i-SAIRAS 2024

# SLAM for SCOUT: A ROS2-Based Multi-Sensor System for Compliant Rovers

[1]*Koch, Johannes Frithjof, [1]Schütt, Manuel, [2]Ryll, Markus

*lead presenter

[1]j.koch@dlr.de, German Aerospace Center, Germany

[2]Technical University of Munich, Germany

## ABSTRACT

This paper investigates a ROS2-based Simultaneous Localization and Mapping (SLAM) system for a ground-based multi-body planetary cave rover. Developed within the German Aerospace Center (DLR), the Space Cave explOration UniT (SCOUT) aims to autonomously explore volcanic cave systems on the Moon and Mars. A robust SLAM system for accurate environmental perception and mapping is therefore a crucial prerequisite for mission success. The presented system is based on visual perception, with data generated by two opposing RGB-D cameras, mounted inside the front and rear modules of the rover. Each of them is accompanied by an efficient LED matrix lighting system specifically designed for the mission environment. The unique architecture of the SCOUT rover, which consists of three flexibly connected aluminum modules with two rimless wheels each, presents novel challenges for SLAM deployment. In particular, its compliant structure, coupled with erratic motion during travel due to the wheel geometry and unknown surface conditions, makes accurate mapping difficult. In this context, the compliant connection between the two cameras requires an additional pose estimation system, based on multiple IMU sensors. To evaluate the developed SLAM system, a portable demonstrator with SCOUT-like characteristics was built. Tests conducted on this platform confirm the system's ability to produce high quality and accurate maps of a karst cave system. This work provides further insight into the specific requirements and challenges of a SLAM system for the SCOUT rover. It also demonstrates the system's effectiveness in mapping a cave system and provides valuable guidance for the future development of autonomous ground-based cave rovers.

## 1 INTRODUCTION

The exploration of underground structures on celestial bodies such as Moon and Mars, is of great importance for future crewed missions. In this context, especially lava tubes have been in the focus of research for the last 50 years [1]. These cave systems are remnants of volcanic activity and can be found on both Moon and Mars. Their long, tubular, and level shape make them suited for human habitation, providing radiation shielding and access to underground recourses. Recently published studies using radar data from various satellite missions show evidence for the existence of these cave systems on Moon and Mars [2]. Satellite images, such as in Figure 1, also show pits, which were suspected to be collapsed cave structures, also called skylights. However, it was unclear if they could provide easy access to lava tubes. One of the

most prominent examples for a skylight is the lunar Mare Tranquillitatis pit (MTP). Earlier this year, Carrer et. al [3] were able to collect evidence for an accessible cave conduit in the MTP, using data from the Lunar Reconnaissance Orbiter, proving the existence of accessible lava tubes on the Moon.

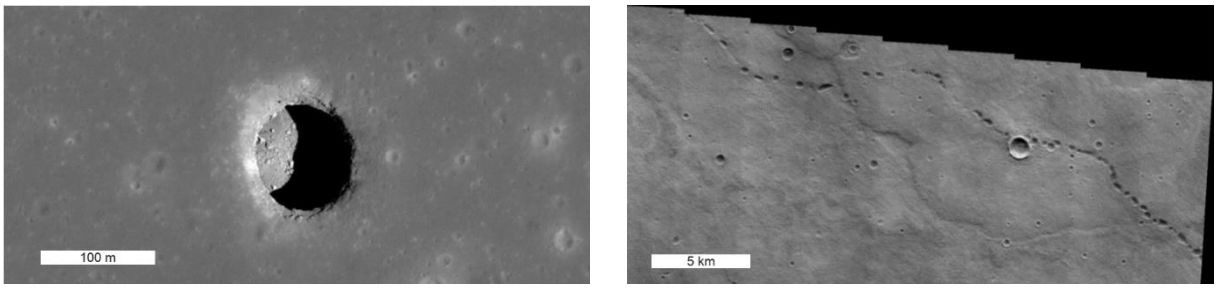


**Figure 1:** Lunar Mare Tranquillitatis pit (left) and line-like structures and pits on the Martian surface (right)


In order to explore these celestial lava tubes, the German Aerospace Center (DLR) is currently developing the Space Cave explOration UniT (SCOUT) – an all-terrain and robust rover, designed specifically for this environment. It consists of three flexibly connected aluminum modules with two rimless wheels each, as shown in Figure 2. This modular design allows for enhanced maneuverability and adaptability in navigating the uneven and challenging terrain of lava tubes, ensuring that SCOUT can effectively explore and map these extraterrestrial environments. [4]

**Figure 2:** SCOUT rover traversing rocky terrain (left) and rendering of one of the outer modules with integrated lighting system and RGB-D camera (right)

The scouting time inside the cave is limited by the rover's battery runtime of up to 10 h [4]. In the case of unexpected signal loss and with the significant time delay for a mission on Mars in mind, SCOUT has to operate fully autonomously. As autonomy is a prerequisite for mission success, a robust and precise Simultaneous Localization and Mapping (SLAM) system has to be implemented locally on the rover's processing unit. Therefore, SCOUT is equipped with two RGB-D cameras, which are mounted at the opposing ends of the rover. These Intel RealSense D435i cameras not only capture depth information but also provide an RGB color video stream. Additionally, a custom designed powerful LED matrix lighting system ensures uniform and bright illumination within the cave structure, facilitating visual data extraction [5]. The lighting system consists of two carrier boards, each equipped with five LEDs. For optimal lighting conditions, the RGB-D camera is positioned between these carrier boards. To increase

the robustness of the rover, all components are tightly integrated into the aluminum modules, as can be seen in Figure 2.

To achieve the objectives of this mission, Koch [6] developed and tested a visual SLAM system for the SCOUT rover as part of his master's thesis. The methods and findings presented in this paper are derived from the extensive testing and analysis conducted during his thesis work.

The initial step in the development process is to select an appropriate SLAM algorithm to build upon. This selection is guided by the following criteria:

- Excellent performance in reliable benchmarking tests
- Active and ongoing development
- Availability under an open-source license
- Ease of integration

These requirements result in the choice of RTAB-Map (v0.21.2) as the visual RGB-D SLAM algorithm for the SCOUT rover. Developed by Labbé and Michaud [7] from the University of Sherbrooke, Real-Time Appearance-Based Mapping (RTAB-Map) is a graph-based SLAM algorithm, tightly integrated into the ROS2 middleware [8]. As the SCOUT rover software is centered around ROS2 Humble, a SLAM algorithm with direct integration was preferred. Being a visual SLAM algorithm, RTAB-Map features out-of-the-box integration of RGB-D cameras through respective ROS topics. Furthermore, the algorithm is split into two nodes: one for odometry and another for the actual SLAM. If desired, this allows the user to replace the odometry part altogether, enabling high architectural flexibility in the system design. Additionally, integration of other sensors, such as Inertial Measurement Units (IMUs) or lidars, is also possible. [7]

One of the key features of the chosen SLAM algorithm is its customizability. Depending on the build instructions, it can support many feature types, odometry strategies, and also brings support for the integration of several RGB-D cameras at once. The results presented below were generated using "ORB-OCTREE" features, a Frame-to-Map (F2M) odometry approach, a detection rate of 3 Hz for the mapping, and a limited minimum feature depth of 0.3 m in accordance with the RealSense camera specs.

## 2 DUAL CAMERA SETUP

For better manual testing of the developed SLAM system on the rover hardware during early test stages, a portable demonstrator shown in Figure 3 was designed. It consists of three flexibly connected aluminum profile segments, imitating the rover's mechanical characteristics. Equally, the cameras are mounted in the same relative positions. Additionally, each demonstrator segment can be replaced by a corresponding SCOUT module without modifications, allowing easy integration of new testing hardware. Furthermore, the demonstrator can constrain link movements around the pitch axis, creating a near-rigid connection between cameras for a simplified SLAM algorithm development.

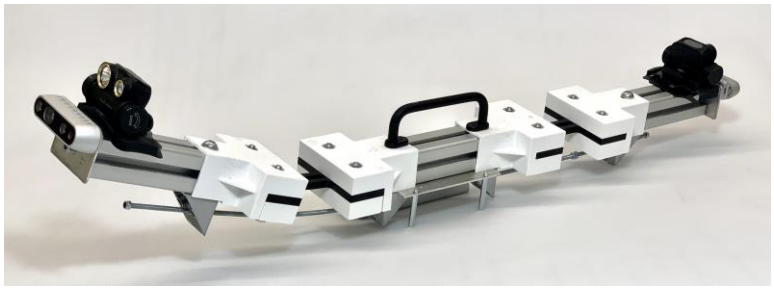

**Figure 3:** Portable demonstrator with fixation system, two mounted Intel RealSense D435i cameras, and mounted headlamps [6]

To avoid estimation of the rover state at all times, consideration was given to run two single-camera SLAM algorithms in parallel and merge their respective maps in a post-processing step. Besides the increased processing power needed, this idea was quickly deemed unfeasible due to the lack of visual overlap between the cameras, their opposite viewing angles, and divergent drift of the two generated maps, making it very difficult to fuse them. Therefore, the focus was placed on the development of a system that merges the visual information of both cameras into a single SLAM algorithm. However, a necessity for this system design to function is to determine the relative position between the cameras as accurately as possible.

To represent the current rover state, a robot model was implemented using ROS2 base tools such as the `robot_state_publisher` package and the `tf2` transform library [9], approximating all flexible segments as pitch and roll joints. The joint states were computed by fusing accelerometer and gyroscopic measurements through a Madgwick filter [10], implemented with the `imu_tools` package [11]. Although the resulting orientation data shows minimal drift during motion, the accuracy of the SLAM algorithm heavily relies on precise orientation information of the cameras relative to each other. To verify the estimated angles, a test was conducted to quantify the accuracy of the integrated IMU of the RealSense D435i camera. In a stationary state, the results showed a standard deviation of 0.05°. Due to a less precise measurement during motion, it was decided to only take measurements during standstill, which necessitates a stop-and-go movement pattern for the rover.

The integrated Intel RealSense D435i cameras can output visual information in the visual and infrared spectrum. Besides providing a greater field of view and a higher refresh rate of up to 90 Hz, the integrated infrared cameras also feature a global shutter in comparison to the rolling shutter of the RGB camera [12]. The source of visual information for the SLAM and odometry algorithms can be freely chosen. A distinction is therefore made below between two modes: *RGB*, using visual information from the visible light spectrum, and infrared (*IR*), which takes advantage of the benefits provided by the infrared cameras.

## 3 ACCURACY ASSESSMENT

In order to obtain objective information on the SLAM system's accuracy and mapping performance, quantitative methods were used. Besides evaluation of the estimated trajectory, the quality of the generated map was also quantified in terms of its scale fidelity.

### 3.1 Trajectory

The accuracy of the estimated path taken over time by the robot, also referred to as its trajectory, can be determined if a ground truth is available. For this purpose, a custom data set was recorded

with the help of ROS bags, while moving the demonstrator through a motion tracking arena on the premises of the Technical University of Munich (TUM). This arena is equipped with a Vicon motion capture system, capable of providing pose estimates at 250 Hz.

Provided the ground truth, two metrics for the quantification of the estimated trajectory were used, as proposed by Sturm et. al [13]. These are the Absolute Trajectory Error (ATE) and the Relative Pose Error (RPE). The ATE is computed by the Root Mean Square Error (RMSE) of the absolute trajectory error matrices $\boldsymbol{F}_i$, calculated from the projected estimated trajectory $\boldsymbol{P}_i$ and the ground truth measurement $\boldsymbol{Q}_i$:

$$\boldsymbol{F}_i = \boldsymbol{Q}_i^{-1}\boldsymbol{P}_i$$

Similarly, the RPE is the RMSE of all relative pose errors $\boldsymbol{E}_i$, computed between adjacent data points:

$$\boldsymbol{E}_i = \left(\boldsymbol{Q}_i^{-1}\boldsymbol{Q}_{i+1}\right)^{-1}\left(\boldsymbol{P}_i^{-1}\boldsymbol{P}_{i+1}\right)$$

Note that in both cases, only the translational components of $\boldsymbol{F}_i$ and $\boldsymbol{E}_i$ are considered.

For the evaluation of the data set, only single-camera SLAM was used to simulate the scenario where one of the cameras fails. Depending on whether the data was post-processed or not, ATE values of 41 mm and 232 mm and RPE values of 230 mm and 199 mm were achieved for the *IR* mode, respectively. Post-processing of the data involves detection of further loop closures and global bundle adjustment, both of which are achieved using tools provided by the RTAB-Map SLAM algorithm.

### 3.2 Map Scale

Due to limited data bandwidth, the output map has to be computed onboard. To estimate the scientific output of the mission, it is important to quantify the quality and level of detail of the map. This was done by mapping a controlled area with placed objects of known sizes and shapes. These object dimensions were then reextracted from the generated mesh, enabling the quantification of a relative scale error $\delta l = \Delta l / l$, with the signed absolute error $\Delta l$ and the true length $l$. Figure 4 shows the resulting absolute and relative linear error of three different dual-camera mapping methods, where both cameras feed one combined SLAM algorithm. Visual information of the infrared spectrum was used for results in the *IR* column, whereas color information was processed for *RGB*. In both cases the connection between the cameras was fixed for better evaluation of the algorithm alone. Opposed to this, *RGB flex* shows results for the color mapping method with a flexible connection between the cameras.

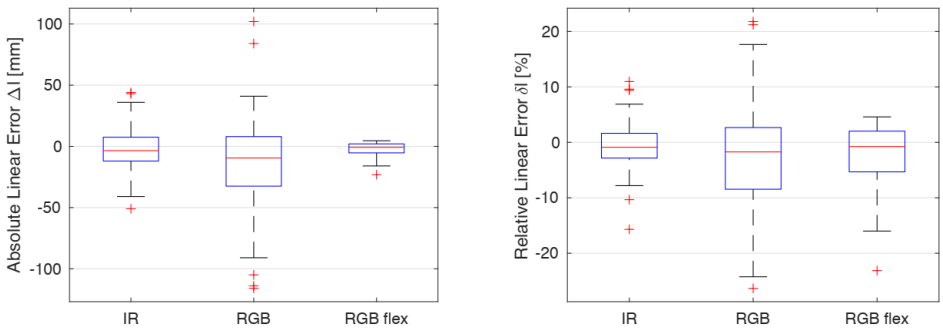


**Figure 4:** Absolute $\boldsymbol{\Delta l}$ and relative linear error $\boldsymbol{\delta l}$ of measured objects for dual-camera *IR*, *RGB*, and *RGB* with flexible interconnection between the cameras (*RGB flex*). Depicted are the mean, upper and lower quartiles, and minimum and maximum values, excluding outliers. [6]

## 4 FIELD TESTING

To get more information on the system's performance in a naturally formed and confined environment, a field test was performed in a karst cave system, located south of Munich. For this first test, the demonstrator was carried along the cave in a slow walking speed with fixed vertebral segments. Since the LED matrix lighting system was not yet completed at the time of testing, one headlamp of the type HM65R from Fenix was mounted on top of the cameras, as shown in Figure 3. However, these headlamps only offer a limited range of illumination and no dynamic dimming in relation to the viewing distance. The scans were recorded as ROS bags and processed retrospectively, simulating real-time conditions. During the scan, the connection between the cameras was fixed, as the focus of this test was to analyze the performance of the visual SLAM algorithm in this specific environment in combination with challenging lighting conditions. Figure 5 shows a visual representation of the live data processing during the scan. After a complete scan, global bundle adjustment and the detection of more loop closures were performed prior to the mesh export to improve the quality of data.

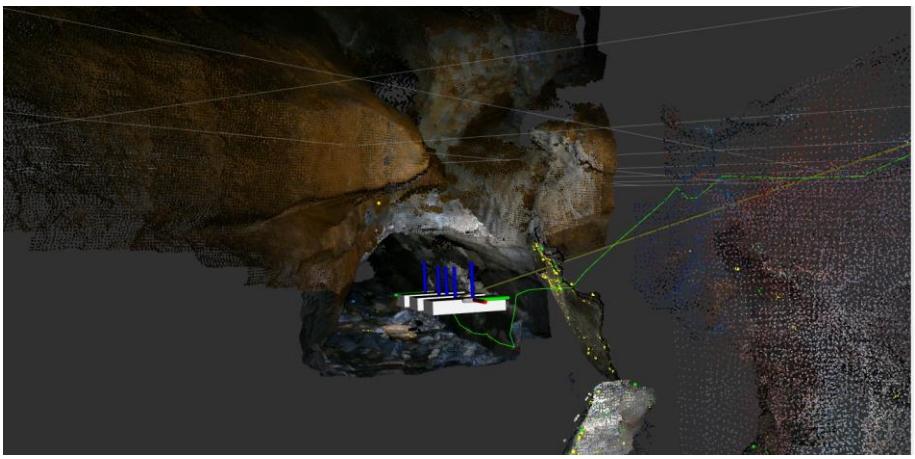

**Figure 5:** Live mapping process of the cave in *RGB* mode, visualized in RViz. Visible are a simple robot model, followed by its path and surrounded by the generated point cloud. [6]

Figure 6 shows two meshes, generated by two individual *RGB* scans. Alignment with a cartographically generated 2D cave map [14] revealed negligible drift in the scans. Furthermore, the total height difference extracted from the scans (4.8 m) closely aligns with the measured height difference of 5.0 m.

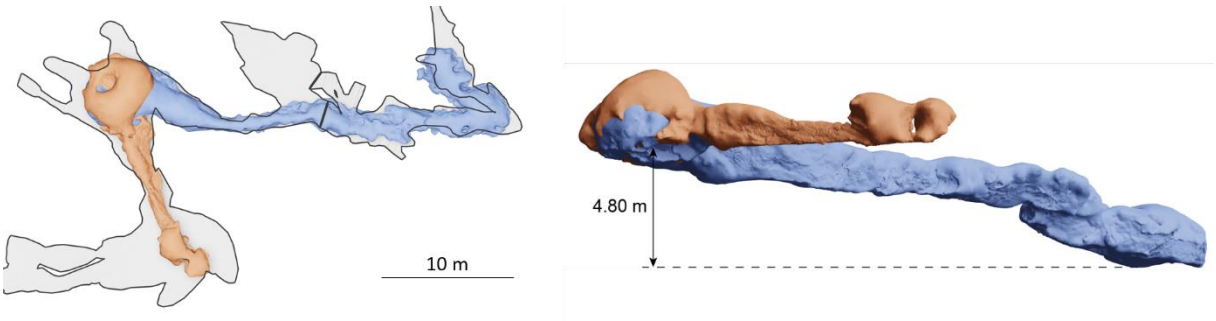


**Figure 6:** Two manually aligned meshes of the mapped cave system from two separate scans. **Left:** Top-view in comparison the outlines of a cartographically generated map [14]. **Right:** Orthographic side-view with maximum elevation difference. [6]

When generating the mesh, there is the option to texturize it. This involves projecting images onto the faces of the mesh, creating a significantly more detailed 3D model. However, as the uncontrolled lighting conditions during the cave scan lead to constant over- or underexposure

of the rock, the texturized mesh can be of poor quality in some areas, as can be seen in Figure 7. Despite untextured and overexposed areas, distinctive points can be made out in both, hinting that the meshes align well.

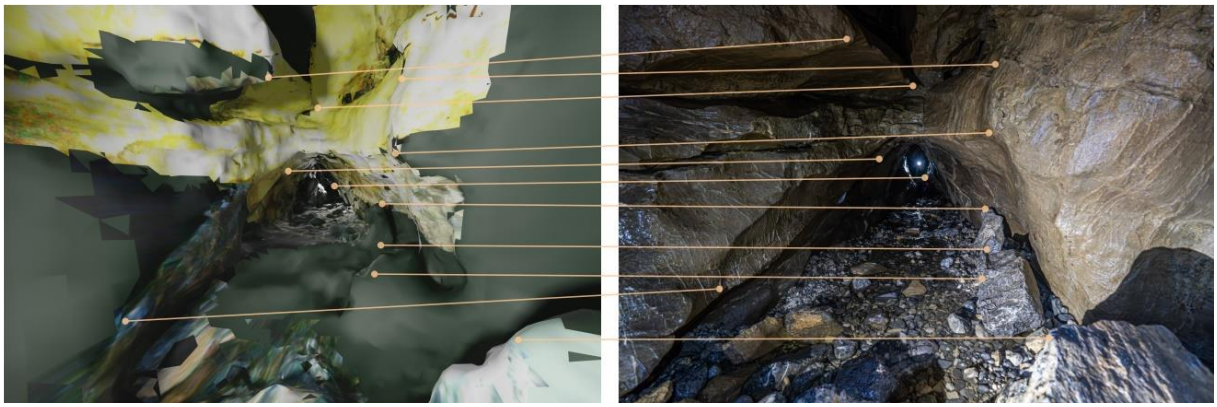

**Figure 7:** Comparison of a photograph (right) and rendered image of generated textured mesh (left). Highlighted are distinctive points, retrievable in both pictures. [6]

Furthermore, a comparison between the meshes generated from scans in *RGB* and *IR* mode can be seen in Figure 8. Noticeably, some of the mesh faces are left untextured. This can result from the specific face being occluded, positioned too far away or angled too broadly from the camera to yield good quality. However, the *IR* scan shows less of these artifacts, thanks to the superior infrared camera properties.

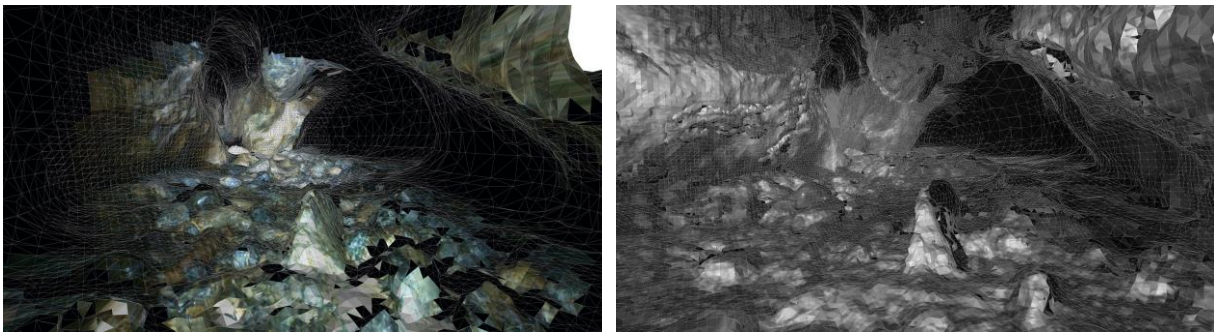

**Figure 8:** Comparison between textured cave meshes captured with *RGB* (left) and *IR* (right) mode [6]

## 5 CONCLUSION

Due to the compliant nature of the SCOUT rover, the implementation of a dual-camera SLAM system was a challenging task due to unknown coordinate transformation between the camera positions in the front and rear module of the rover. The chosen method fuses information from both cameras online in a single SLAM approach utilizing IMU data for angle estimation between the rover modules, which minimizes system load and enhances mission duration. A Stop and go movement pattern is needed to create detailed maps with high accuracy, enabling a valuable scientific outcome of the mission.

The *IR* method shows a clear advantage in texturizing the generated mesh with higher quality images due to its wider field of view and global shutter, resulting in more detailed meshes with less artifacts. Despite requiring greyscale texturization, the *IR* method remains superior for accurate odometry algorithms and performs better in small-scale accuracy tests compared to *RGB*. A hybrid setup using *IR* images for odometry and RGB-D streams for colored map generation will be implemented and tested in the future.

This work paths the way for the SCOUT rover autonomous navigation system, enabling future development through a standardized and expandable ROS environment, which will significantly reduce development efforts and accelerate progress towards a final product.

## REFERENCES


[1] R. Greeley, "Lava tubes and channels in the lunar Marius Hills," *The moon*, vol. 3, no. 3, pp. 289–314, Dec. 1971, doi: 10.1007/BF00561842.

[2] X. Qiu and C. Ding, "Radar Observation of the Lava Tubes on the Moon and Mars," *Remote Sens.*, vol. 15, no. 11, Art. no. 11, Jan. 2023, doi: 10.3390/rs15112850.

[3] L. Carrer, R. Pozzobon, F. Sauro, D. Castelletti, G. W. Patterson, and L. Bruzzone, "Radar evidence of an accessible cave conduit on the Moon below the Mare Tranquillitatis pit," *Nat. Astron.*, pp. 1–8, Jul. 2024, doi: 10.1038/s41550-024-02302-y.

[4] R. Lichtenheldt, E. Staudinger, S. Adeli, J.-P. de Vera, G. Giudice, and M. Baqué, "A MISSION CONCEPT FOR LAVA TUBE EXPLORATION ON MARS AND MOON –THE DLR SCOUT ROVER," in *Lunar and planetary Science Conference*, Virtuell, 2021. Accessed: May 22, 2023. [Online]. Available: https://www.hou.usra.edu/meetings/lpsc2021/pdf/1443.pdf

[5] M. Schütt, D. Lizio, and J. F. Koch, "A Lava Tube Lighting Solution for the SCOUT Rover's Visual System," presented at the i-SAIRAS, Brisbane, Australia, 2024.

[6] J. F. Koch, "Integration and Analysis of a SLAM Platform for the DLR Scout Rover Using Two Compliantly-Connected RGB-D Cameras," Master Thesis, Technische Universität München, Oberpfaffenhofen, 2023. [Online]. Available: https://elib.dlr.de/201188/

[7] M. Labbé and F. Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *J. Field Robot.*, vol. 36, no. 2, Art. no. 2, 2019, doi: 10.1002/rob.21831.

[8] M. Labbé, *rtabmap_ros*. (Oct. 19, 2023). C++. IntRoLab. Accessed: Oct. 20, 2023. [Online]. Available: https://github.com/introlab/rtabmap_ros

[9] T. Foote, "tf: The transform library," in *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, Apr. 2013, pp. 1–6. doi: 10.1109/TePRA.2013.6556373.

[10] S. O. H. Madgwick, "An efficient orientation filter for inertial and inertial/magnetic sensor arrays," 2010, Accessed: Oct. 16, 2023. [Online]. Available: https://courses.cs.washington.edu/courses/cse474/17wi/labs/l4/madgwick_internal_report.pdf

[11] *IMU tools for ROS*. (Nov. 15, 2023). C++. CCNY Robotics Lab. Accessed: Nov. 15, 2023. [Online]. Available: https://github.com/CCNYRoboticsLab/imu_tools

[12] Intel, "Intel RealSense D400 Series Product Family Datasheet," Intel, Datasheet 337029–013, Jun. 2023. Accessed: Jun. 14, 2023. [Online]. Available: https://www.intelrealsense.com/wp-content/uploads/2023/03/Intel-RealSense-D400-Series-Datasheet-March-2023.pdf

[13] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura-Algarve, Portugal: IEEE, Oct. 2012, pp. 573–580. doi: 10.1109/IROS.2012.6385773.

[14] R. Lichtenheldt, "Cave Map 1271-1 Angerlloch," 2024.