Benchmarking ground truth trajectories with robotic total stations

Effie Daum[1,2], Maxime Vaidis[1] and François Pomerleau[1]

***Abstract*— Benchmarks stand as vital cornerstones in elevating Simultaneous Localization and Mapping (SLAM) algorithms within mobile robotics. Consequently, ensuring accurate and reproducible ground truth generation is vital for fair evaluation. A majority of outdoor ground truths are generated by Global Navigation Satellite System (GNSS), which can lead to discrepancies over time, especially in covered areas. However, research showed that Robotic Total Station (RTS) setups are more precise and can alternatively be used to generate these ground truths. In our work, we compare both RTS and GNSS systems' precision and repeatability through a set of experiments conducted weeks and months apart in the same area. We demonstrated that RTS setups give more reproducible results, with disparities having a median value of $8.6\,\mathrm{mm}$ compared to a median value of $10.6\,\mathrm{cm}$ coming from a GNSS setup. These results highlight that RTS can be considered to benchmark process for SLAM algorithms with higher precision.**

I. INTRODUCTION

Benchmarks play a crucial role in enhancing SLAM algorithms and real-time location algorithms in mobile robotics [1], [2]. It is essential to ensure the accuracy and reproducibility of the ground truth used for fair comparisons between evaluated algorithms [3]. However, outdoor ground truths, primarily generated by GNSS, can lead to disparities between experiments conducted at different times in the same environment, as shown in Figure 1. These variations in GNSS positions result from various sources, such as satellite constellations, ephemerid, and atmospheric conditions. They may cause significant biases when evaluating trajectories through benchmarks [4]. Recently, our research has demonstrated that RTS can generate ground truths in six-Degrees Of Freedom (DOF) with millimeter-level accuracy [5], [6]. Building on these findings, we evaluate the feasibility of using RTSs to generate ground truth trajectories for objective benchmarking of SLAM algorithms. We compare the precision and repeatability of a RTS and GNSS system by analyzing their data taken simultaneously during different deployments.

II. BENCHMARKING STANDARDIZED EXPERIMENTS

A. Standardization of RTS and GNSS protocol

The experiments were conducted following a standardized protocol to ensure accurate and reproducible results. The process began by allowing the RTS surveying instruments to acclimate to the outdoor temperature. Once ready, the instrument was leveled to ensure proper alignment. Three prisms were mounted at different heights on the robotic platform to optimize visibility and tracking. Each prism was associated with its respective RTS, also positioned at different heights to avoid obstructing visibility and to facilitate the extrinsic calibration of the sensors. An essential aspect of the experiment is the extrinsic calibration, where a set of eight to twelve static ground control points is measured around the RTS in a circular configuration to express all the data in a common frame [7]. Finally, after each deployment, we performed a final extrinsic calibration in the laboratory by measuring the positions of prisms and sensors on the robotic platform using a RTS. The same procedure was applied during each experiment to collect consistent and standardized data during all deployments.

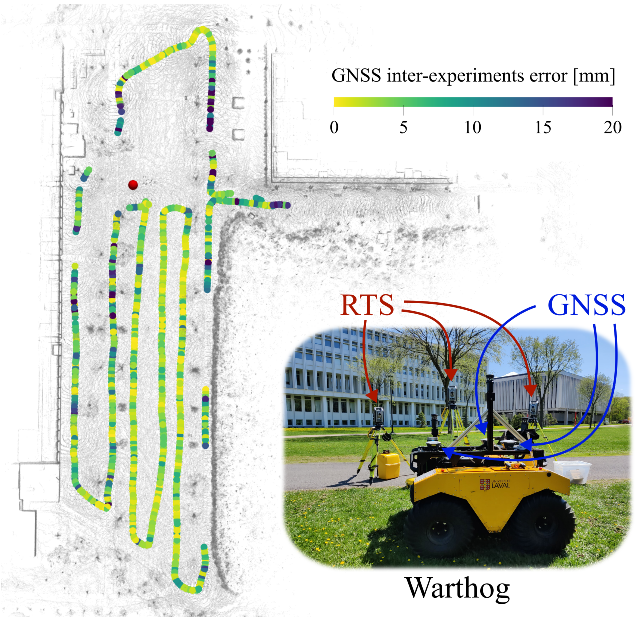

Fig. 1: A RTS setup and GNSS antennas were used to record the trajectory of a *Warthog Clearpath* platform on the Université Laval campus. The color bar displays the average GNSS disparities obtained between two identical trajectories done at different times. The red sphere marks the location of our static Real Time Kinematic (RTK) GNSS reference antenna.

The data obtained from each deployment were processed using our pipeline.[1] The pipeline incorporates various filtering techniques to enhance the precision of the ground truth. These filtering methods contribute to minimizing noise and errors, ultimately improving the reliability of the generated trajectories. As RTS positions are not taken synchronously, we used the parameters described by Vaidis *et al.* [6] to perform linear interpolation of the positions. A point-to-point

---

[1] Northern Robotics Laboratory, Université Laval, Québec City, Québec, Canada, {maxime.vaidis, francois.pomerleau}@norlab.ulaval.ca

[2] École Supérieure des Géomètres-Topographes, Le Mans, France, effie.daum.auditeur@lecnam.net

[1] `https://github.com/norlab-ulaval/RTS_project`

minimization is used to reconstruct the full pose of the vehicle by measuring three rigid points [5]. By leveraging this comprehensive processing pipeline and utilizing the same parameters, the experiment aimed to achieve more precise results, thus facilitating the evaluation and validation of the robotic platform's localization and mapping performance.

Additionally, three GNSS antennas are mounted on top of the prisms to achieve precise positioning. A fourth static antenna, located nearby with known global geodesic coordinates, provides real-time corrections to the three mobile antennas on the robotic platform for RTK-positioning. The RTK method allows obtaining real-time measurements for the trajectory of the moving platform. Establishing a radio connection between the static antenna and the three mobile antennas, along with setting mask parameters, is crucial for the system's GNSS method. By using the same point-to-point minimization method as for the RTS solution [5], the robot ground truth trajectory can be determined in the GNSS frame, through the extrinsic calibration of the GNSS antennas done in the laboratory.

B. *Metrics*

An inter-distance metric is used to evaluate the precision of each system. This metric is computed with the distance between each synchronous triplet of RTS target position (inter-prism distances) or GNSS antenna position (inter-GNSS distances) obtained during an experiment. Each of these distance triplets is then compared to their RTS calibrated distance, i.e., the position of the prisms or GNSS antennas rigidly installed on the robot. Moreover, an inter-experiment metric is used to quantify the difference in precision obtained between two experiments done at different times. Two positions taken during different experiments are assessed to be in close range by computing their nearest neighbor distance. Then, each inter-distance triplet of the RTS prisms or GNSS antenna positions that matched spatially are subtracted to compute this metric. The results represent the disparities in precision in-between the two different trajectories taken at a different time, as shown in Figure 1 for a set of GNSS data.

III. RESULTS

The RTS setup is composed of three *Trimble* S7 surveying instruments that track three *Trimble MultiTrack Active Target* MT1000 prisms, operating at a measurement rate of 2.5 Hz. The prisms are mounted on a *Clearpath Warthog* unmanned ground vehicle, along with three *Emlid RS+* GNSS antennas. To analyze the disparities of the different setups, eleven experiments were conducted weeks and months apart on the same area of the Université Laval campus, for a total of 16 km of GNSS and RTS-tracked prism trajectories.

The Figure 2 illustrates the inter-prism and inter-GNSS metric errors, indicating that the RTS acquisition system achieves median sub-centimeter precision at 6.8 mm, while the GNSS system provides a median precision around 1.35 cm. The GNSS precision aligns with results from an RTK method, showing within 2 cm accuracy in static scenarios [8]. These outcomes are especially promising considering the dynamic nature of the robotic platform. It's worth noting that the inter-distances error highlights the higher precision of the RTS acquisition system compared to the GNSS system. This discrepancy can be attributed to the relatively low error of the RTS acquisition system versus the absolute error of GNSS related to the satellites' constellation, even in open-sky and large-space environments where the experiments were done.

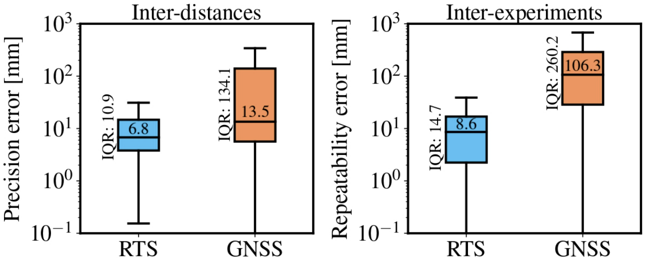

Fig. 2: Error resulting from **(a)** inter-prisms and inter-GNSSs metrics and, **(b)** inter-experiments metrics presented in Section II-B. The results from the RTS are depicted in blue, while those from the GNSS are represented in orange. The median error is displayed at the center of each box, and the Interquartile Range (IQR) is depicted on the side.

Reproducibility between the experiments is assessed by computing the nearest neighbor distance. Points falling within a 2 m range are considered reproducible between the different experiments. As evident in Figure 2, the reproducibility appears consistent for the RTS setup. This showcases that precision remains consistent across all experiments with a median margin of 8.6 mm. However, the GNSS system has higher disparities at a median level of 10.6 cm. The ground truth trajectory generated is displayed in Figure 1 with the color gradient showing the inter-experiments error and grey points representing the SLAM generated map. These differences highlight that GNSS is less prompt to give reproducible ground truth trajectories with lower uncertainty in the test conditions.

IV. CONCLUSION

In this paper, we successfully integrated both RTS and GNSS ground truth acquisition systems for trajectory reconstruction. RTSs offers a valuable solution for benchmarking due to their higher precision, their median reproducibility around 8.6 mm, and applicability as shown in Figure 2. Moreover, they can be used in both indoor and outdoor environments compared to GNSS. Results for GNSS are as expected, with higher disparities at a median level of 10.6 cm, making it a relevant subsidiary to obtain reproducible trajectories. However, it's important to note that RTS has certain limitations, such as line of sight dependency, higher cost, and post-processing requirements. Despite these drawbacks, combining both RTS and GNSS systems presents a favorable trade-off. This approach enables us to generate accurate ground truth trajectories and enhance reproducibility, thereby improving the overall benchmarking process for SLAM algorithms. Future works should consider the complementarity of both systems for six-DOF trajectory reconstruction.

REFERENCES

[1] M. Burri, J. Nikolic, P. Gohl, *et al.*, “The EuRoC micro aerial vehicle datasets,” 10, vol. 35, SAGE Publications, Jan. 2016, pp. 1157–1163.

[2] M. Helmberger, K. Morin, B. Berner, N. Kumar, G. Cioffi, and D. Scaramuzza, “The HILTI SLAM challenge dataset,” 3, vol. 7, IEEE, 2022, pp. 7518–7525.

[3] E. Maset, L. Scalera, A. Beinat, D. Visintini, and A. Gasparetto, “Performance investigation and repeatability assessment of a mobile robotic system for 3d mapping,” *Robotics*, vol. 11, no. 3, 2022.

[4] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 2012, pp. 3354–3361.

[5] M. Vaidis, P. Giguere, F. Pomerleau, and V. Kubelka, “Accurate outdoor ground truth based on total stations,” in *2021 18th Conference on Robots and Vision (CRV)*, IEEE, May 2021.

[6] M. Vaidis, W. Dubois, E. Daum, D. LaRocque, and F. Pomerleau, “Uncertainty analysis for accurate ground truth trajectories with robotic total stations,” in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, Detroit, U.S.A., 2023.

[7] M. Vaidis, W. Dubois, A. Guénette, J. Laconte, V. Kubelka, and F. Pomerleau, “Extrinsic calibration for highly accurate trajectories reconstruction,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, London, England, 2023.

[8] Y. Morales and T. Tsubouchi, “DGPS, RTK-GPS and Starfire DGPS performance under tree shading environments,” in *2007 IEEE International Conference on Integration Technology*, IEEE, 2007, pp. 519–524.