The YCB Object and Model Set: Towards Common Benchmarks for Manipulation Research,” in *Proceedings of the International Conference on Advanced Robotics (ICAR)*, 2015.
- [4] PSYONIC, “Ability Hand for Research Users,” online resource, 2026.
- [5] Vicon Motion Systems, “Vicon Help: Motion Capture Documentation,” online resource, 2026.