State Estimation: Investigating the Issues During Indoor-Outdoor Transitions

Damla Leblebicioglu*, Francis Jacob Kalliath*, Arun Madhusudhanan*,
Shankara Narayanan Vaidyanathan*, Tejaswini Dilip Deore*
Team D-FAST

***Abstract*— Understanding the state of the robot is a crucial element for autonomous navigation. Several papers investigate SLAM (Simultaneous Localization and Mapping) pipelines in either outdoor or interior contexts, or both. However, only a few works have looked at the difficulties that might develop when moving from an outdoor to an indoor setting or vice versa. In this project, we have explored some frequent problems faced when performing global state estimation during such environment transitions. This project used all three key sensors covered in the Robotics Sensing and Navigation course: Camera, IMU, and RTK-GPS.**

I. Introduction

We expect robots to have the ability to make decisions and act on their own, and this fictitious (no more!) fantasy has never been more closer to reality than in the previous decade. Autonomous navigation can be understood in a general sense using the five-step pathway, also known as the Sense-Think-Act (SPA) paradigm [1], which is:

- Vehicle interacts with the physical world.
- Sense - Collects data using multiple sensors.
- Perceive - Interprets this data to build the map of the environment, understands where the obstacles are, and determines its location with respect to landmarks.
- Plan - Plans the optimal path from the current location to the desired point.
- Act - Acts according to its decision and hence causes an interaction.

According to this loop, the robot needs to determine its state, which is a set of quantities such as position, orientation, and velocity, that, if known, fully describe that robot's motion over time.

We can use sensors like Cameras, IMUs, GPS, and LiDARs (Light Detection and Ranging) to understand the state of the robot. However, using these sensors separately makes it difficult to achieve robust state estimation in long-term navigation. Because many of these sensors have complementary qualities, sensor fusion has been the go-to strategy in recent years because it allows us to obtain improved accuracy despite the fact that some of the sensors have scenario-specific drawbacks (like the multipath error in GPS when moving in an urban canyon scenario).

Visual-Inertial Navigation algorithms, which combine the measurements provided by cameras and IMUs, have been shown to have high accuracy and robustness. The most popular ones in the last 3 years have been ORB-SLAM 3 [2], BASALT [3], VINS-Fusion [4], and Kimera [5]. Although the visual place recognition modules perform loop closure in some of the Visual Inertial Odometry pipelines, they still have the issue of significant drift in long-term trajectories. GPS has been known to provide globally drift-free localization. Despite it being accompanied by noise, there have been attempts to fuse the GPS measurements with the VINS algorithm estimates to achieve a drift-free and globally aware solution. We have had multiple methods display their prowess in outdoor settings. However, relatively few approaches [6], [7], [8], have attempted to address the issues that arise when there is a brief GPS loss in complicated contexts.

In this study, our aim has been to understand the challenges usually faced when using a sensor suite with a camera, IMU and GPS in a complex environment involving urban canyons, indoor environment, and indoor-outdoor transitions.To analyze VINS algorithms separately and observe their advantages and challenges in long-term navigation, we performed experiments using ORB SLAM 3 [2] in carefully selected environments across our university that pushed each of the sensors to their failure cases. To assess global location accuracy, we collected data using an RTK GPS and the NTRIP (Network Transport of RTCM via Internet Protocol) technology, which enabled us to connect to a remote base station and removed the hindrance of having to set up our own base station. To test the algorithms that do tightly coupled GPS-Camera-IMU fusion for state estimation, we used the current state of the art, GVINS [6]. GVINS is an optimization-based method that fuses visual-inertial data with multi-constellation GNSS raw measurements. In this work, the authors have been able to provide a 6-DOF global drift-free estimation considering situations where GNSS signals could be intercepted or totally unavailable. We tested this algorithm on the complex environment dataset released by the Hong Kong University of Science and Technology - Aerial Robotics Group [9].

II. Experimental Setup

We conducted experiments on both our own hardware setup and a complex environment dataset released by HKUST. Our primary aim was to conduct experiments on GVINS [6] for both the dataset and the live data from our hardware setup. Our attempts to conduct experiments in a

*Equal contribution to the work

tightly coupled manner using this method failed due to an issue in implementing their feature extractor module. Our attempts to contact the authors weren't successful too. So, we collected data using our hardware setup in an uncoupled manner to investigate the challenges in VINS methods and GPS separately and analyze the need for sensor fusion, which for our sensor suite, is a tightly coupled global estimation module.

A. Dataset

The popular datasets used for SLAM, such as EuRoC [10] and KITTI [11], involved specific scenarios - either indoors or outdoors. There were very few publicly available datasets that involved a complex environment that involved indoor-outdoor transitions. We implemented GVINS on the complex environment dataset [9] released by HKUST. This dataset covers different scenarios, such as areas with dim or bright light, indoor-outdoor transitions, and areas with GPS inaccessibility.

B. Hardware Setup

We conducted experiments on our Hardware setup to analyze live data. To understand the challenges, we divided our experiments into four categories:

- **Open Outdoor**- A non-covered area with better sky exposure to have the best GPS measurements.
  Location: Columbus Garage Rooftop (Path as shown in Fig. 2)
- **Urban Outdoor**- A partially-open area covered by buildings to observe some common issues in GPS and drift in VINS estimation.
  Location: A selected region around the campus (Path as shown in Fig. 3)
- **Semi-Indoor**- Path involving transitions from outside to inside a building. This is our complex environment involving lighting changes, accumulated inertial errors, and GPS inaccuracies.
  Location: Path from Snell Library to Ruggles station via ISEC building. (Path as shown in Fig. 4)
- **Pure Indoor**- Path involving complete GPS failure.
  Location: Underground tunnels connecting different buildings (Starting from Curry Student Center and going back to the same place.Path as shown in Fig. 5)

These ensured that we were able to consider scenarios where one of the sensors (GPS/IMU/Camera) would fail and the other would be able to complement this failure and would provide support in the scenario of fused state estimation.

As shown in Fig. 1, our hardware setup contains a ZED stereo camera, an inbuilt synchronized IMU, and an u-blox ZED-F9P GNSS receiver. Images are received at a frame rate of 15 Hz, while RTK-GPS signals are received at a rate of 10 Hz. IMU measurements are streamed with a frequency of 200 Hz. The ZED-F9P has an internal RTK engine, and the corrections for the receiver were obtained from the base station in Clarksburg, Massachusetts, through the internet (NTRIP Technology), via rtk2go.com. Using NTRIP, we could find accurate fix solutions much more quickly without setting up our base station.

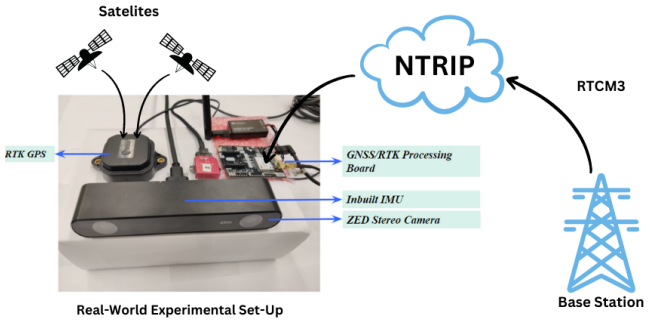

Fig. 1. **Our Real-world experiment setup consists of a ZED stereo camera, an inbuilt synchronized IMU, and an u-blox ZED-F9P GNSS receiver. The corrections from the remote base station for the GNSS receiver were obtained through the internet (NTRIP Technology). NTRIP consists of two main subsystems: NTRIP Caster at the server side & NTRIP Client at the receiver side. The task of the NTRIP caster is to receive the data stream from the base station and rebroadcast it to the NTRIP Client over a specified TCP (Transmission Control Protocol) port. This setup helped to eliminate the need to set up our own base station.**

So far, ORB SLAM3 has the best accuracy in terms of VINS algorithms [12]. To evaluate local state estimations, we utilized trajectory data from running ORB SLAM3 with GPS data obtained concurrently. This was done to observe the situations where VINS is more beneficial than GPS and vice versa.

III. Analysis

A. Visualization

The trajectories of experiments were visualized in Google Maps using a MATLAB function "paby/plot_google_map" [13]. The latitude & longitude obtained from the GNSS receiver can be plotted directly using the function since they are represented in the global frame. However, the trajectory from VINS must be pre-processed before plotting since it is represented in a local frame. The starting point of the VINS trajectory must first be changed to match that of the RTK GPS trajectory so that both tracks begin at the same place. Then the heading is aligned so that the first straight line from RTK GPS and VINS trajectory are oriented in the same direction. The aligned coordinates of the VINS trajectory were converted to UTM coordinates and then to latitude & longitude for visualizing the trajectory in Google Maps.

B. Real-World Experiment

1) Open Outdoor: This experiment is conducted at the top of the Columbus Garage, an outdoor environment in an "L" shaped route. This is a typical outdoor environment with an open area. In this experiment, we attempted to capture a scenario where we have a fix state in the RTK GPS permanently to enable easy qualitative and visual evaluation of the GPS trajectory. From Fig. 2, we can observe a significant drift in the VINS trajectory (red line) compared to

the GPS trajectory (blue line). This RTK GPS trajectory can be taken as the ground truth. This was an expected result. VINS data is produced in a Local reference frame, and the origin is when the VINS data begins the estimation. That's why trajectory evolves according to that starting point, and drift grows at each pose (as seen in Fig. 2. Up to a certain point, RTK GPS and VINS are aligned in a smooth manner. After some time, since the drift error gets more prominent at each pose, the VINS trajectory starts to move away from the RTK trajectory).

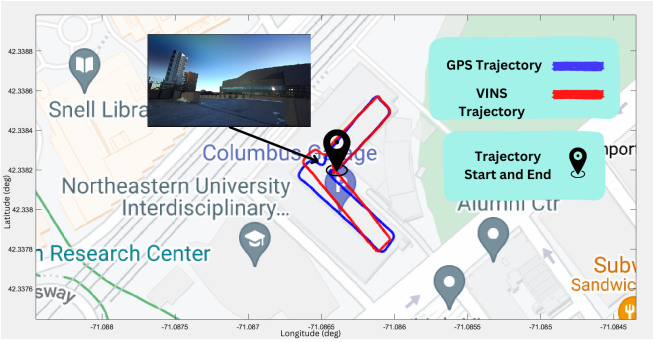

Fig. 2. **The trajectory of RTK GPS and VINS in the Open Outdoor experiment**

2) Urban Outdoor: The path of this experiment started in front of Hayden hall, then towards Ruggles-T station, Huntington Ave, and back to the same place via Forsyth Street. We sought to observe the cumulative drift that affects VINS estimations over a longer period of time. Here the GPS remains in fix state for most of the trajectory and shifts to float in very few regions due to the buildings. In this scenario, we observed that the GPS data (blue line) collected in the urban outdoor environment is accurate to a great extent and the drift is also minimum. VINS data (red line), on the other hand, is correct for the first half of the whole course, but drift can be seen in the second half of the route (it grows at each step gradually). These observations can be seen in Fig. 3. IMU data is always susceptible to drift because of the sensor noise, bias instability, scale factor error, measurements in the local frame, and misalignment of the sensor itself. Hence, there is a difference in the GPS and the VINS path. Thus we need to fuse GPS estimates to minimize this drift in VINS estimates.

3) Semi Indoor: The Semi Indoor dataset trail begins outside the Snell library and leads to ISEC. It can be observed from Fig. 4 that the data collected outdoors is accurate because GPS (blue line) consistently has the fix. But as soon as we entered ISEC, we observed that GPS lost the fix and data showed inaccurate readings. GPS was highly corrupted and unavailable in the indoor environment. To overcome this challenge, we use the data from the VINS (red line) and provide accurate position estimation as it performs very precisely in indoor conditions for a limited duration of time. In general, utilizing GPS for global positioning results in erroneous trajectory owing to satellite orbit inaccuracy, multi-path effect (reflections/scattering of the signal resulting

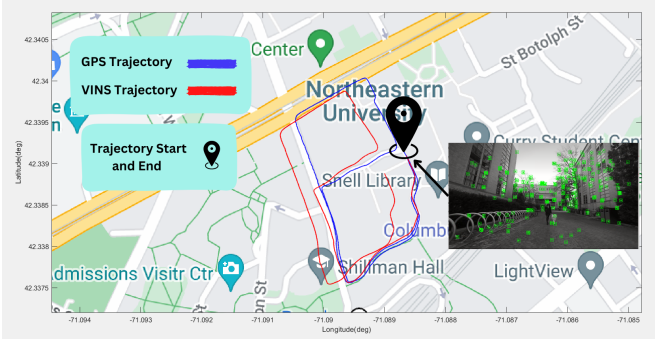

Fig. 3. **The trajectory of RTK GPS and VINS in the Urban Outdoor experiment**

in oscillations in signal intensity), and inaccurate atmospheric delay modeling. Another observation is the features being lost due to bright sunlight. When we exited the ISEC building and started moving towards the Ruggles station, the sudden brightness due to the sunlight caused a loss in feature tracking, but the IMU assisted in continuing the VINS trajectory for a short while. So, the VINS trajectory remains accurate. Thus utilization of IMU measurements for short immediate observations, camera measurements for rich visual information, and GPS for global positioning is a minimum sensor suite required to handle indoor-outdoor transitions.

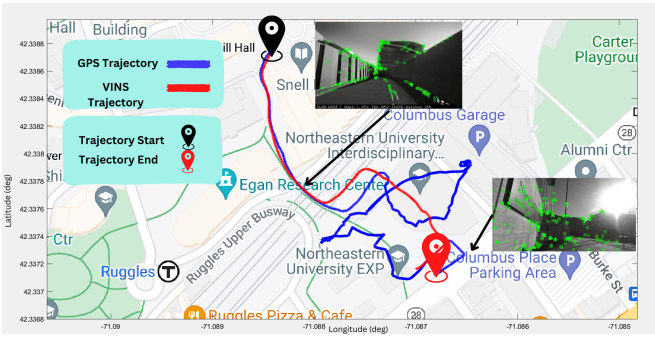

Fig. 4. **The trajectory of RTK GPS and VINS in the Semi Indoor experiment**

4) Pure Indoor: This dataset was collected in the tunnels starting from Curry Student Center and tracing back the same path. Since RTK GPS completely loses signal in indoor environments, Fig. 5 shows only the VINS trajectory (green dots). This experiment is done to show the poor quality and failure of GPS signals in harsh environments where the receiver is blocked by buildings. These conditions are where VINS would help the system maintain its trajectory for a short while.

C. Dataset

We evaluated GVINS for complex environment dataset [9]. This experiment's path covers several tough conditions that may lead a single-sensor-based system to fail. Fig. 6 shows GVINS and GNSS trajectories, and we can identify the places where GVINS and GNSS complement each other.

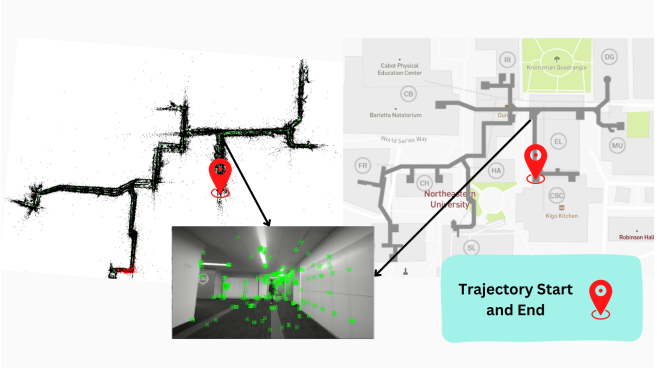

Fig. 5. **The trajectory of VINS in the Pure Indoor experiment**

GVINS functions well in indoor zones, as seen in Fig. 6, where the RTK signal exhibits incorrect behavior. Moreover, for larger distances, accumulated drift becomes inevitable in VINS, where GNSS has shown its ability to reduce accumulated drift. In areas with bright sunlight, the camera fails to detect features, and at such places, the data from IMU and RTK aids in positioning, and the GVINS trajectory continues to provide the right trajectory. Hence we can say that a tightly coupled optimization system fusing measurements from the camera, IMU, and GNSS receiver attains both local smoothness and global consistency as mentioned in their work.

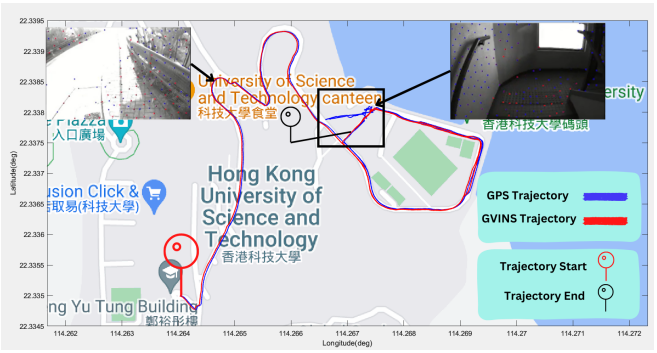

Fig. 6. **The trajectory of RTK GPS and GVINS in the complex environment dataset**

IV. Conclusion and Future Work

In this study, we analyzed the problems and possible sensor failures faced during pose estimation with VINS and RTK GPS measurements in multiple environments. We conducted real-world experiments to evaluate the performance of GPS and VINS under different circumstances, and our results show that even though VINS measurements provide locally accurate pose estimation, drift error coming from local frame data expands gradually over a longer duration of time. GPS signal loses its accuracy and becomes highly corrupted in urban areas, around high-rise buildings, and in indoor places. However, the fusion of GPS with visual-inertial methods can provide a more robust state estimation in complex environments.

In future work, we will try to fuse the measurements of GNSS, IMU, and camera together from our hardware setup to have a locally accurate and globally drift-free trajectory estimation. We expect to get our queries resolved by the GVINS team so that we can apply the GVINS algorithm and verify its effectiveness/robustness compared to other sensor fusion approaches [6].

We observed from the NUANCE dataset that the GPS data was collected from the BU-353 GPS module. We would love to have an opportunity to set up this NTRIP-based RTK-GPS module for the NUANCE car to utilize its higher accuracy.

Acknowledgement

This project would not have been possible without the exceptional support of our Professor, Kris Dorsey. We are also grateful to the Teaching Assistants whose invaluable support has helped us work with the sensors. We also thank Prof. David M. Rosen for the additional hardware we borrowed from his lab for this project.

References

[1] M. Siegel, "The sense-think-act paradigm revisited," in *1st International Workshop on Robotic Sensing, 2003. ROSE'03.* IEEE, 2003, pp. 5–pp.

[2] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[3] V. Usenko, N. Demmel, D. Schubert, J. Stückler, and D. Cremers, "Visual-inertial mapping with non-linear factor recovery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 422–429, 2019.

[4] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," 2019.

[5] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: an open-source library for real-time metric-semantic localization and mapping," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020. [Online]. Available: https://github.com/MIT-SPARK/Kimera

[6] S. Cao, X. Lu, and S. Shen, "Gvins: Tightly coupled gnss–visual–inertial fusion for smooth and consistent state estimation," *IEEE Transactions on Robotics*, 2022.

[7] J. Liu, W. Gao, and Z. Hu, "Optimization-based visual-inertial slam tightly coupled with raw gnss measurements," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 612–11 618.

[8] W. Lee, P. Geneva, Y. Yang, and G. Huang, "Tightly-coupled gnss-aided visual-inertial localization," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9484–9491.

[9] C. Shaozu, "Gvins-dataset public," https://github.com/HKUST-Aerial-Robotics/GVINS-Dataset, 2022.

[10] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, 2016. [Online]. Available: http://ijr.sagepub.com/content/early/2016/01/21/0278364915620033

[11] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[12] D. Sharafutdinov, M. Griguletskii, P. Kopanev, M. Kurenkov, G. Ferrer, A. Burkov, A. Gonnochenko, and D. Tsetserukou, "Comparison of modern open-source visual SLAM approaches," *CoRR*, vol. abs/2108.01654, 2021. [Online]. Available: https://arxiv.org/abs/2108.01654

[13] Yehuda, "zoharby/plot_google_map," https://github.com/zoharby/plot_google_map, 2022.