Benchmark of visual and 3D lidar SLAM systems in simulation environment for vineyards

Ibrahim Hroob, Riccardo Polvara, Sergi Molina,
Grzegorz Cielniak, and Marc Hanheide

Lincoln Center for Autonomous Systems, University of Lincoln, UK
{ihroob,rpolvara,smolinamellado,gcielniak,mhanheide}@lincoln.ac.uk

Abstract. In this work, we present a comparative analysis of the trajectories estimated from various Simultaneous Localization and Mapping (SLAM) systems in a simulation environment for vineyards. Vineyard environment is challenging for SLAM methods, due to visual appearance changes over time, uneven terrain, and repeated visual patterns. For this reason, we created a simulation environment specifically for vineyards to help studying SLAM systems in such a challenging environment. We evaluated the following SLAM systems: LIO-SAM, StaticMapping, ORB-SLAM2, and RTAB-MAP in four different scenarios. The mobile robot used in this study equipped with 2D and 3D lidars, IMU, and RGB-D camera (Kinect v2). The results show good and encouraging performance of RTAB-MAP in such an environment.

Keywords: Agricultural Robotics, visual SLAM, 3D lidar SLAM

1 Introduction

Precision agriculture relies on collecting data from multiple sensors to help improving farm management and crop yield. Better management of the farm requires continuous monitoring of the plant health and soil condition, discovering diseases at an early stage and reducing chemical treatment. To achieve these goals, one solution would be to use a mobile robot to autonomously inspect the plants and the crops. In that context, the mobile robot must have an accurate representation of the farm to accurately localize itself and navigate to the goals. For this reason, many solutions have been proposed to overcome the localization problem in an outdoor environment. The most common solution is to use Real-Time Kinematic Global Positioning Systems (RTK-GNSS) [1], however, this solution is quite expensive and requires good coverage of base stations to guarantee accuracy. Other solutions rely on consumer-grade GNSS with fusing the output with different onboard sensors to enhance localization accuracy [2].

GNSS is not always available and the signal may not be reliable due to environmental conditions; loss of signal for the autonomous robot may lead to catastrophic failures. Therefore, alternative solutions have been proposed based on the Simultaneous Localization and Mapping (SLAM) concept [3], where robot pose is estimated using sensory input, while at the same time building map of the environment. Various SLAM systems have been developed in that regard [4]. The selection of a suitable SLAM system depends on multiple factors, such as type (indoor or outdoor) and scale of the environment. Another factor is the sensors' cost, for example, some systems rely on data from relatively expensive 3D lidars whilst others use data from cheap consumer-grade monocular cameras. However, most of the developed solutions were targeting either the indoor environments or outdoor urban environments [3]. Nevertheless, SLAM in agricultural applications is still a growing field due to challenges related to harsh environmental conditions, seasonal changes in appearance, and repeated visual features in large open fields.

The main contributions of this paper are (i) Releasing to the public an open-source realistic vineyard simulator, offering uneven terrain and five different stages of plant growth[1]. (ii) Comparing and benchmarking 4 SLAM systems in an environment with repeating structure and appearance. The algorithms chosen for this study represent the sate-of-the-art visual and 3D-lidar systems including LIO-SAM [5], StaticMapping [6], ORB-SLAM2 [7], and RTAB-MAP [4].

2 Related work

There has not been much work done in comparing various SLAM systems specifically for the vineyard environment, nevertheless, there have been many research papers dealing with analyzing and comparing SLAM methods for indoor static environments. In this section, we review the most relevant papers in this area.

The authors in [8] evaluated the trajectory generated from different ROS-based SLAM algorithms in a typical office indoor environment. The mobile robot was equipped with a 2D laser scanner, a monocular and stereo camera. The evaluation was on a specifically acquired data-set. The authors used the estimated trajectory from the best performing 2D lidar SLAM as the ground truth for visual SLAM systems. The results were good and encouraging for RTAB stereo with Root Mean Square Error (RMSE) of 0.163 m, and for ORB-SLAM monocular with RMSE of 0.166 m. A precise laser tracker was used in [9] for accurate ground truth to evaluate the accuracy of the map and the expected trajectory generated from the three most common 2D SLAM algorithms, gmapping, hector_slam, and google cartographer. The results showed that in this particular scenario, Google Cartographer is the most accurate algorithm compared to others.

Part of the work in [4] is an evaluation of the trajectory performance between different sensor configuration of RTAB-MAP (stereo and lidar), LSD-SLAM (stereo), ORB-SLAM2 (stereo) and SOFT-SLAM (stereo) in the outdoor KITTI dataset [10]. At this dataset, the authors stated both lidar and stereo configuration have followed well the ground truth. However, in some sequences with not a complex structure, stereo setup systems outperform the lidar configuration of RTAB-MAP. Stereo RTAB with ORB feature detection has performed well

[1] github.com/LCAS/bacchus_lcas

in 9 out of 11 sequences, but that configuration was computationally expensive and was not able to meet real-time constraint on their hardware setup.

An Extended Information Filter (EIF) was used for mapping an agricultural environment and localizing a mobile robot [11]. The system makes use of a 2D laser scanner and monocular camera to detect olive tree stems. The tests were done in a real agricultural environment. However, the authors stated that they find some errors in identifying features and detecting loop closure. In [12] the authors evaluated visual SLAM methods, such as ORB-SLAM2 and S-PTAM, against visual-inertial SLAM system as S-MSCKF on the Rosario dataset [13]. The result showed poor accuracy and robustness compared to an indoor or urban environment, where those algorithms are designed for. Another study [14] comparing three visual SLAM algorithms in an orchard of fruit trees. The results showed that ORB-SLAM2 is the most accurate system with its loop closure ability.

Based on the above literature, there is a lack of comparison between SLAM methods in vineyards environment. Therefore, more research needs to be done in that field.

3 SLAM Algorithms

The SLAM systems used in this work can be classified into two main categories: visual SLAM and 3D lidar SLAM. This section briefly describes the four tested algorithms, RTAB-MAP and ORB-SLAM2 under visual SLAM, LIO-SAM, and StaticMapping under 3D lidar SLAM. Those systems have been chosen since they are the state-of-the-art, and the most popular algorithms by the time writing this paper. However, StaticMapping is not very popular yet. A brief list of sensors employed by each system is shown in Table 1 including the front-end back-end algorithms used by each method. The sensors that are used by each system in our setup are marked as bold text in the table.

3.1 RTAB-Map

Real-Time Appearance Based Mapping (RTAB-Map) is a graph-based SLAM approach [4] that supports input from RGB-D, Stereo, and lidar sensors. It combines two main algorithms which are *loop closure detector* and *graph optimizer*. The system uses the bag-of-words concept for loop closure detection by determining if the new image comes from a previously visited location or a new location; If the hypothesis of the new image is above a certain threshold, the new location will be added to the map as a new graph constraint. Then, in the background, the map graph is optimized to reduce the drift error in the overall map [15]. To achieve real-time performance for large scale environments, the system has a memory manager that limits and control the number of locations that are used for loop closure detection [4]. The system implements two standard odometry methods, Frame-To-Map (F2M) and Frame-To-Frame (F2F) using 3D visual features. In F2M, the system registers the new frame against upon local

Table 1: The SLAM systems used for benchmarking together with supported sensors, front end and back end algorithms.

System	**Sensors support**	**Front-end**	**Back-end**
LIO-SAM	**3D lidar + IMU**, GNSS (Optional)	ICP	GTSAM
Static-Mapping	**3D lidar**, IMU (Optional), GNSS (Optional)	ICP	GTSAM
ORB-SLAM2	Monocular, **RGB-D** or Stereo cameras	ORB features extraction. PnP RANSAC for motion estimation.	g2o
RTAB-MAP	**RGB-D** or Stereo camera (Mainsensors). 2D or 3D lidar (Optional to enhance map build from main input sensors)	Visual odom: GFTT/BRIEF for feature detectoin, NNDR for feature matching, PnP RANSAC for motion estimation. lidar odom: ICP	GTSAM (default) g2o, TORO

map, while the F2F registers the new frame to the last key-frame. RTAB-MAP can generate 2D and 3D occupancy grid map with dense point cloud, which is very useful for robotics applications. Furthermore, there is a full integration of this algorithm in Robot Operating System (ROS) as rtabmap_ros[2] package.

3.2 ORB-SLAM2

ORB-SLAM2 is a feature-based visual SLAM method that can create a sparse 3D map, which can be used with monocular, stereo, and RGB-D cameras to compute the camera trajectory. This algorithm has three main threads running in parallel, which enable real-time performance. The first thread is for tracking the camera pose in the new frames by finding feature matches in the local map. The second thread for local map management and optimization by applying local Bundle Adjustment (BA). The final thread is for loop closure detection and pose-graph optimization; this process is mainly for correcting the accumulated drift [7]. The system uses the bag-of-words DBoW2 concept [16] for place recognition, loop closure, and localization. However, when mapping a large-scale environment, the processing time of loop closure detection and graph optimization will increase as the map grows. This leads to a significant delay when making loop closure corrections after being detected. This system does not generate an occupancy grid map, which makes it difficult to use directly in real robotics applications [4]. ORB-SLAM2 does not have full integration with ROS, it only subscribes to the camera topics and there are no output topics. However, it offers a visualizer for the trajectory and a sparse point cloud.

[2] wiki.ros.org/rtabmap_ros

3.3 LIO-SAM

LIO-SAM is a factor graph tightly-coupled lidar inertial odometry via smoothing and mapping system [5]. The main input sensors to the system are 3D lidar and 9-axis IMU, but it can also use data from GNSS sensors for absolute measurement and map correction. LIO-SAM estimates lidar motion during the scan by using the raw IMU data. Then, for point cloud de-skewing, it assumes a nonlinear motion model. The novelty of this algorithm is that it uses the idea of keyframes and sliding windows from visual SLAM systems. In this way, the scan registration is performed at the local window scale instead of the global map improving the real-time performance significantly. On the other hand, the old scans are used for pose optimization. The IMU data is critical for this system to work properly. LIO-SAM is fully integrated into ROS. The generated output map and trajectory could be saved on a disk after finishing the scan.

3.4 StaticMapping

StaticMapping is a 3D lidar SLAM algorithm with optional IMU, odometry and GNSS inputs [6]. The back end of this algorithm uses M2DP [17] global descriptor for loop closure detection, and iSAM2 [18] for smoothing and mapping. This is an offline map-building algorithm from the recorded data.

4 Evaluation

This section describes our simulation environment, the different scenarios to test the algorithms, and the metrics we used for evaluation. Finally, we present and discuss the results.

4.1 Environment

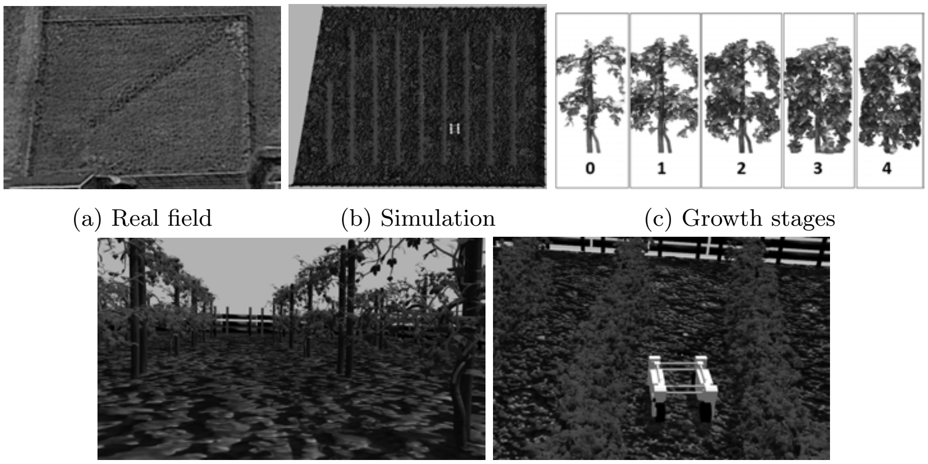
We created a digital twin of an actual vineyard located at the University of Lincoln Riseholme campus under ROS/Gazebo as shown in Figure 1. The virtual environment offers realistic uneven terrain, plus multiple growth stages of the vine plants and the crops. The vineyard has nine rows that are 18 m long with a 3 m distance between the rows. The mobile robot used is Thorvald produced by Saga Robotics[3] equipped with the following sensors: two 2D Hokuyo laser scanners, 3D lidar VLP-16 by Velodyne, ROS-IMU plugin and a Kinect V2 camera. A video of the environment can be seen in the link below[4].

4.2 Testing scenarios

We designed four different scenarios in which we evaluated the 4 different SLAM methods we aim to compare.

[3] sagarobotics.com/

[4] Field: youtu.be/L9ORZNyWdT0. Uneven terrain: youtu.be/L9ORZNyWdT0



(a) Real field (b) Simulation (c) Growth stages

(d) Robot front camera view (e) The final growth stage

Fig. 1: Digital twin of the vineyard environment.

- Scenario 0 (S0): Move in a straight line. That is mainly to evaluate the drift of the generated trajectory. Trajectory length is 25 m.
- Scenario 1 (S1): Send the robot to inspect a row and get back from the same row. Trajectory length is 48.6 m.
- Scenario 2 (S2): Send the robot to inspect a row and get back from the adjacent row. Trajectory length is 54.6 m.
- Scenario 3 (S3): Inspect multiple rows. That is to simulate a real-life inspection scenario with multiple loop closures. Trajectory length is 101.2 m.

4.3 SLAM algorithms configurations

As mentioned earlier, the tested SLAM systems have some integration with ROS, either partially like ORB-SLAM2, or fully as in LIO-SAM. The configuration file for each algorithm has been modified to work with our setup. (a) For ORB-SLAM2, we used the RGBD configuration and modify the camera parameters in the setting file based on the used camera. (b) In RTAB-MAP, the configurations are passed through the launch file. We used the `rgbd_sync` node to synchronize the RGB-D camera data before passing them to `rtabmap_ros` node. A laser scan data was passed to construct a 2D occupancy grid map. (c) StaticMapping is an offline system, it can only construct the map from recorded data. The default parameters were used and `accumulate_cloud_num` was set to 1. (d) For LIO-SAM we used the default configurations but disabled the GNSS optimization.

4.4 Metrics

The output trajectory of a SLAM system can be evaluated by finding the absolute distance between the estimated trajectory and the ground truth. The

Absolute Trajectory Error (ATE) is defined as the average deviation from the ground truth trajectory[19].

$$\text{ATE}_{rmse} = \left(\frac{1}{n} \sum_{i=1}^{n} \|\text{trans}(Q_x^{-1} SP_i)\|^2 \right)^{\frac{1}{2}}$$

, where i is the time sample or frame, SP_i is the spatial translation at time i, $trans$ is the translation error, and Q_x is the ground truth pose. To find the statistical metrics of ATE, we used the open-source library *evo*[5] to calculate the following metrics: Maximum, Mean, Median, Root Mean Square Error (RMSE), and Standard deviation (STD).

4.5 Results

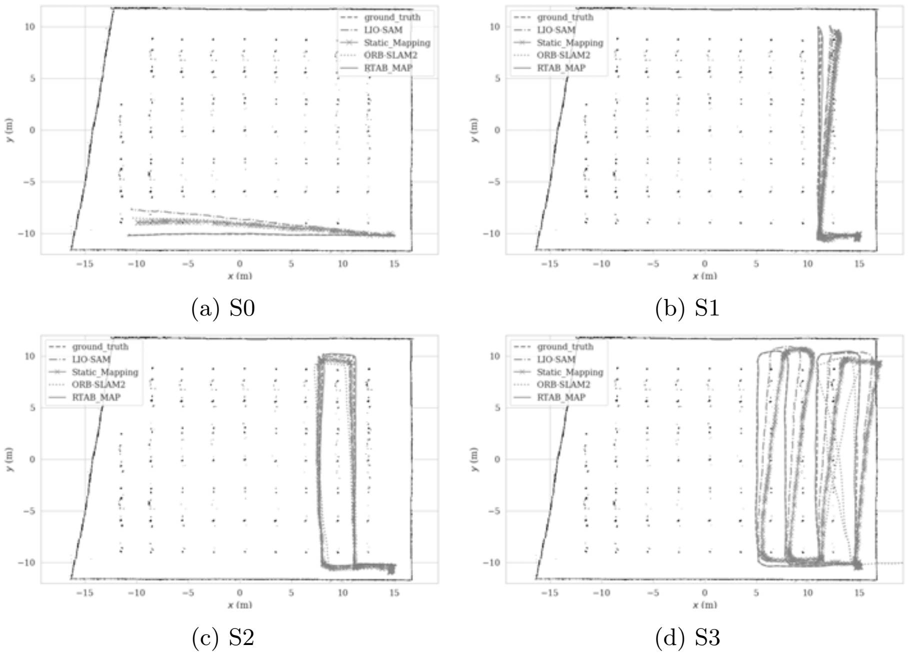
The error metrics of the experiments are summarized in Table 2. Figure 2 shows the output trajectories of the four algorithms compared to the ground truth. As it can be observed from Table 2 RTAB achieves superior performance with the lowest RMSE across all scenarios compared to other systems.

Table 2: ATE statistical metrics for various SLAM systems on 4 different scenarios

	S0				**S1**				**S2**				**S3**			
	LIO-SAM	STATIC	ORBSLAM2	RTAB-MAP	LIO-SAM	STATIC	ORBSLAM2	RTAB-MAP	LIO-SAM	STATIC	ORBSLAM2	RTAB-MAP	LIO-SAM	STATIC	ORBSLAM2	RTAB-MAP
MAX	1.65	1.55	1.62	**0.20**	1.32	1.85	1.42	**0.08**	0.83	0.79	1.34	**0.20**	0.77	2.64	8.95	**0.12**
MEAN	1.00	0.97	0.81	**0.08**	0.53	0.77	0.86	**0.05**	0.20	0.42	0.69	**0.12**	0.36	1.23	4.35	**0.07**
MEDIAN	1.11	1.03	0.81	**0.05**	0.40	0.29	1.04	**0.05**	0.18	0.40	0.78	**0.12**	0.34	0.98	3.91	**0.06**
RMSE	1.13	1.07	0.95	**0.09**	0.68	1.04	0.96	**0.05**	0.22	0.45	0.84	**0.14**	0.40	1.51	5.22	**0.07**
STD	0.53	0.46	0.50	**0.05**	0.43	0.70	0.42	**0.01**	0.10	0.17	0.47	**0.06**	0.19	0.88	2.90	**0.03**

In the first scenario S0, all SLAM methods except RTAB-MAP reported big drift in the estimated trajectory, with a maximum value of 1.65 m from LIO-SAM, while RTAB-MAP has the minimum drift with a maximum value of 0.20 m and RMSE of 0.09 m. For scenario S1, RTAB-MAP has the minimum drift and smallest RMSE with 0.05 m, the performance of the remaining algorithms were close. There is no loop closure in this scenario even though the robot got back to the same place, which is due to the difference in the camera viewpoint. In scenario S2, the second-best algorithm after RTAB-map is LIO-SAM, followed by StaticMapping, and finally ORB-SLAM2 with the largest RMSE of 0.84 m.

[5] github.com/MichaelGrupp/evo



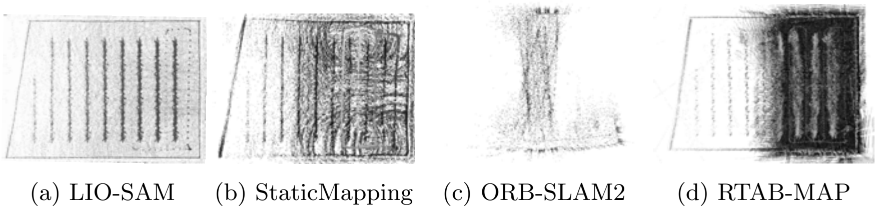
(a) S0

(b) S1

(c) S2

(d) S3

Fig. 2: Output trajectory from various SLAM systems in different test scenarios.



(a) LIO-SAM (b) StaticMapping (c) ORB-SLAM2 (d) RTAB-MAP

Fig. 3: Output maps and estimated trajectory from different SLAM system for scenarios 3

The final scenario is the most interesting and realistic one, where the robot traverses multiple rows before going back to its starting point. This scenario with multiple loop closures represents a very interesting benchmark for the algorithms, as represented in Figure 2. ORB-SLAM2 has failed to create a reliable map, we think that might be due to the repeated visual appearance of the environment; on the other hand, LIO-SAM and StaticMapping did not completely fail. However, the drift of the trajectory is too large to be used for robot navigation. This is because, in a vineyard, the distance between two rows is usually between 2 to 3 meters, in which if the mobile robot fails to localize itself accurately it can cause damage to the crops. The map generated for scenario 3 from the different systems is shown in Figure 3.

Even though both ORB-SLAM2 and RTAB-MAP are using the bag-of-words concept for loop closure detection, the pose-graph optimization is different. In ORB-SLAM2, global bundle adjustment is used for the pose optimization process after loop closure. So, since there are lots of visual features shared between the keyframes due to similar appearance, this algorithm fails to estimate reliable trajectory as represented in Figure 2d and Table 2. On the other hand, RTAB-MAP is much more robust to false loop closures, since it checks the transformation in the graph after the optimization process, if the translation variance was too large, the loop closure is rejected. The lidar methods did not suffer from repeated feature issue, due to the large field of view of the 3D lidar.

5 Conclusion

In this work, we compared four state-of-the-art visual and 3D lidar SLAM algorithms in a challenging simulated vineyard environment with uneven terrain. The main challenge for the visual SLAM system in such an environment is represented by a repeated pattern of appearance and less distinct features. This may result in false loop closures. The state of the art ORB-SLAM2 failed when the robot moved across multiple rows with multiple loop closures, we think this may be due to an identical visual appearance between vineyard rows. RTAB map is much more robust to invalid loop closure. The trajectory generated from the RTAB-MAP algorithm was the most accurate in our test scenarios. The estimated trajectories from LIO-SAM and StaticMapping suffer from big drift but the trajectories shape is acceptable, we believe in real-world tests, this drift could be fixed with the availability of GNSS signal. For future work, we will add some unique features within vineyard rows to test the reliability of ORB-SLAM2. Those unique features could be to have some variation in the plant models instead of having them identical across all the rows. In addition, we would like to test the localization ability within the generated maps from the SLAM systems. Finally, we will test those methods in a real vineyard.

Acknowledgement

This work has been supported by the European Commission as part of H2020 under grant number 871704 (BACCHUS).

References

1. Nørremark, M., Griepentrog, H., Nielsen, J., Søgaard, H.: The development and assessment of the accuracy of an autonomous GPS-based system for intra-row mechanical weed control in row crops. Biosystems Engineering **101**(4) (December 2008) 396–410
2. Imperoli, M., Potena, C., Nardi, D., Grisetti, G., Pretto, A.: An effective multi-cue positioning system for agricultural robotics. IEEE Robotics and Automation Letters **3**(4) (October 2018) 3685–3692

3. Aguiar, A.S., dos Santos, F.N., Cunha, J.B., Sobreira, H., Sousa, A.J.: Localization and mapping for robots in agriculture and forestry: A survey. Robotics **9**(4) (November 2020) 97
4. Labbé, M., Michaud, F.: Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. Journal of Field Robotics **36**(2) (2019) 416–446
5. Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Daniela, R.: Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE (2020) 5135–5142
6. Liu, E., atinfinity: Edwardliuyc/staticmapping: Release for doi (May 2021)
7. Mur-Artal, R., Tardos, J.D.: ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-d cameras. IEEE Transactions on Robotics **33**(5) (October 2017) 1255–1262
8. Filipenko, M., Afanasyev, I.: Comparison of various SLAM systems for mobile robot in an indoor environment. In: 2018 International Conference on Intelligent Systems (IS), IEEE (September 2018)
9. Yagfarov, R., Ivanou, M., Afanasyev, I.: Map comparison of lidar-based 2d SLAM algorithms using precise ground truth. In: 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE (November 2018)
10. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE (June 2012)
11. Cheein, F.A., Steiner, G., Paina, G.P., Carelli, R.: Optimized EIF-SLAM algorithm for precision agriculture mapping based on stems detection. Computers and Electronics in Agriculture **78**(2) (September 2011) 195–207
12. Comelli, R., Pire, T., Kofman, E.: Evaluation of visual slam algorithms on agricultural dataset. (September 2019)
13. Pire, T., Mujica, M., Civera, J., Kofman, E.: The rosario dataset: Multisensor data for localization and mapping in agricultural environments. The International Journal of Robotics Research **38**(6) (April 2019) 633–641
14. Capua, F.R., Sansoni, S., Moreyra, M.L.: Comparative analysis of visual-SLAM algorithms applied to fruit environments. In: 2018 Argentine Conference on Automatic Control (AADECA), IEEE (November 2018)
15. Altuntas, N., Uslu, E., Cakmak, F., Amasyali, M.F., Yavuz, S.: Comparison of 3-dimensional SLAM systems: RTAB-map vs. kintinuous. In: 2017 International Conference on Computer Science and Engineering (UBMK), IEEE (October 2017)
16. Galvez-López, D., Tardos, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Transactions on Robotics **28**(5) (October 2012) 1188–1197
17. He, L., Wang, X., Zhang, H.: M2dp: A novel 3d point cloud descriptor and its application in loop closure detection. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE (October 2016)
18. Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., Dellaert, F.: iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering. In: 2011 IEEE International Conference on Robotics and Automation, IEEE (May 2011)
19. Prokhorov, D., Zhukov, D., Barinova, O., Anton, K., Vorontsova, A.: Measuring robustness of visual SLAM. In: 2019 16th International Conference on Machine Vision Applications (MVA), IEEE (May 2019)