Evaluation of an Inertial and Optical Sensors Based Mapping and Localization System

Stefan Hensel[1], Marin B. Marinov[2](✉), and Max Schmitt[1]

[1] Department for Electrical Engineering, University of Applied Sciences Offenburg,
Badstraße 24, 77652 Offenburg, Germany
[2] Department of Electronics, Technical University of Sofia,
8, Kliment Ohridski Blvd., 1756 Sofia, Bulgaria
`mbm@tu-sofia.bg`

Abstract. , The visual-inertial mapping and localization system *maplab* is analyzed by its implementation and subsequent evaluation. The mapping or localization is based on environmental feature detection. In addition to creating maps, there is also the option of fusion of several maps and thus mapping extensive areas and using them for further analysis of data. In this way, various software tools can be used to optimize the existing data sets.

Two sensor components are needed: an inertial measuring unit (IMU) and a monochrome camera, which are combined by a hardware rig and put into operation for the analysis of the visual-inertial system. System calibration is crucial for precision and system functioning and is based on nonlinear dynamic state estimation. This ensures the best possible estimate of the position of the environmental feature and the map. *Maplab* is particularly suitable for mapping rooms or small building complexes as the implementation and evaluation of the results in different application scenarios show. Special emphasis is laid on the evaluation of larger scenarios, in which is shown, that the system is struggling to keep up geometric consistencies and thus provide an accurate map.

Keywords: maplab · Robot Operating System (ROS) · Simultaneous Localization and Mapping (SLAM)

1 Introduction

Robot-assisted system developments and the improved performance of algorithms in the software area made it possible to use unmanned vehicles in almost any sphere of human activities. The process of development leads to a broad spectrum of possible autonomous systems with the ability to operate unassisted. As a result, there is the possibility of applications into areas that cannot be reached by man.

It is important to have an accurate visual idea of the surroundings for navigation in unstudied areas. This requires image processing and interpretation algorithms that create a digital image from the real world. In this way, it is possible to use the available

D. Perakovic and L. Knapcikova (Eds.): FABULOUS 2022, LNICST 445, pp. 3–15, 2022.
https://doi.org/10.1007/978-3-031-15101-9_1

environmental information later. Current developments in the field of robotics allow for the effective design of efficient mapping and localization systems.

This work evaluates the effectiveness of the visual-inertial mapping and localization system maplab. Procedures for improving the system in the processes of mapping and localization are also studied.

ETH Zurich has provided the main frame lab as a basis for further research projects in the area of navigation of mobile robot platforms [1]. A monochrome camera, and an inertial measurement unit (IMU), with the necessary software packages, are available for evaluation. These two components comprise a system and its analysis and the evaluation of the results are used to obtain an assessment as a so-called SLAM (Simultaneous Localization and Mapping) system.

A mobile robot platform is available to test possible maplab applications, which is additionally equipped with a reference mapping system based on 3D laser point clouds.

2 State of the Art and Related Work

Robot mapping and localization have been studied for several years beginning with the seminal papers of [2]. After tackling the problem with tools of dynamic state estimation (see e.g. [3]), a paradigm shift to graph-based SLAM for online and offline processing could be observed [4, 5]. A sub-field of the area is the use of optical systems to solve the SLAM problem, pointing to named visual SLAM [6, 7].

As camera systems and computer vision show bottlenecks due to the amount of processed data, the slower processing is compensated by the incorporation of inertial sensors to improve the accuracy of the orientation estimation. This is basically a sensor fusion process resulting in so-called visual-inertial SLAM approaches [8, 9].

Recent work in the was focusing on the improvement of computational speed [10] and the handling of large-scale maps, e.g. for autonomous driving.

This contribution evaluates the maplab framework described in [11]. The system is a combination of a hardware visual-inertial setup and a given software framework. Emphasis was set on designing a cheap and mobile device, as described in previous work [12] and adopting the maplab system to the sensor system. Another contribution is the comparison and verification of the setup by another state-of-the-art localization framework, that uses expensive sensors in form of 3D LiDARs to localize. A thorough evaluation was done for certain scenarios and setups, assessing usability, precision, and needed resources for the system.

The rest of the paper is structured as follows: Section 3 describes the system setup with a focus on the software parts, to differentiate from previous publications in [13].

In Sect. 4, a description of the researched scenarios is given, explaining choice and setup. Section 5 is the main part of the publication, giving a detailed explanation and insight into the system evaluation before Sect. 6 concludes the paper.

3 System Setup and Components

3.1 The Maplab Framework

The maplab framework is a visual-inertial SLAM system. A variety of tools can be used to solve the SLAM problem. These include [13]:

(i) Creation and localization in maps, (ii) Multi-session map combination, (iii) Loop closure detection, (iv) Deep reconstruction, (v) Visualization of maps.

3.2 Robot Operating System

A robot operating system (ROS) is available to ensure communication and exchange of data between the camera or IMU and the computer. The widespread and frequent use of this system in robot applications formed a huge collection of software tools and libraries [14, 15]. These allow the user to install simply and easily the necessary drivers for various sensors. A so-called *rosbag* is a file format in ROS for data storing [16].

3.3 ROVIO and ROVIOLI

The *maplab* system uses the ROVIO (RObust VIsual Odometry) framework [17]. The extension with a localization module leads to ROVIOLI (RObust VIsual Odometry with Localization Integration). ROVIO is, so to speak, the basic building block for the detection and tracking of environmental features.

All the necessary data for the lower-level Kalman filter must be available within a certain time for ROVIOLI to correctly detect and track environmental features. According to the selected data rates of the electronic components, this "time window" is $t = 50\,\mathrm{ms}$.

It is possible to operate the camera with automatic exposure time and to map or position it in an environment despite possible strong exposure changes with the help of specific extensions of the camera driver. The schematic structure of the visual-inertial system is shown in Fig. 1. [12] gives details on the hardware implementation and calibration of the Camera-IMU system.

3.4 Mobile Platform Hardware

As sensor carrier is used the Husky mobile robot platform from Clearpath Robotics. It is a mobile base that can be equipped with different sensors, controllers, and batteries for power supply due to its payload of 75 kg. Its compact dimensions, wheelbase, wheel height, and skid-steer drive allow for it to be used both outdoors and indoors and make it possible to evaluate mapping for structured and natural environments. The maximum speed is 1 m/s, and the runtime of one battery charge is about 3 h.

4 Creating Scenarios for Evaluation

To achieve the best possible quality of evaluation this work adheres to the basic structures of the standards for evaluation of Gesellschaft für Evaluation e.V. (DGEval for short) [18]. In these standards, it is particularly emphasized that all available information must be presented, which is essential for a high-quality and correct evaluation. Extensive analyses are performed and evaluated for this process to cover the widest possible spectrum of use cases. This section presents different scenarios and discusses the experience gained with maplab in this work. The investigations and evaluations of the results should answer the following questions:

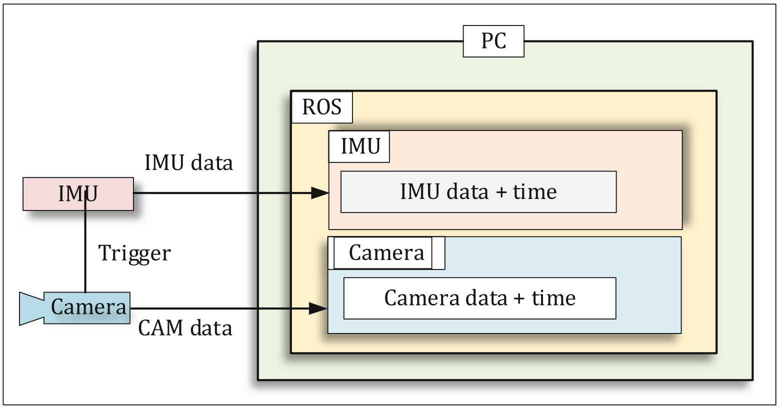

Fig. 1. Construction of the hardware of the visual-inertial system.

- Is maplab suitable as a new SLAM method in addition to existing mapping systems?
- Does maplab have sufficient mapping and localization precision?
- What is the performance of maplab inside and outside of buildings?
- How well does the optimization algorithm work?
- Is the framework suitable to equip mobile robot platforms with it and localize their position?
- What possible applications are there, considering the evaluation results?
- What hardware capabilities are required for the calculations?
- Is there a way to continue the research in the SLAM field with maplab?

4.1 Approach for Creating Evaluation Scenarios

Experience with the visual-inertial system has been gained by creating multiple rosbags that record camera and IMU data. While the use of the IMU is simple and straightforward, the electrical properties of the camera present difficulties.

This should be taken into account when working with this component. A particular problem here is the sensitivity of the camera to the strong light incidence in the lens.

If there are significant contrasts in lighting incidents in a scene, the camera software ensures the best possible representation of the bright surroundings. The resulting increase in exposure time means that environmental characteristics are no longer visible in dimly lit environments. However, these play an essential role in mapping and localization with maplab. The resulting lack of environmental features leads to a loss of information in certain areas. The result is interruptions in the continuity of the map.

For effective loop detection and the merging of several maps, it is also necessary to examine scenarios with identical starting and endpoints.

4.2 Presentation of the Created Scenarios

The following scenarios have been selected for high-quality evaluation, which allows for the use of most of the tools available in maplab.

a) Analysis and evaluation of the loop closure detection in maplab
The aim is to determine the performance of maplab in terms of mapping and recognition of already known environmental features and of increasing the geometric consistency of the maps.
b) Creation of the so-called multi-session maps
The aim is to determine the performance of the algorithm for merging multiple maps.
c) Investigation of the localization ability using a mobile robot platform
The aim here is to determine the precision of the positioning of a mobile robot for solving the SLAM problem.

The choice of scenarios is intended to cover as many application cases as possible and to study the main features of maplab (see [1]). Maplab offers the option of displaying the route covered during mapping. In addition to showing the distance in a single mapping process, it is also possible to show the sum of the distance in the case of a multi-session. Results of the evaluation of the accuracy in measuring distances are given in [19].

5 Evaluation of Mapping Performance

5.1 Analysis and Evaluation of Loop Closure Detection in Maplab

With loop detection, the recognition of already known environmental features should be guaranteed and thus an increase in the geometric consistency of the entire map should be achieved.

To determine the maximum performance of this ability, a distance traversed several times is recorded with the visual-inertial system. As soon as a certain number of loops is reached, the recording of the data at the starting point is stopped. The existing, changing difference between the starting point and the identical endpoint should now be minimized and displayed by carrying out the loop closure detection several times. Based on this analysis, a statement can then be made about the performance of the recognition of already known environmental features in maplab.

A laboratory room at the Offenburg University of Applied Sciences is used as the environment to be mapped. Starting from a starting point, the room is traversed several times in a circle and thus areas that have already been detected are entered several times. This process is finally terminated at the former starting point. The length of the route included during mapping is about 150 m. The starting point has the coordinates

$$s_0 = \left[x_0\, y_0\, z_0\right]^T = [0\,0\,0]^T. \tag{1}$$

After the area has been mapped with maplab, the coordinates of the last data point can be determined. This has the coordinates:

$$p_0 = \left[p_{x,0}\, p_{y,0}\, p_{z,0}\right]^T = [-1.0525\; 1.0042\; -\; 0.2228]^T. \tag{2}$$

The distance between the start and endpoint can be determined using Eq. (2).

$$d_{3d,n} = \sqrt{\left(p_{x,n} - x_0\right)^2 + \left(p_{y,n} - 0\right)^2 + \left(p_{z,n} - z_0\right)^2}. \tag{3}$$

Here n indicates the number of loop closure detection executions. If you now calculate the distance from the point p_0 to s_0 using (2) before loop closure detection is carried out, this results in a value of $d_{3d,n} = 1.4717\,\text{m}$.

After applying the algorithm for finding loop closure one gets a new coordinate for the endpoint caused by the loop detection. This endpoint p_1 has the form:

$$p_1 = \left[p_{x,1}\, p_{y,1}\, p_{z,1}\right]^T = [-1.1518\ 0.0991 - 0.0076]^T \tag{4}$$

Here, too, the difference to the origin of the map s_0 can be calculated. The result is a value for $d_{3d,1}$, which is presented in Table 1 for reasons of clarity.

A significant minimization of the distance can already be seen by looking at the coordinates of the point p_1. It is preferable to perform the optimization and loop detection process a second time. This again results in a new endpoint coordinate p_2 with the values:

$$p_2 = \left[p_{x,2}\, p_{y,2}\, p_{z,2}\right]^T = [0.0001 - 0.0014\ 0.0001]^T \tag{5}$$

The final distance between s_0 and p_2 can be seen in Table 1.

Table 1. Illustration of the effects of loop closure detection.

Number of runs n of an LCD	0	1	2
Distance $d_{3d,n}$, m	1.4717	0.0181	0.0017

In this analysis, the good performance of the loop detection can be seen very well. From a former initial distance between the starting point and the identical endpoint of 1.4717 m, this is reduced significantly to an absolute error of 1.7 mm.

In this way, the algorithms of the loop closure detection and the optimization process can be evaluated. It turns out that a significant improvement or minimization of distance $d_{3d,n}$ can be achieved with an increasing number of loop detections and with the implementation of the optimization process. In addition to the recognition of already known features, this increases geometric consistency, as already stated. However, it should be mentioned that this analysis takes place under ideal conditions.

By repeatedly entering already known scenes, a particularly large collection of descriptors is available, which can then be compared for matches in loop closure detection. As a result, the position of the environmental features and the location related to them can be estimated with considerably greater precision.

5.2 Ecording Multiple Multi-sessions

One of maplab's most distinctive capabilities is map fusion. This option allows for mapping large areas. Changing perspectives or external environmental influences such as weather or exposure changes should not play a role in the alignment of all maps to each other [1].

Mapping of a Laboratory Room

The result of merging maps in a small room is examined by recording four separate sessions in a laboratory room. In this application, the focus should be on the quality and precision of the merging algorithm. The general presentation of the inventory of the laboratory, i.e., the arrangement of the laboratory tables or the depiction of the measuring devices, is decisive for the evaluation of the resulting map. To ensure that the starting point remains consistent in all individual recordings, a simple template was created on which the entire camera IMU system can be placed and aligned almost identically for all sessions. To detect as many environmental features as possible at the beginning of each session, the starting point is chosen between two laboratory tables, with the visual-inertial system aligned in the middle of the room. At the same time, the starting point will be the beginning of the necessary start and end sequence, which ends after about 6 m.

Each of the four individual recordings includes a different part of the laboratory. The laboratory in which the mapping is performed is shown in Fig. 2. As can be seen in Fig. 2, there is a window in front on the left that floods the room with daylight. Here, as explained, attempts are made to keep the strongest light source behind the visual-inertial system.

Opposite the window front is the corridor of the university building, which is spatially separated by a glass front. To avoid the influence of scattered light, the entire camera IMU system is preferably aligned towards the center of the room.

In contrast to the usual approach of mapping by recording *rosbags*, in this use case mapping is performed using the *rostopics* of the camera and IMU. Thus, instead of evaluating the visual and inertial data afterward, the detected features are immediately mapped. The advantage of this procedure is the immediate display of the extraction of visual-inertial information from the environment because of the detection of environmental features. This also makes it possible to recognize the current exposure in the picture. This enables, on the one hand, focusing on features that are still to be detected and, on the other hand, reacting to strong light sources. This leads to an increase in the information content in the recording.

Fig. 2. Photo of the laboratory room.

However, the only disadvantage of mapping using rostopics is that the recordings are not available as rosbags. For this reason, subsequent reevaluation of these visual and inertial data will no longer be possible.

First, the result is presented, which is obtained when all four maps are loaded into the maplab console. The maps were neither aligned to each other nor were an optimization process carried out.

In Fig. 3, in addition to the individual maps, which are shown in different colors, the course of the mapping process can be seen in a blue line. To illustrate the performance of the fusion algorithm and its accuracy, the dartboard on the left of the image is used as the object. Due to its coloring, this has a high contrast with very low reflection.

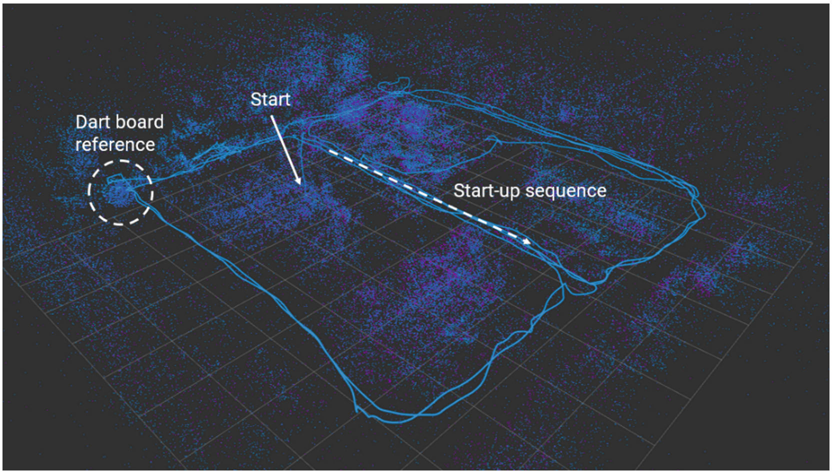

Fig. 3. Result of the mapping of the measurement and sensor technology laboratory in an unoptimized and unaligned condition.

A side view of this dartboard can be seen in Fig. 4.

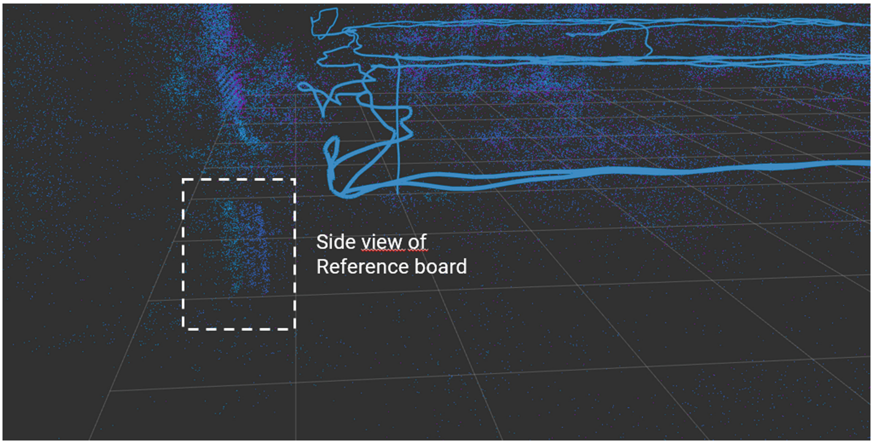

Fig. 4. Side view of the dartboard in an unoptimized and not aligned condition.

If you look at the two preliminary results in Fig. 3 and Fig. 4 of the fusion, there is a slight offset. Since both Fig. 2 and Fig. 3 were taken from a similar perspective, conclusions can be drawn about the general performance of maplab.

For the evaluation of the fusion algorithm, the optimization process is now carried out to optimize the display of the combined map and to obtain only those features that are most meaningful for a possible localization. Special mention should be made of the importance of the keyframing command, which significantly reduces the complexity of the map. The result of the mapping of the laboratory after the fusion is presented in Fig. 5.

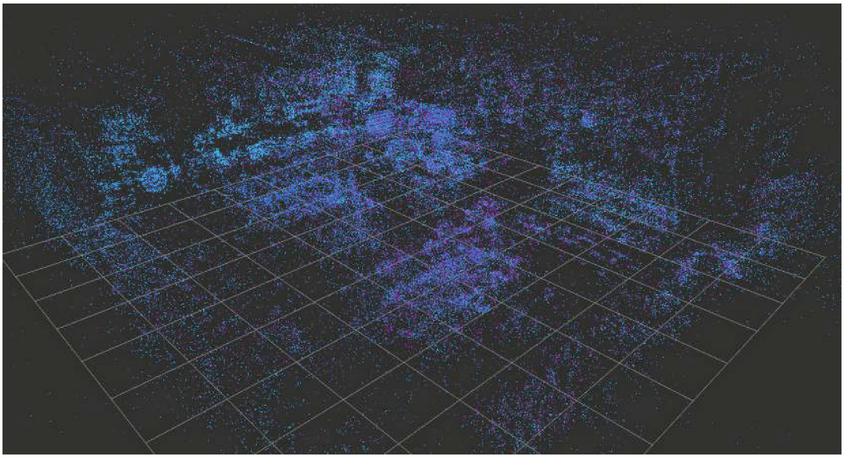

Fig. 5. Result of the mapping of the laboratory after the fusion.

If the two maps in Fig. 3 and Fig. 5 are compared, a slight improvement in the representation of the individual features can be seen. To clarify this, the dartboard is again used as a reference.

If the two recordings of the dartboard from Fig. 4 and Fig. 6 are compared, a clear minimization of the offset can be seen, which results in a more accurate mapping result. Incidentally, it is even possible to guess the position of the darts (see marking in Fig. 6).

Looking at the results of the map fusion and the effects of the optimization process, maplab potential for mapping environments can be predicted. In addition to mapping closed spaces, there is the possibility of detecting and mapping the inventory and, relatively small objects.

5.3 Evaluation of Maplab Localization Capability

One of the crucial tasks for solving the SLAM problem is the localization of a mobile robot. Positioning relative to a map is to take place using a current or already carried out mapping of the environment. The robot platform “Husky” is used to evaluate the performance of the localization algorithm. A rosbag containing all visual and inertial data is also recorded here for data evaluation. What is special about this route is the

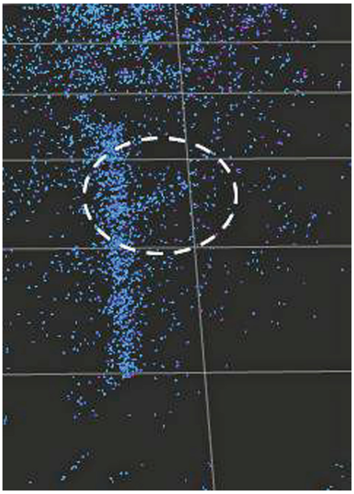

Fig. 6. Side view of the dartboard after the fusion.

high number of features already detected, which has a high information density due to repeated mapping. As a result of this large data collection, a significant increase in localization capability for the robot platform should be expected.

To carry out the localization, a map is created that only contains the ground floor of the building complex of the Offenburg University of Applied Sciences. It turns out that reducing the number of decks increases the fusion precision. This resulting map is used as a reference against which localization is performed. The map is shown in Fig. 7. In addition to the starting point, the associated route of the robot is also entered there.

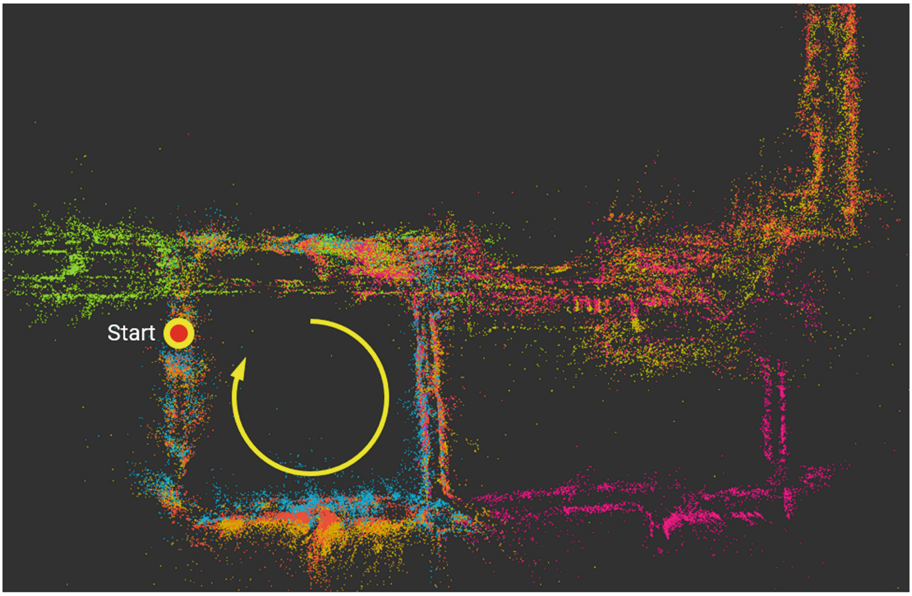

Fig. 7. Result of the mapping using the fusion of ground floor data.

To test the performance of the localization algorithm, the starting point of the recording is changed. It is possible to display positively detected features of the reference map during the extraction of the environmental features from the rosbag with the help of ROVIOLI. The procedure here is identical to that for loop closure detection. If a feature is successfully recognized, the robot is positioned relative to the known map. This is represented by a red dot. To clarify the route of the robot, the colors of the point clouds in Fig. 7 are unified to increase the contrast. In Fig. 8 the predicted route is shown based on the recognized environmental features.

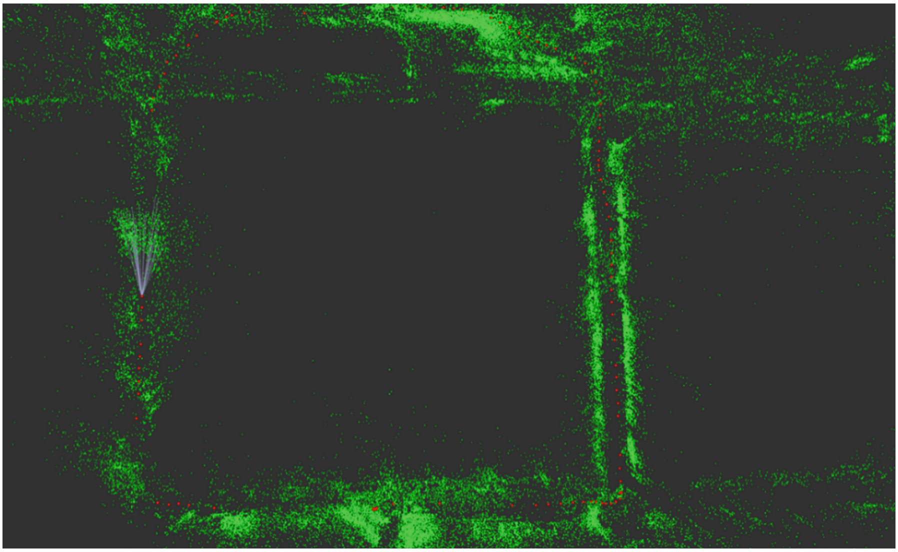

Fig. 8. Positioning of the robot relative to the ground floor map.

It is also possible to display the feature vectors recognized by ROVIOLI at runtime relative to the current position (see Fig. 9).

To perform a preliminary assessment of the possibility of localization using maplab, a surprisingly positive result can be achieved based on the investigations carried out. In addition to the retrieval and exact positioning relative to the ground floor map and the possibility of displaying positively detected features, a good localization capability is guaranteed. This is increased by the possibility of using a “live source”, whereby positioning at runtime is conceivable. This contributes significantly to solving the SLAM problem.

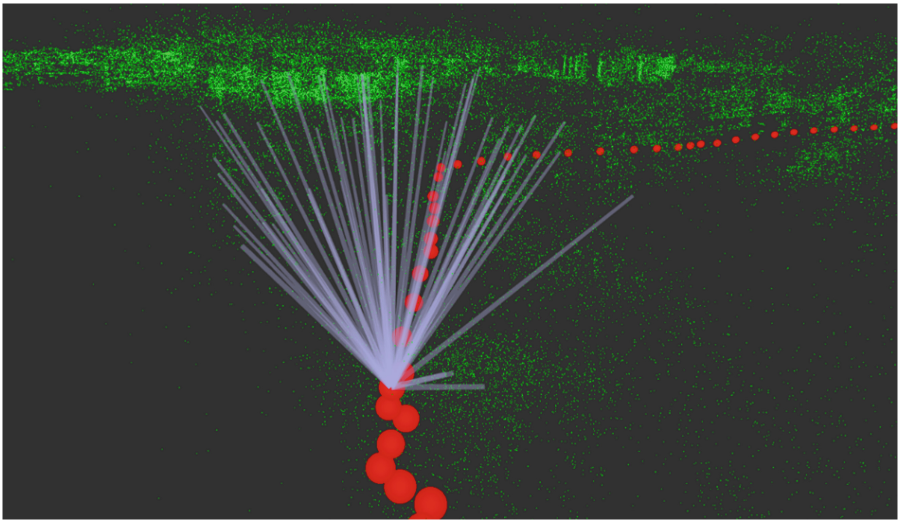

Fig. 9. Representation of the vectors of recognized features relative to the current position.

6 Conclusions and Future Work

To conclude the performance of maplab in terms of mapping and localization, the paper presents several possible uses and the results obtained.

Looking at the results of the individual mapping [19] with the fusion map in Subsect. 5.2 there are significant differences in accuracy and geometric consistency. With an increasing number of maps or area inspections, the density of features within the resulting map can be increased many times over, but at the same time, the risk of producing geometric inconsistencies is increased. However, if you look at the result of the mapping in Subsect. 5.3, this is mapped according to reality with certain errors. A reduction in the number of maps for fusion can lead to an increase in geometric consistency.

Based on the surprisingly positive outdoor mapping the statement can be made that maplab is not only suitable for use in building complexes but also for mapping areas outside such premises.

With further research projects and extensions of the current maplab software, it would be conceivable to equip mobile robot platforms or even unmanned aerial vehicles (UAV) with this SLAM system.

As a result, and based on the experience gained in dealing with maplab, there are a variety of application scenarios. One conceivable possibility would be the automated following of pre-programmed trajectories in buildings.

Acknowledgments. This research is supported by the Bulgarian National Science Fund in the scope of the project “Exploration the application of statistics and machine learning in electronics” under contract number КП-06-Н42/1.

References

1. Schneider, T., et al: Maplab: an open framework for research in visual-inertial mapping and localization, CoRR, vol. abs/1711.10250, 2017
2. Durrant-Whyte, H., Bailey, T.: Simultaneous localisation and mapping (SLAM): Part I the essential algorithms. IEEE Robot. Autom. Mag. 1–9 (2006)
3. Dissanayake, G., Durrant-Whyte, H., Scheding, S.: An experiment in autonomous navigation of an underground mining vehicle. IEEE Trans. Robot. Autom. **15**, 85–95 (2000)
4. Grisetti, G., Kümmerle, R., Stachniss, C., Burgard, W.: A tutorial on graph-based SLAM. IEEE Intell. Transp. Syst. Mag. **2**, 31–43 (2010)
5. Hess, W., Kohler, D., Rapp, H., Andor, D.; Real-time loop closure in 2D LIDAR SLAM. In: IEEE International Conference on Robotics and Automation (ICRA) (2016)
6. Nüchter, A.: 3D Robotic Mapping - The SLAM Problem with Six Degrees of Freedom, Springer, Heidelberg (2009)
7. Cadena, C., et al.: Past, Present, and future of simultaneous localization and mapping: towards the robust-perception age. IEEE Trans. Robot. **32**(6), 1309–1332 (2016)
8. Weiss, C., Achtelik, W.M., Siegwart, R.: Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments. In: IEEE International Conference on Robotics and Automation (ICRA) (2012)
9. Bloesch, M., Omari, S., Hutter, M., Siegwart, R.: Robust visual inertial odometry using a direct EKF-based approach. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2015)
10. Zhang, J., Singh, S.: LOAM: lidar odometry and mapping in real-time. In: Robotics: Science and Systems Conference, Berkeley (2014)
11. Schneider, T., Dymczyk, M., Fehr, M., Egger, K., Siegwart, R.: Maplab: an open framework for research in visual-inertial mapping and localization. IEEE Robot. Autom. Lett. **3**, 1418–1425 (2018)
12. Hensel, S., Marinov, M.B., Schmitt, M.: System setup for synchronized visual-inertial localization and mapping. In: 2020 XXIX International Scientific Conference Electronics (ET), Sozopol, Bulgaria (2020)
13. Corke, P.: Robotics, Vision and Control. Springer, Berlin (2011)
14. Blasdel et al.: About ROS.Version (2020). https://www.ros.org/about-ros/
15. Quigley, M., et al.: ROS: an open-source robot operating system. In: ICRA Workshop on Open Source Software (2009)
16. Bjelonic, M.: YOLO ROS: real-time object detection for ROS, 2016–2020. https://github.com/leggedrobotics/darknet_ros
17. Bloesch, M., Omari, S., Hutter, M., Siegwart, R.: Robust visual-inertial odometry using a direct EKF-based approach. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–3 October 2015
18. Deutsche Gesellschaft für Evaluation e.V., Standards für Evaluationen, Köln: Redaktion: Dr. Wolfgang Beywl, Zimmermann-Medien (2002)
19. Hensel, S., Marinov, M.B., Schmitt, M.: Experimental setup for investigation and evaluation of a mapping and localization system. In: Proceedings of the 9th FDIBA Conference - Challenges of the Digital World, Sofia, Bulgaria, 28–29 November 2019