Comparative Study of Visual SLAM-Based Mobile Robot Localization Using Fiducial Markers

Jongwon Lee[1], Su Yeon Choi[1], David Hanley[2], and Timothy Bretl[1]

Abstract— This paper presents a comparative study of three modes for mobile robot localization based on visual SLAM using fiducial markers (i.e., square-shaped artificial landmarks with a black-and-white grid pattern): SLAM, SLAM with a prior map, and localization with a prior map. The reason for comparing the SLAM-based approaches leveraging fiducial markers is because previous work has shown their superior performance over feature-only methods, with less computational burden compared to methods that use both feature and marker detection without compromising the localization performance. The evaluation is conducted using indoor image sequences captured with a hand-held camera containing multiple fiducial markers in the environment. The performance metrics include absolute trajectory error and runtime for the optimization process per frame. In particular, for the last two modes (SLAM and localization with a prior map), we evaluate their performances by perturbing the quality of prior map to study the extent to which each mode is tolerant to such perturbations. Hardware experiments show consistent trajectory error levels across the three modes, with the localization mode exhibiting the shortest runtime among them. Yet, with map perturbations, SLAM with a prior map maintains performance, while localization mode degrades in both aspects.

I. INTRODUCTION

The use of fiducial markers—square-shaped planar artificial landmarks with a black-and-white grid pattern—has been favored for the application of visual simultaneous localization and mapping (SLAM) in the scenario when these markers can be deployable in the given environment. This preference emerges due to the robustness and accuracy exhibited by fiducial marker-based SLAM approaches [1-4] over canonical approaches using visual features (e.g., ORB-SLAM [5]) or pixel values (e.g., DSO [6]). The spectrum of fiducial marker-based SLAM spans from methods exclusively utilizing fiducial marker detection outcomes [1-3] to hybrid techniques incorporating both marker detections and features [4]. As our focus here centers on fiducial marker-based SLAM, it is important to note that the subsequent analysis dedicates solely to SLAM with fiducial markers.

Within this context, three distinct operational modes of fiducial marker-based SLAM are discernible. The first mode, *SLAM*, entails a comprehensive approach that estimates the robot's pose while mapping out the surrounding environment.

The second mode, *SLAM with a prior map*, starts from a pre-existing map as an initial reference value. In cases where the map is a priori known and its states remain fixed, the third mode, *localization*, exclusively estimates the robot's pose.

A noticeable gap in the existing literature lies in the absence of comparisons between these three modes, through quantifiable metrics such as error analysis and processing speed. While certain fiducial marker-based SLAM approaches, such as SPM-SLAM [1] and TagSLAM [2], report error and processing speed the *SLAM* mode, they do not provide corresponding data for the *SLAM with a prior map* and *localization* modes.

Another crucial aspect yet to be addressed pertains to the resilience of the *SLAM with a prior map* and *localization* modes in response to variations in marker map quality. Indeed, the performance of these modes is influenced by the fidelity of the marker map. If the map is considered "ideal" or at least of a high standard, both modes are anticipated to yield comparable or superior error performance while placing a lesser computational burden in contrast to the *SLAM* mode. This attribute proves particularly advantageous in scenarios requiring real-time robot localization under resource-constrained, onboard platforms. However, if the map quality deviates from such conditions, it may introduce degradation of the performance of each operational mode.

In this paper, we aim to provide an evaluation of the three different operational modes for robot localization based on SLAM with fiducial markers, in terms of absolute trajectory error and processing speed. In particular, for the last two modes (*SLAM with a prior map* and *localization* modes), we evaluate their performances by perturbing the accuracy of the prior map to study the extent to which each mode is tolerant to such perturbations.

We present an overview of the structure and content of our paper as follows. Section II outlines our data collection and experimental setup to perform fiducial marker-based SLAM in three different operational modes. We then proceed to report the metrics comparing the three modes in Section III, which include the absolute trajectory error and the runtime for the optimization process per frame. We specifically provide the runtime for the optimization process per frame because it is the major component that contributes to variations in processing speed across the three modes, unlike the other components which are more or less the same such as detecting fiducial markers. Finally, Section IV summarizes our findings, provides concluding remarks, and discusses the implications of our work.

[1]Jongwon Lee, Su Yeon Choi, and Timothy Bretl are with the Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (Email: {jongwon5, suyeonc3, tbretl}@illinois.edu).

[2]David Hanley is with the School of Informatics and the Translational Healthcare Technologies Group in Centre for Inflammation Research, Institute for Regeneration and Repair, University of Edinburgh, UK (Email: dhanley@ed.ac.uk).

II. EXPERIMENTS

A. Data collection

We collected an indoor dataset comprising multiple sequences of camera images and Vicon Mocap data as ground truth using a hand-held device equipped with an Intel Realsense D435i and reflective markers. During data collection, five 36h11 AprilTags [7] with side lengths of 0.2m were placed within the environment, as depicted in Fig. 1.

The dataset consists of eight sequences: one sequence lasts 60 seconds, dedicated to generating a prior map using the *SLAM* mode for the *SLAM with a prior map* and *localization* modes; the other seven sequences last 30 seconds each, serving all three modes. The distinguishing factor between these two categories of sequences is that the former captures all AprilTags to create a comprehensive prior map, crucial for the *SLAM with a prior map* and *localization* modes, while the latter does not require this complete mapping of all the AprilTags placed.
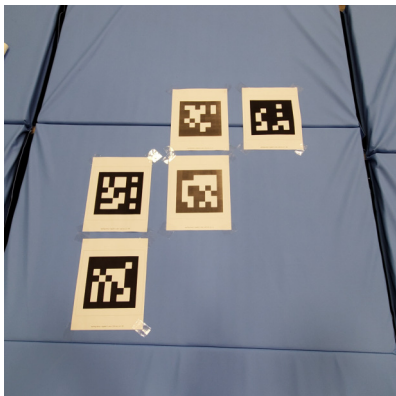

Fig. 1: For data collection, an experimental setup was arranged with the placement of five 36h11 AprilTags, each having side lengths of 0.2m.

B. Implementation details

We used the existing WOLF codebase [3] to execute the *SLAM* mode (constructing a marker map from scratch), the *SLAM with a prior map* mode (loading a pre-saved marker map for initial estimation and performing SLAM), and the *localization* mode (using a pre-saved marker map as a fixed reference for localization). As a preliminary step, the *SLAM* mode was employed to create a prior map for both the *SLAM with a prior map* and *localization* modes. It is important to note that this marker map is not flawless, as it results from estimating the poses of fiducial markers rather than using their true references, inherently harboring potential sources of error even before intentional perturbation is introduced. For the *SLAM with a prior map* and *localization* modes, we systematically perturbed the positions of every fiducial marker along directions at random within the map, varying from 0.1m to 0.5m in every 0.1m. This approach allowed us to investigate the tolerance of each method to variations in the marker map's quality. We assessed the absolute trajectory error using an open-source evaluation tool [8] and the runtime for the optimization process per frame through Ceres Solver [9]. All executions and evaluations were carried out on an octa-core Intel i7-10700 CPU operating at 2.90 GHz, with 32 GB of RAM.

III. RESULTS

Table I presents the absolute trajectory error for each of the three localization modes across all sequences. Specifically, for the *SLAM with a prior map* and *localization* modes, we provide results by perturbing the positions of fiducial markers within the prior marker map. Again, the prior map is imperfect and hence is inherently regarded to be perturbed even before intentional perturbation is introduced. The perturbations range from 0.1m to 0.5m, with increments of 0.1m, along directions at random.

When no perturbation is applied, all three modes yield results that exhibit minimal differences within a few centimeters. However, as perturbations are introduced to the *localization* mode, errors increase to tens of centimeters even with the smallest perturbation (i.e., $\delta\mathbf{p} = 0.10$), while the *SLAM with a prior map* mode consistently maintains results within a few centimeters. This aligns with the common understanding that the *SLAM with a prior map* mode is capable of recovering reference marker pose values within the map by concurrently updating both localization and mapping outcomes, whereas the *localization* mode cannot as the map is regarded accurate and hence is fixed during optimization, thereby limiting updates to only the localization results.

Table II presents the runtime for the optimization process per frame for each of the three localization modes across all sequences. Similar to the error analysis discussed earlier, we also present results for the *SLAM with a prior map* and *localization* modes by introducing inherent perturbations to the positions of fiducial markers within the prior marker map. Again, it is worth noting that the prior marker map itself already contains undesired perturbations due to its inherent imperfection.

The *SLAM with a prior map* mode does not exhibit any distinct runtime trend in relation to the extent of perturbation applied. However, in a general sense, it shows a slightly longer runtime (up to about 20%) than the *SLAM* mode. Conversely, the *localization* mode shows shorter runtime (up to about 20%) than the *SLAM* mode when no perturbation is introduced. Yet, as the perturbation level increases, the *localization* mode experiences a corresponding rise in runtime, culminating in up to about 40% longer runtime at $\delta\mathbf{p} = 0.50$ compared to the *SLAM* mode. This behavior is indicative of the *localization* mode grappling with ill-posed problems stemming from inaccurate map quality.

IV. CONCLUSIONS

In summary, this paper conducted a comparative analysis of three distinct robot localization modes based on visual SLAM with fiducial markers: *SLAM*, *SLAM with a prior map*, and *localization with a prior map*. We evaluated these

TABLE I: Comparison of absolute trajectory error for SLAM, SLAM with prior map, and localization modes. Perturbations ($\delta\mathbf{p}$) applied to fiducial marker positions in the prior map for SLAM with prior map and localization modes.

sequence	length [m]	SLAM [m]	SLAM with a Prior Map [m] $\delta\mathbf{p}$ [m]						Localization [m] $\delta\mathbf{p}$ [m]					
			0.00	0.10	0.20	0.30	0.40	0.50	0.00	0.10	0.20	0.30	0.40	0.50
1	8.07	0.07	0.05	0.05	0.07	0.08	0.08	0.08	0.06	0.23	0.22	0.44	1.18	1.21
2	8.34	0.08	0.07	0.07	0.08	0.04	0.10	0.10	0.08	0.25	0.31	0.60	1.58	1.69
3	8.21	0.08	0.08	0.08	0.10	0.10	0.09	0.10	0.10	0.21	0.28	0.45	1.28	1.34
4	8.46	0.14	0.14	0.14	0.15	0.15	0.14	0.14	0.15	0.45	0.43	0.58	0.96	2.35
5	8.19	0.13	0.13	0.13	0.13	0.11	0.13	0.13	0.14	0.40	0.52	0.44	1.28	2.43
6	8.22	0.10	0.08	0.09	0.10	0.10	0.10	0.10	0.13	0.39	0.42	1.44	2.04	2.50
7	8.57	0.12	0.12	0.11	0.12	0.12	0.11	0.11	0.11	0.49	0.42	0.61	1.78	1.54
average		0.10	0.10	0.10	0.11	0.10	0.11	0.11	0.11	0.35	0.37	0.65	1.44	1.87

TABLE II: Comparison of runtime of pose-graph optimization process per frame for SLAM, SLAM with prior map, and localization modes. Perturbations ($\delta\mathbf{p}$) applied to fiducial marker positions in the prior map for SLAM with prior map and localization modes.

sequence	SLAM [ms]	SLAM with a Prior Map [ms] $\delta\mathbf{p}$ [m]						Localization [ms] $\delta\mathbf{p}$ [m]					
		0.00	0.10	0.20	0.30	0.40	0.50	0.00	0.10	0.20	0.30	0.40	0.50
1	5.07	5.26	5.79	5.00	5.75	6.23	6.42	3.05	3.53	3.90	4.56	5.33	6.06
2	4.92	5.81	5.15	5.92	6.06	5.70	6.15	4.69	5.28	5.09	6.26	7.52	8.35
3	5.31	6.43	5.71	5.50	5.20	6.30	6.11	4.60	4.33	4.71	4.90	6.20	7.12
4	5.73	6.05	5.91	6.77	6.35	6.63	6.66	4.80	4.49	7.66	4.99	6.29	7.60
5	6.23	6.19	7.34	6.27	7.42	7.02	6.81	5.81	6.33	6.83	8.06	8.61	9.76
6	6.08	6.56	4.64	6.69	6.61	7.46	6.30	4.78	7.44	6.15	7.43	9.92	9.75
7	7.95	8.80	9.23	7.45	8.06	9.11	7.94	5.22	6.66	8.13	9.19	8.47	8.99
average	5.90	6.44	6.25	6.23	6.49	6.92	6.63	4.71	5.44	6.07	6.48	7.48	8.23

modes in terms of trajectory error and runtime for the optimization process. Specifically, we introduced perturbations to the map for the *SLAM and localization with prior map* modes to examine their impact on the aforementioned metrics.

When no perturbations were introduced, our hardware experiments show that all three modes exhibited similar levels of trajectory error, while the *localization* mode showed the shortest runtime. However, in scenarios involving map perturbations, the *SLAM with a prior map* mode maintained its trajectory error and runtime performance at levels comparable to those in the absence of perturbation. On the other hand, the *localization* mode experienced deteriorating trajectory error and runtime performance as the magnitude of map perturbation increased.

REFERENCES

[1] R. Munoz-Salinas, M. J. Marin-Jimenez, and R. Medina-Carnicer, "Spm-slam: Simultaneous localization and mapping with squared planar markers," *Pattern Recognition*, vol. 86, pp. 156–171, 2019.

[2] B. Pfrommer and K. Daniilidis, "Tagslam: Robust slam with fiducial markers," *arXiv preprint arXiv:1910.00679*, 2019.

[3] J. Sola, J. Vallvé, J. Casals, J. Deray, M. Fourmy, D. Atchuthan, A. Corominas-Murtra, and J. Andrade-Cetto, "Wolf: A modular estimation framework for robotics based on factor graphs," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4710–4717, 2022.

[4] R. Munoz-Salinas and R. Medina-Carnicer, "Ucoslam: Simultaneous localization and mapping by fusion of keypoints and squared planar markers," *Pattern Recognition*, vol. 101, p. 107193, 2020.

[5] 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.

[6] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[7] M. Krogius, A. Haggenmiller, and E. Olson, "Flexible layouts for fiducial tags," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1898–1903.

[8] M. Grupp, "evo: Python package for the evaluation of odometry and slam." https://github.com/MichaelGrupp/evo, 2017.

[9] S. Agarwal, K. Mierle *et al.*, "Ceres solver," 2012.