Vision Challenge Report

(ROB 330)

Alexandru Otlacan (aotlacan) and Akash Sarada (asarada) Tian Yi Yang (tianyiy)

Overview

This report documents our implementation and results for the Vision Challenge. The challenge consisted of three technical components: camera calibration, AprilTag detection, and color detection. The report concludes with a reflection on our performance and potential improvements.

Task 1: Camera Calibration in ROS

Camera Intrinsic Parameters

During our implementation, the robot experienced difficulties successfully completing a full camera calibration using live image data. As a result, for the purposes of this challenge, we used the camera calibration parameters provided by the lab staff. These parameters were distributed as a compressed `.tar.gz` file, which we unzipped and integrated directly into our ROS2 workspace. Using the provided calibration ensured that our vision pipeline remained functional and consistent for AprilTag detection and color estimation.

Although we relied on the provided calibration, we outline below the correct procedure for performing monocular camera calibration in ROS2, which we would follow if recalibrating the robot from scratch.

Camera calibration in ROS2 is typically performed using the monocular calibration tools with a planar checkerboard target. The process involves collecting a large set of images of the checkerboard at varying distances, orientations, and positions within the camera frame. These images are processed to estimate the intrinsic parameters of the pinhole camera model, which describe how 3D points in the camera coordinate frame are projected onto the 2D image plane.

The primary intrinsic parameters estimated during calibration are:

- **Focal lengths** (f_x, f_y): These parameters scale points along the horizontal and vertical image axes and are related to the camera focal length and the physical pixel size of the image sensor.

- **Principal point** (c_x, c_y): The pixel coordinates corresponding to the optical center of the camera, which is ideally near the center of the image.
- **Distortion coefficients**: Radial distortion coefficients (k_1, k_2, k_3) account for barrel or pincushion distortion introduced by the lens, while tangential distortion coefficients (p_1, p_2) model lens misalignment relative to the image plane.

Together, these parameters define how real-world geometry is mapped into image coordinates and are critical for accurate pose estimation, projection, and metric reasoning in vision-based robotics.

Reprojection Error

Reprojection error is a quantitative measure of calibration quality. It represents the average pixel distance between observed checkerboard corner locations in the image and the projected locations of the corresponding 3D points using the estimated camera model. A lower reprojection error indicates that the estimated parameters accurately model the camera.

Our calibration resulted in a low average reprojection error, indicating a good calibration. In practice, reprojection errors on the order of a fraction of a pixel to a few pixels are generally considered acceptable for mobile robotics applications.

Checkerboard Detections and Calibration Progress

Figures 1 and 2 show checkerboard detections at two different timesteps during calibration. During calibration, we looked for consistent and accurate corner detections across a wide range of checkerboard poses, including varying distances, orientations, and positions within the image frame. A good calibration is one that is able to correctly identify sharp edges and corners, while having multiple different viewpoints of the full calibration checkerboard.

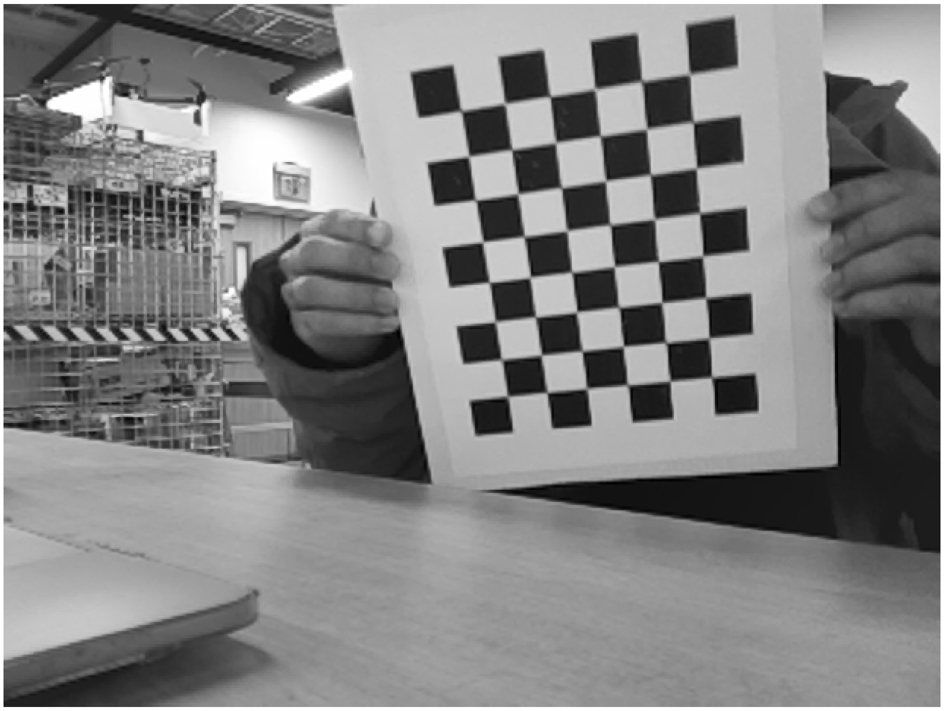

Figure 1: Checkerboard detection during camera calibration at timestep 1.

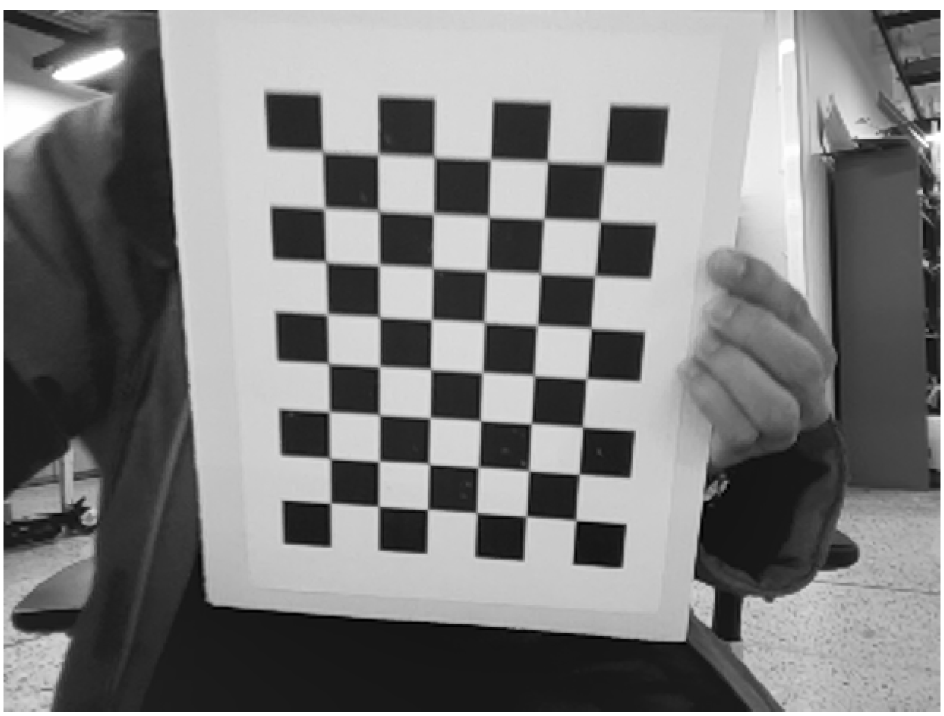

Figure 2: Checkerboard detection during camera calibration at timestep 2.

Task 2: AprilTag Detection and Localization

Advantages and Disadvantages of Fiducial Markers

Fiducial markers such as AprilTags provide a robust and computationally efficient method for object detection and pose estimation.

Some advantages of fiducial markers include their reliability and ease of estimation. Fiducial markers do not change based on lighting conditions or viewing angle, allowing the robot to know exactly where in the map it is when it discovers the marker. Furthermore, since it does not move, when the robot detects the marker, it can fix any errors as it knows its true location. When viewing the marker, the robot can easily estimate its location on the map by triangulating its viewpoints with the marker's location.

However, some limitations do exist, namely camera limitations and the need to be placed in a specific location and maintained. These fiducial markers must be placed by a human, and must be maintained, allowing the robot to easily view the markers. However, these markers don't easily blend into the environment, making it unsuitable for certain conditions. Furthermore, since the robot relies on vision to detect the markers, camera limitations like distortion, extreme viewing angles, or long distances degrade the detection performance of the robot.

Fiducial markers are most appropriate in structured environments, calibration tasks, and controlled laboratory or warehouse settings, where robots and humans work together to keep operations alive.

Experimental Observations

In our lab experiments, AprilTag detection worked well when the tag was well-lit, approximately facing the camera, and within a reasonable distance. Detection accuracy was highest when the tag occupied a sufficient number of pixels in the image.

Detection failures occurred when the robot was at a relatively extreme angle to the tag, or when motion blur was present. Performance also degraded when the tag appeared very small in the image due to distance.

AprilTag Detection Example

Figure 3 shows an accurate AprilTag detection, including the detected tag ID and its estimated location relative to the camera frame.

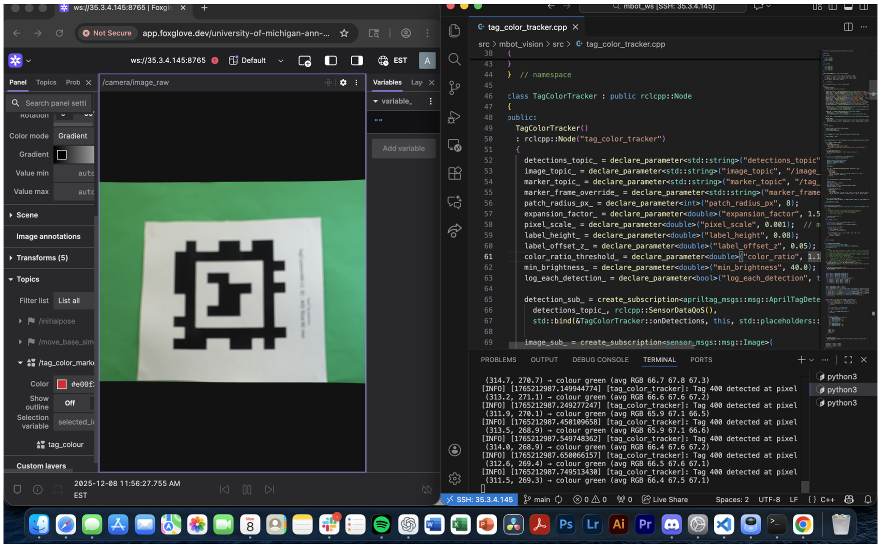

Figure 3: Accurate AprilTag detection showing tag ID and pose estimation.

Use of AprilTag Pose for Navigation

The pose returned by AprilTag detection cannot always be used directly for robot navigation. Measurement noise, camera calibration error, and latency can lead to inaccurate pose estimates. Additionally, the pose is reported in the camera frame and must be transformed into the robot base or map frame. For navigation, filtering, frame transformations, and consideration of robot kinematics and safety constraints are required.

Task 3: Color Detection

Method and Parameters

Our color detection method sampled pixel values in a local image region associated with a detected AprilTag. Average RGB values were computed and compared against predefined thresholds to classify the observed color. Parameters included minimum brightness thresholds and channel ratio thresholds, which were tuned empirically by observing color readings under different lighting conditions.

Limitation

A key limitation of this approach is sensitivity to lighting variation. Changes in ambient lighting or shadows can significantly affect RGB values, potentially leading to incorrect color classification. More robust approaches could involve HSV color space thresholding or adaptive color models.

Reflection

Overall, our implementation successfully met the objectives of the Vision Challenge. Camera calibration and AprilTag detection were reliable and integrated effectively into the system. If we were to repeat the challenge, we would spend more time quantitatively tuning color detection thresholds and exploring illumination-invariant color representations. Additionally, incorporating temporal filtering for AprilTag pose estimates would further improve robustness for navigation tasks.