Late Breaking Results on Visual S-Graphs for Robust Semantic Scene Understanding and Hierarchical Representation

Ali Tourani[1], Hriday Bavle[1], Jose Luis Sanchez-Lopez[1], Rafael Muñoz Salinas[2], and Holger Voos[1]

I. Introduction and Motivation

The primary advantage of vision sensors in Visual SLAM (VSLAM) is that they supply rich visual information from surroundings, making them attractive methods compared to Light Detection And Ranging (LiDAR)-based tools. Accordingly, by processing the visual data, semantic data can be extracted and employed for improved path planning and map reconstruction. Semantic data, *i.e.,* high-level information acquired from the environment, expand the range of the applications that can employ the reconstructed maps, such as object identification and dynamic object removal. Among various solutions, employing fiducial markers such as ArUcos [1] can aid in providing accurate pose estimation, supplying reliable features in low-texture conditions, and encoding semantic information into the environment.

II. Problem Statement and Proposed Solution

Some VSLAM frameworks such as UcoSLAM [2] and TagSLAM [3] employ fiducial markers to create purely geometric map representations. However, they do not contribute to encoding meaningful semantic information to improve the accuracy of pose estimates and reconstructed maps.

The current work introduces an improved version of our previous marker-based approach introduced in [4] with a wider range of sensors and semantic concepts support built upon ORB-SLAM 3.0 [5]. It employs pose information of the markers placed on walls and doorways to estimate the $3D$ equation of their planes, while the markers only keep their correspondence information of the semantic entities. The proposed approach, compared to its baseline and other similar works, produces higher-accuracy reconstructed maps with semantic entities, including walls, rooms, corridors, and doorways. The pipeline of the mentioned approach is depicted in Fig. 1.

III. Evaluation Results

To evaluate the effectiveness of the proposed approach, a dataset of frames captured by an RGB-D camera and a LiDAR as the ground truth was collected using a *Boston Dynamics Spot®* robot. The robot traversed various corridor

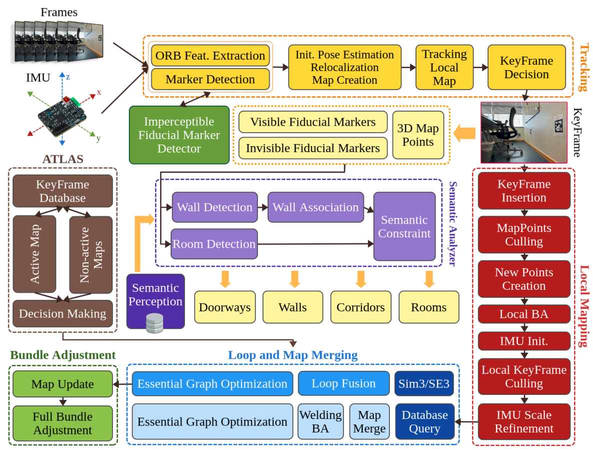

Fig. 1. The primary system components of the proposed approach.

and room setups in real-world circumstances where several walls and doorways were labeled with printed $8cm \times 8cm$ ArUco markers. By analyzing the Root Mean Square Deviation (RMSE) and Standard Deviation (STD) values of the proposed, the baseline (*i.e.,* ORB-SLAM 3.0), ground truth framework [6], UcoSLAM, and Semantic UcoSLAM [4], the proposed framework works better in terms of map reconstruction accuracy, along with providing high-level semantic entities and hierarchical representations.

References

[1] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[2] R. Muñoz-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.

[3] B. Pfrommer and K. Daniilidis, "Tagslam: Robust slam with fiducial markers," 2019. [Online]. Available: https://arxiv.org/abs/1910.00679

[4] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, R. M. Salinas, and H. Voos, "Marker-based visual slam leveraging hierarchical representations," *arXiv preprint arXiv:2303.01155*, 2023.

[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] H. Bavle, J. L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos, "S-graphs+: Real-time localization and mapping leveraging hierarchical representations," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4927–4934, 2023.

[1]Authors are with the Automation and Robotics Research Group, Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, Luxembourg. Holger Voos is also associated with the Faculty of Science, Technology, and Medicine, University of Luxembourg, Luxembourg. {ali.tourani, hriday.bavle, joseluis.sanchezlopez, holger.voos}@uni.lu

[2]Author is with the Department of Computer Science and Numerical Analysis, Rabanales Campus, University of Córdoba, Spain. rmsalinas@uco.es