Individual Lab Report 09

Manyung Emma Hon

October 28th, 2025

Team H: MAGIC

Team Members:
Kartik Agrawal, Shreya Ragi, Megan Lee, Kailash Jagadeesh

Table of Contents

Individual Progress

This cycle focused on fully integrating perception-driven pose estimation with manipulation execution, validating it in hardware, and closing the loop between planning, control, and scene understanding. The main accomplishment was achieving reliable, repeatable grasping using perception outputs for object localization and orientation (yaw), without relying on manually specified grasp poses. This represents a major step toward autonomous dual-arm manipulation driven by vision.

Building on Kartik's improved pose estimation, we sync up on arm orientations and yaw rotation. I then incorporated yaw estimation into the manipulation pipeline. The system now correctly spawns a full pose for the target object consisting of x, y, and yaw in the world frame. That yaw information is propagated through the finite state machine (FSM) and used to generate grasp approaches aligned to the object's real-world rotation on the table. The yaw-aware grasping logic was implemented both in software and tested in hardware. We verified that the arm can select a grasp aligned with the detected handle orientation within the robot's reachable yaw band, as seen in Figure 1 for simulation. After integration, we achieved successful grasping in hardware as well, with no damage to the object, confirming that the grasp approach and closure are gentle and well aligned.

One important physical constraint we formally captured this cycle is the robot's yaw reachability envelope. Through testing, we determined that the arm can only stably grasp objects whose yaw is approximately aligned within a 45-degree nominal band, with a tolerance of about ±5 degrees before joint limits, self-collision risk, or loss of manipulability become significant. The FSM itself is now synchronized end-to-end with pose estimation. Each state has explicit entry conditions that include requirements for valid transforms, recent pose estimates, and reachable yaw. The FSM also maintains consistent exit criteria, so we always know whether we transitioned because of success, recovery, or abort. That consistency matters for debugging and for safety. We validated this by running complete cycles where perception publishes x/y/yaw, the FSM evaluates feasibility, the planner produces a Cartesian trajectory with waypoints, and the arm executes the grasp. The result: repeatable pickup behavior with no collisions and no object damage.

We also synced both arms with the same pose estimation information. The shared world model now ensures that both manipulators see the same x, y, and yaw for the object. By doing this, we reduced the risk of one arm "believing" a different state of the world than the other. This synchronized view is a critical step toward coordinated bimanual behaviors, because it establishes a single geometric truth source instead of letting each arm guess independently.

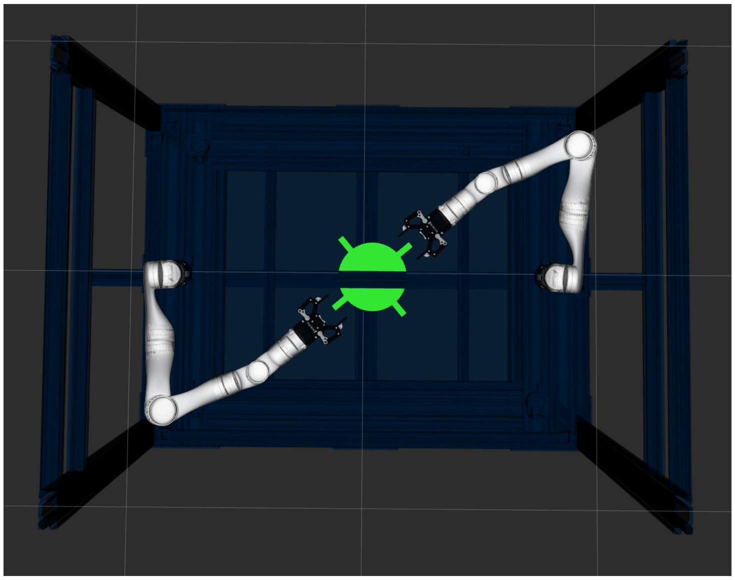

Figure 1: object spawning at yaw = 40, x = 0.03, y = 0 with arms successfully reaching the grasp point

Challenges

Even though localization is now much more accurate, there are still failure modes. The pose estimation currently assumes that calibration between the camera and the world/robot base is correct. While Kartik's distortion and tilt calibration got us much closer, we still observe small errors in some test poses. These are now low enough for reliable grasping in nominal setups, but they will compound if the object is at the edge of reach or rotated near the edge of our allowable yaw band.

Another limitation is the yaw reachability itself. We confirmed that while we can compute yaw for any detected object, we cannot always act on arbitrary yaw values. The arm physically cannot approach certain rotations cleanly, especially those that force uncomfortable wrist orientations. This is an inherent kinematic constraint, not a software bug. We have mitigations in software by gating execution on the 45° ± 5° feasible envelope, but that still means there are real-world configurations where we can see the object, know exactly where it is and how it's oriented, and still be unable to grasp it with the current hardware and trajectories.

A further ongoing challenge is full-scene 3D understanding. We would like to build robust object reconstructions, not just 2D pose plus yaw. That's where the reconstruction pipeline comes in and where we still have open problems. The object we are manipulating is mostly black and white, and those same tones are present in the workspace. That low texture and low color diversity make stereo depth estimation and segmentation-based mask extraction (including automatic segmentation methods like sam2) unreliable. Large sections of the object either blend into the table or wash out. As a result, the current reconstruction is incomplete, and the sides that are not directly visible remain either noisy or missing. This prevents us from closing the loop on full 6-DoF grasp planning and also limits any downstream reasoning that depends on full geometry.

Teamwork

Kartik Agrawal

He addressed core perception quality issues (distortion, tilt) and produced stable x, y, and yaw outputs. He then integrated those into the manipulation subsystem so that pose estimation isn't running in isolation; it is part of the live control loop.

Shreya Ragi and Kailash

They drove the reconstruction and transform infrastructure. Kailash completed camera to world calibration for the reconstruction pipeline and worked on integrating transforms from both arms into that same world frame. Shreya used that transform chain to attempt multi-view reconstruction of the object. The pipeline currently takes the nearest transform in time for each frame to align captures into a consistent scene. This temporal association step is important because it means each camera frame can be mapped into the shared reconstruction frame without drifting too far in time from the "ground truth" arm/world transform. Using this setup, we achieved an initial 3D reconstruction of the object.

Megan Lee

She continued improving planning reliability. Her Cartesian planning with waypoints and interpolation is what allows us to execute smooth approach paths instead of jumping straight to the grasp pose. This is important for protecting the object and avoiding abrupt wrist rotations.

Collaboration

Kartik, Megan, and I worked together to complete the manipulation and pose estimation integration. Shreya and Kailash worked together to solve the transformation.

Future Plan

By the next progress review, we plan to have full integration ready with a fall validation demo dry run. I will focus on adding steps on FSM to improve object reconstruction and streamline manipulation for the stretch object.