Camera-based Localization of an Unmanned Aerial Vehicles

Khushi Katara
B21ME032
katara.1@iitj.ac.in

Rishika Bera
B21ME052
bera.4@iitj.ac.in

Tamanna Maheshwari
B21ME068
maheshwari.12@iitj.ac.in

***Abstract*—This proposal addresses the need for precise unmanned aerial vehicle (UAV) localization in scenarios where Global Navigation Satellite System (GNSS) signals are degraded. We introduce a Camera Localization system for UAVs, focusing on 2-D image space measurements and overcoming challenges associated with traditional approaches. The project integrates YOLO for object detection, Deep SORT for tracking, and monocular SLAM with novel bundle adjustment processes incorporating GPS and Digital Elevation Model (DEM) data. This aims to enhance UAV navigation precision and reliability in diverse environments, overcoming limitations of both relative and absolute visual localization methods. Contributions include implementing YOLOv5, integrating Deep SORT, and optimizing algorithms for UAV systems, advancing the field of autonomous UAV navigation.**

***Index Terms*—Keyword- SLAM, GPS, GIS, Deep SORT, YOLOv5, AVL, RVL, VO**

I. Introduction

To navigate autonomously in scenarios like military and defense, post-disaster assessment, infrastructure inspection, precision agriculture, flight testing, and so on, UAVs depend highly on the Global Navigation Satellite System (GNSS), such as GPS and Beidou, to acquire their locations. However, signals from GNSS satellites could be degraded when it is close to vegetation, water bodies, hostile environments, and inside structures. This Camera Localization of Unmanned Aerial Vehicle situations would make it impossible to get precise latitude and longitude estimates of the UAV during flight. An auxiliary or redundant localization approach is needed to ensure the UAV navigation when it fails to receive the GNSS signals. Primarily, two approaches are employed for UAV localization: relative visual localization (RVL) and absolute visual localization (AVL). While RVL involves frame-by-frame localization, AVL references a fixed frame for localization. However, achieving accurate localization presents challenges, particularly in RVL systems utilizing Visual Odometry (VO) and Simultaneous Localization and Mapping (SLAM). This results in error accumulation and localization drift over time. This drift poses significant hurdles in attaining accurate geo-localization.

II. Literature Review

GPS localization provides only a 3D position, SLAM has the advantage of providing 6 degrees of freedom (DoF) localization at a higher frequency and a 3D point cloud describing the surrounding environment. Nevertheless, the resulting localization is not geo-referenced and suffers from accumulation error [3]. Landmarks are conspicuous objects which can mark locality. Landmark-based localization has already been applied in UAV applications. The challenge is that the matching might fail due to the changing image-capturing conditions, such as seasonal and perspective changes. Even though deep learning tools can make generalizations for the changed features, training the CNN to remember all the reference maps accurately is impractical due to the limited GPU memory size. In addition, it is also worth noting that UAV needs to fly at a favorable altitude to acquire images with a resolution comparable to the reference map, which might disable UAVs from conducting AVL at a lower altitude [1]. Though popularly used methods rely on calculating the pose of the drone after detecting the landmark and then transforming the pose from 3-D camera coordinates into 3-D world coordinates, it might lead to unnecessary errors if the camera is not calibrated properly. Some of these steps can be skipped with the help of visual servoing. The control needs to be modified to get the 2-D image coordinates as input; this will make the process more robust to tiny calibration errors, which will be significant while transforming into world coordinates [2].

III. Summary Of Work Done

A. Part 1

1) Object tracking with DeepSORT: In this, we implemented DeepSORT object tracking algorithm to detect objects in moving frames. DeepSORT is an advanced version of SORT(Simple Online and Realtime Tracking). With DeepSORT, this algorithm can work well in case of occlusion of objects. We have cloned the deepSORT repository from GitHub directly, which contains all the modules (for example Kalman Filter, pre-trained model, IoU matching, linear assignment, nn matching, etc), that were needed for implementation. Then a Python code is written which implements a tracking system using the DeepSORT algorithm. It utilizes pre-trained models and feature extraction techniques to track objects in video frames, updating the tracks based on new detections. Following this, one more Python file is generated which contains the code that combines object detection, which is done by YOLOv8, and object tracking techniques to annotate and track objects in input video, producing an output video file with annotated object tracks.

B. Part 2

1) SLAM-based localization integrated it with bundle adjustment processes that incorporate GPS measurements and 'Digital Elevation Model' (DEM) data: We have implemented the code of SLAM integrated with bundle adjustment in MATLAB. We first did the **Map initialisation** in order to find out the camera pose and 3-D points for an initial reference and for that we used **triangulation and homography techniques**. After that, we did **tracking** and in this we iterated through the frames until a key frame is found and then using that key frame the camera pose was determined using **Perspective-n-Point algorithm**. Whenever a key frame is found we applied the **triangulation algorithm** to maintain and update the 3-D map of the environment and it was also used in **similarity pose optimisation**(used to correct the drift in the camera poses.) which comes after the loop closure detection for which the similarity between the two key frames was required to be determined. Finally, the optimised camera trajectory was was compared with the ground truth to evaluate the accuracy of the ORB SLAM. The average Root Mean Square Error of this algorithm was **0.14917**.

C. Part 3

1) Simulation: This project features integration of various technologies within a Gazebo simulation environment. The ArduPilot package, along with MAVROS, SITL, GPS-enhanced SLAM, and DeepSORT based object detection systems, have all been successfully installed and configured to orchestrate an autonomous drone system. The objective is for the drone to autonomously take off from a designated marker, perform area mapping, and then return precisely to its starting point. The necessary software installations are complete, and the code for initiating drone takeoff has been developed and implemented. The system is fully prepared for operational testing, with all components theoretically ready to function as intended. Despite this readiness and the successful setup of all components, the drone has not been able to take off as expected. This unexpected issue has currently stalled further progress and testing of the drone's autonomous mapping and return capabilities.

IV. Results

V. Contributions

1) Khushi Katara: I integrated the DeepSORT object tracking algorithm into the project, enhancing its capability to detect objects in dynamic environments, particularly adept at handling occlusion scenarios. I developed Python scripts to implement a robust tracking system, seamlessly combining object detection using YOLOv8, resulting in annotated object tracks in the output video.

2) Rishika Bera: Will orchestrate the establishment of the Gazebo simulation environment for UAVs, ensuring fidelity and realism in simulated scenarios, integrate ROS with Gazebo, enabling comprehensive communication and control functionalities for simulated UAVs, develop sophisticated UAV models and sensor configurations within Gazebo, facilitating accurate simulation-based testing of algorithms.

3) Tamanna Maheshwari: I have implemented the monocular visual SLAM algorithm for UAV localization and to optimize it, I have integrated it with local Bundle adjustment, loop closure detection and map optimization. Hence, I was able to acheive the 6 DoF localisation of UAV. However, I faced some challenges in using GPS data as ground truth in order to do the global optimisation of the UAV path.

[1]–[3]

References

[1] Mostafa Hegazy, Riby Abraham Boby, and Alexandr Klimchik. Localization of a drone for landing using image-based visual servoing with image moments. pages 129–136, 2024.

[2] Dorra Larnaout, Vincent Gay-Bellile, Steve Bourgeois, and Michel Dhome. Vehicle 6-dof localization based on slam constrained by gps and digital elevation model information. In *2013 IEEE International Conference on Image Processing*, pages 2504–2508, 2013.

[3] Meng D. Zhao S. An B. (2023) Ma, L. Visual localization with a monocular camera for unmanned aerial vehicle based on landmark detection and tracking using yolov5 and deepsort.. *International Journal of Advanced Robotic Systems.*