HILTI SLAM Challenge Submission: VILENS and SLAM

Marco Camurri*, Lintong Zhang*, David Wisth*, Maurice Fallon*

I. Introduction

This report provides an overview of our system used to complete the HILTI SLAM Challenge. We used a combination of VILENS odometry with a loop closure module to estimate the sensor pose in real time. Different VILENS configurations of sensors and modules have been used extensively on legged robots [1], wheeled vehicles, drones, and handheld devices [2], [3] (see Fig. 2) at the Oxford Robotics Institute, University of Oxford.

II. Method

The main module used for estimation is the VILENS odometry algorithm [2], [1], which computes a high frequency state estimate based on lidar, camera, and IMU measurements. Because VILENS does not perform loop closures, a separate SLAM module based on [4] produces low frequency estimates which are used for global pose graph optimization. The motion-corrected lidar scans and the odometry estimate from VILENS are the inputs to the SLAM system. A system overview of our method is shown in Figure 3.

III. VILENS Odometry Module

VILENS is a multi-sensor fusion algorithm which uses a sliding window optimization based on factor-graphs, as shown in Fig. 1. Its modular design allows for different sensor inputs to be easily added and removed depending on the application. As detailed in Section V, for the challenge we have fused prior (black), preintegrated IMU (orange), stereo visual tracking (yellow) and lidar odometry (magenta) factors.

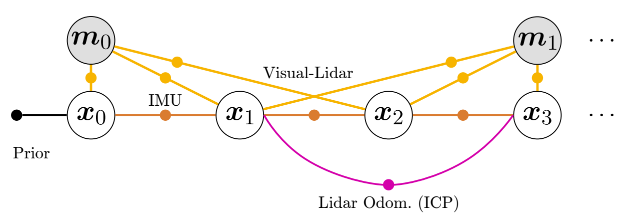

Fig. 1. Sliding-window factor graph structure, showing prior, visual, ICP, and preintegrated IMU factors.

IV. Loop Closure Module

We use the SLAM module with lidar loop closure verification described in [4] to perform global pose optimization with the iSAM2 solver [5]. The module uses the motion-corrected lidar scans from VILENS to construct a map which is used to propose geometric loop closures. These are incorporated in a global pose graph optimization and output at 1 Hz.

All authors are with the Oxford Robotics Institute Department of Engineering Science, University of Oxford, UK {mcamurri, lintong, davidw, mfallon}@robots.ox.ac.uk.
*All authors contributed equally to this submission.

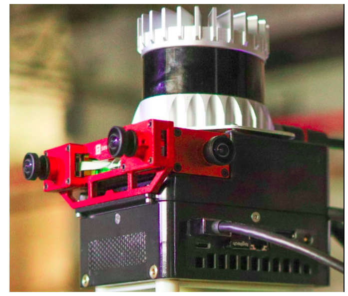

Fig. 2. The VILENS algorithm has been extensively tested on a Handheld Alphasense Ouster device (HALO) at the Oxford Robotics Institute [3]. All measurements are processed by the on-board Intel NUC PC.

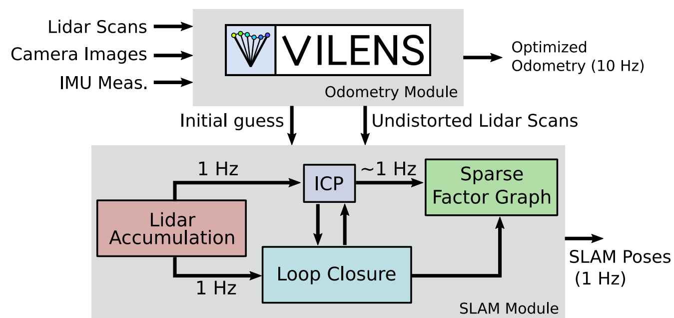

Fig. 3. System diagram composed of high frequency odometry and SLAM.

V. Sensor Modalities

This section will describe the specific settings and sensor inputs used for our submission to the HILTI SLAM Challenge.

Our system is flexible and can be easily configured to use different modalities. Given the moderate complexity level of the HILTI SLAM Challenge dataset, we were able to use our baseline configuration to produce accurate estimation. Specifically, we used the method described in Section II to combine inputs from the Alphasense front stereo pair (*cam0*, *cam1*), the ADIS IMU (which is the highest quality among the IMUs available), and the Ouster Lidar. Due to missing data from both the Alphasense IMU and the ADIS IMU, for some sequences of the dataset the Ouster IMU was used instead. The LIVOX lidar was not used.

VILENS supports other advanced features such as lidar point cloud feature tracking, multi-camera odometry, and kinematics which were not used for the HILTI datasets (see [2], [3] for more details). Note that we used the same parameters throughout all sequences.

VI. Performance

The odometry system outputs the optimized state estimate from the factor-graph at camera keyframe frequency (10 Hz

	Mean APE RMSE		
Dataset	**SVO2**	**GraphSLAM**	**VILENS + SLAM**
Basement_1	0.815	0.279	**0.054**
Basement_4	2.609	0.366	**0.050**
Campus_2	8.948	0.353	**0.122**
Construction_Site_2	2.986	0.741	**0.124**
Lab_Survey_2	0.074	0.053	**0.017**
uzh_tracking_area_run2	1.927	0.350	**0.184**

TABLE I
PERFORMANCE COMPARISON

in this dataset), while a forward-propagated state is produced at IMU frequency (819.2 Hz for the ADIS IMU). The SLAM system detects loop closures and adds a new pose to the global optimization at 1 Hz.

The rosbags were processed in real-time (through ROS) so the total processing time was the same as the duration of the rosbags themselves. The system ran on a consumer-grade laptop with an Intel Core i7-9850H CPU (2.60GHz) and 16GB memory. No GPU was used for the evaluation.

For the purpose of the HILTI Challenge, a further refinement step was performed in post processing by fusing the VILENS output and the SLAM output in a full batch optimization fashion. This yield high frequency estimates while retaining the loop closure capability. Note that this refinement step did not involve any measurement processing, but just the fusion of the two outputs.

In table I, we present the comparisons between the provided results from SVO2 [6], GraphSLAM [7] and our VILENS with loop closure. Note the table only contains half of the HILTI dataset with provided ground truth. Because these trajectories contain a limited number of loop closures, the results from VILENS without loop closure were identical.

VII. CONCLUSION

Our submission, which is based on VILENS and a loop closure module provides accurate estimation across the entire HILTI SLAM Challenge dataset. Our evaluation on the results with ground truth provided shows a substantial reduction in ATE error compared to the reference trajectories provided.

REFERENCES

[1] D. Wisth, M. Camurri, and M. Fallon, “VILENS: Visual, inertial, lidar, and leg odometry for all-terrain legged robots,” *arXiv preprint arXiv:2107.07243*, 2021.

[2] D. Wisth, M. Camurri, S. Das, and M. Fallon, “Unified Multi-Modal Landmark Tracking for Tightly Coupled Lidar-Visual-Inertial Odometry,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1004–1011, 2021.

[3] L. Zhang, D. Wisth, M. Camurri, and M. Fallon, “Balancing the budget: Feature selection and tracking for multi-camera visual-inertial odometry,” *arXiv preprint arXiv:2109.05975*, 2021.

[4] M. Ramezani, G. Tinchev, E. Iuganov, and M. Fallon, “Online lidar-slam for legged robots with robust registration and deep-learned loop closure,” in *IEEE International Conference on Robotics and Automation*, 2020, pp. 4158–4164.

[5] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, “iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering,” in *IEEE International Conference on Robotics and Automation*, 2011, pp. 3281–3288.

[6] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, “Svo: Semidirect visual odometry for monocular and multicamera systems,” *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2017.

[7] K. Koide, J. Miura, and E. Menegatti, “A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement,” *International Journal of Advanced Robotic Systems*, vol. 16, 02 2019.