The S3LI Vulcano Dataset: A Dataset for Multi-Modal SLAM in Unstructured Planetary Environments

Riccardo Giubilato
German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
riccardo.giubilato@dlr.de

Marcus Gerhard Müller
German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
marcus.mueller@dlr.de

Marco Sewtz
German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
marco.sewtz@dlr.de

Laura Alejandra Encinar Gonzalez
German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
KTH Royal Institute of Technology
laura.encinargonzalez@dlr.de

John Folkesson
KTH Royal Institute of Technology
Deptartment of Intelligent Systems
johnf@kth.se

Rudolph Triebel
German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
rudolph.triebel@dlr.de

arXiv:2601.19557v1 [cs.CV] 27 Jan 2026

***Abstract*—We release the S3LI Vulcano dataset, a multi-modal dataset towards development and benchmarking of Simultaneous Localization and Mapping (SLAM) and place recognition algorithms that rely on visual and LiDAR modalities. Several sequences are recorded on the volcanic island of Vulcano, from the Aeolian Islands in Sicily, Italy. The sequences provide users with data from a variety of environments, textures and terrains, including basaltic or iron-rich rocks, geological formations from old lava channels, as well as dry vegetation and water. The data (`rmc.dlr.de/s3li_dataset`) is accompanied by an open source toolkit (`github.com/DLR-RM/s3li-toolkit`) providing tools for generating ground truth poses as well as preparation of labelled samples for place recognition tasks.**

TABLE OF CONTENTS

1. INTRODUCTION 1
2. RELATED WORKS 2
3. THE S3LI VULCANO DATASET 3
4. EXAMPLE APPLICATIONS 5
5. CONCLUSIONS 7
ACKNOWLEDGEMENTS 7
REFERENCES 7
BIOGRAPHY 9

1. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) components are necessary for mobile systems to achieve high-level autonomy, as they provide the ability to perform localization on a simultaneously growing map from an entirely unknown environment. While having achieved in the past year a fairly elevated level of maturity for traditional artificial environments (e.g., cityscapes or indoor environments), methods that achieve high performances in fully unstructured settings, such as for planetary-like scenes, belong to an open and active area of research [1], [2], [3].

979-8-3315-7360-7/26/$31.00 ©2026 IEEE

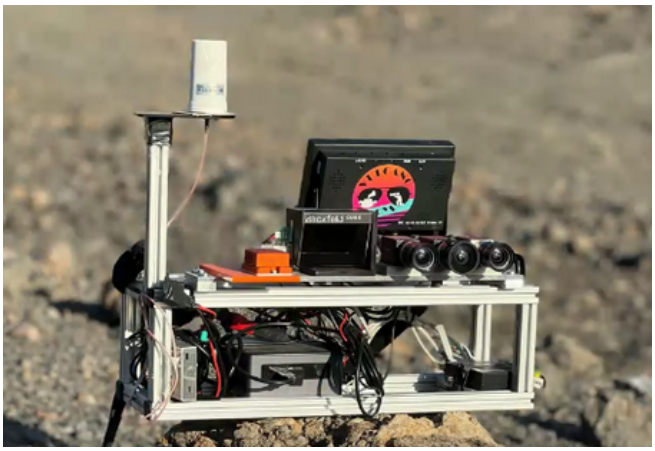


Figure 1: Impression of the **S3LI** (RGB **S**tereo, **S**olid-**S**tate **L**iDAR, **I**nertial) sensor setup, captured on the *Gran Cratere della Fossa*, the active center of Vulcano, vulcan from the homonymous island from the Aeolian Islands, Sicily. On the setup it is visible the white GNSS antenna, computer unit from an Intel NUC and touchscreen to operate the sensor setup.

Development and testing of SLAM algorithms in challenging outdoor environments is difficult due to these places of interest being often hardly reachable (e.g., volcanoes or deserts), and a solution to this is to rely on publicly available data. The majority of data available to challenge and evaluate the accuracy of multi-modal SLAM algorithms, however, often consider environments that are affected by anthropogenic factors. The introduction of human-made elements, both in urban but also natural contexts, often introduce perceptual and structural regularities that bias the performance of state estimation and mapping. When recording data from moving platforms, such as in the case of autonomous driving, the repetitiveness of the driving direction impose significant limitations to the variety of viewpoints to expect, significantly helping place recognition algorithms in re-detecting previously visited scenes to establish "Loop Closures" within SLAM. In planetary exploration settings, however, mobile

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. DOI: 10.1109/AERO66936.2026.11519802

systems operate in highly unstructured natural environments, which open large possibilities to traverse environments, posing significant challenges to the aforementioned tasks. Natural environments are characterized by extreme perceptual aliasing and abundance of ambiguous features, which, exacerbated by unconstrained traversing opportunities, pose significant limitations to SLAM algorithms to provide an accurate pose to the observer with respect to the origin of their map representation.

To foster research in the fields of localization and mapping in planetary settings, and provide data to researchers to develop, test and tune algorithms for this purpose, we release a multi-modal dataset recorded from an hand-held device, that mimicks the height and point of view of a mobile robotic explorer. The sensor setup, named **S3LI** (**S**tereo, **S**olid-**S**tate **L**iDAR **I**nertial) comprises a pair of RGB cameras in horizontal stereo configuration, a compact LiDAR with a MEMS-actuated reflective mirror, an industrial-grade IMU and a GNSS antenna. All sensors are carefully calibrated on the field, and time-synchronized via Precision Time Protocol (PTP). A differential GNSS solution is also provided for evaluation and global referencing. Contrarily to existing multi-modal datasets for SLAM, our sensor setup is intended to explore future possibilities for perception systems suitable for planetary exploration thanks to traditional camera sensors and an alternative LiDAR technology to common spinning mirrors.

The dataset is recorded in the island of Vulcano, Eolian Islands, Sicily, characterized by peculiar geological properties resulting from the fusion of several volcanoes, of which one currently active. The island presents a large variety of natural environments with appearance and structures that resemble settings for future scientific exploration missions. Several sequences observe different traversable terrains and geological features spanning from lava paths or sedimentary structures, posing different challenges towards localization accuracy and long-term consistency of mapping pipelines, offering furthermore various options towards efficient fusion of complementary modalities. In addition to this, we release sequences that observe basaltic lava structures, water and vegetation, introducing dynamic elements that disturb traditional algorithmic pipelines.

This paper comprises a collection of Robot Operating System (ROS) recordings of raw sensor data, including calibration sequences. Accompanying the paper, we release a complementary toolkit as an open-source tool to replicate calibration procedures and post-processing of the data, e.g., generating differential GNSS solutions. We furthermore release example scripts and configuration files to execute existing state of the art SLAM algorithms as an aid towards the utilization of the data.

2. Related Works

Several datasets have been released over the years to address the development and testing of SLAM performances. In literature, the predominant case for visual- or LiDAR-based SLAM algorithms relates to indoor or artificial scenes [22], with references to applications such as Augmented Reality, or robotics for constructions [23], targeting often short sequences in very controlled conditions [24]. A large body of literature relates otherwise to the case of autonomous driving [25], [26], presenting long sequences from the perspective of a moving car, traversing static or dynamic scenarios and across diverse weather and lighting characteristics [27]. This large amount of data, however, fails to address specific conditions which are typical of what robots encounter in the field. First, road-bound vehicles can traverse environments in two directions, imposing significant constraints on the diversity of viewpoints and perspectives and affecting positively the performances of place recognition algorithms for detection of loop closures [28]. Secondly, artificial outdoor environments, such as urban settings, introduce unique elements, both from a visual as well as structural perspective, that facilitate localization. To address the case of a robot rowing in the field, several datasets have been proposed, of which a selected sample is here discussed.

Unstructured and Natural Terrestrial Environments

A relatively large body of literature exists that relates to natural unstructured terrestrial environments. Several datasets address the case of mapping in forest scenes, with target use-cases of either survey of trees and forest digitalization or large-scale navigation from wheeled vehicles. Wild-Places [12] addresses the task of LiDAR-based place recognition in kilometer-long sequences. The authors provide LiDAR submaps, accompanied by RGB imagery, with significant overlap, to investigate loop closure detection or relocalization in the context of multi-session mapping. Similarly, Great Outdoors [16] present data captured in forests, this time from the perspective of a rover system, and adding semantic annotation to RGB data. Botanic Garden [13], instead, addresses the case of under-canopy localization and mapping in the specific context of botanical gardens, providing precise time-synchronization and calibration between RGB and multiple LiDAR sensors, as well as semantic annotations. More variety, in the context of motion patterns, as well as weather and lighting conditions, is introduced by the ROVER [11] dataset, capturing sequences in parks and gardens recorded from the perspective of a robotic platform. While exposing significant challenges with respect to aliasing and ambiguity of appearance, these datasets often present regularities introduced by driving along paths or unpaved roads, therefore facilitating relocalization tasks. Furthermore, the abundance of vegetation makes these data not relevant for benchmarking of algorithms targeted at planetary-like environments

Planetary-Like Environments

Several attempts at gathering data which is representative of planetary environments have been made, generally targeting locations which, for their geological properties and overall appearance, resemble parts of the lunar or martial soil. The MADMAX [4] dataset, collected in three locations in the Moroccan desert, provide images and depth from a stereo setup, with accurate D-GNSS ground truth and orientations. While providing diverse sequences of particular motion patterns, aimed at traversing or mapping tasks, the presence of a single sensing modality limits the applicability to uni-modal settings. Collected in a geographically close location, the Erfoud [8] dataset provides, in addition to visual and multi-spectral data, LiDAR scans from two rovers, but is limited in variety of environments and scenes. Collected in artificial planetary analogous environments, the Canadian Planetary Emulation dataset [6] as well as the Katwijk beach dataset [9] provide rover sequences with stereo, or multi-camera, and LiDAR sequences or fully or partially artificial settings, aiming at replicating natural rocky structures that can be observed in martian scenes. Similary to the Erfoud dataset, however, the variety of scenes is limited to the circumscribed test environment and, true or artificial, geological features.

Table 1: Evaluation of selected related works. Here is only considered data related to mobile robots, or hand-held platforms, excluding the case of road-bound vehicles.

Name	Platform	Env.	Mono	Stereo	IMU	Lidar	GT
Planetary-like Environments							
MADMAX [4]	Hand-held	Desert	✓(M)	✓(M)	✓	✗	D-GNSS
S3LI-Etna [5]	Hand-held	Volcano	✓(M)	✓(M)	✓	✓(SS)	D-GNSS
Canadian Planetary Emulation [6]	Rover	Planet. testbed	✗	✗	✗	✓(M)	D-GNSS
TAIL-Plus [7]	Quadruped	Sandy beach	✓(C)	✓(C)	✓	✓(M)	D-GNSS
Erfoud [8]	Rover	Desert	✓(M)	✓(M)	✓	✓(M)	D-GNSS
Katwijk Beach [9]	Rover	Sandy beach	✓(C)	✓(C)	✓	✓(M)	D-GNSS
Mars C-SLAM [10]	Multi-rover	Planet. testbed	✗	✗	✓	✓(M)	D-GNSS
Unstructured Natural Terrestrial Environments							
ROVER [11]	Rover	Gardens	✓(C)	✓(C)	✓	✗	Prism
WildPlaces [12]	Hand-held	Forest	✓(C)	✓(C)	✓	✓(M)	SLAM
Botanic Garden [13]	Rover	Gardens	✓(C)	✓(C)	✓	✓(M+SS)	SLAM
DiTer++ [14]	Quadruped	Gardens	✓(C)	✗	✓	✓(M)	SLAM
GreenBot [15]	Rover	Forest	✓(C)	✓(C)	✓	✓(M)	D-GNSS
Great Outdoors [16]	Rover	Forest	✓(C)	✓(C)	✓	✓(M)	D-GNSS
Artificial Structured Environments							
IndoorMCD [17]	Platform	Households	✓(C)	✓(C)	✓	✗	Tracker
M2DGR [18]	Rover	Indoor	✓(C)	✓(C)	✓	✓(M)	Multiple
S3E [19]	Rover	Uni. Campus	✓(C)	✓(C)	✓	✓(M)	D-GNSS
Multi-Floor [20]	Rover	Indoor	✓(C)	✓(C)	✓	✓(M)	SLAM
Other							
SubT-MRS [21]	Multiple	Multiple	✓(C)	✓(C)	✓	✓(M+SS)	Multiple
S3LI-Vulcano	Hand-held	Multiple	✓(C)	✓(C)	✓	✓(SS)	D-GNSS

The TAIL dataset [7] targets quadruped robots in beach settings, provides multi-modal inputs and revisit of scenes across diverse daytimes, but limited in the scene diversity.

3. The S3LI Vulcano Dataset

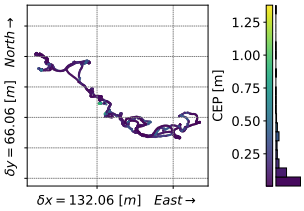
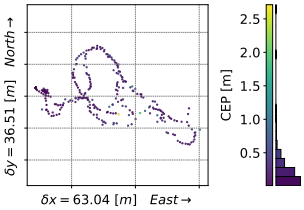
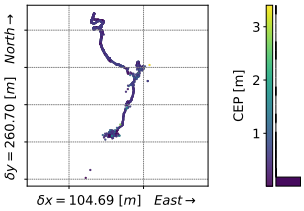
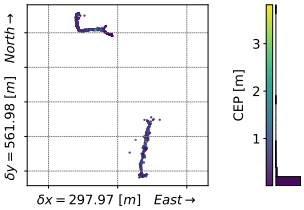
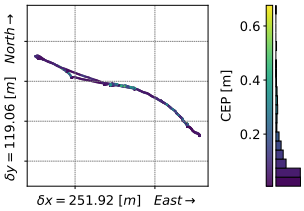
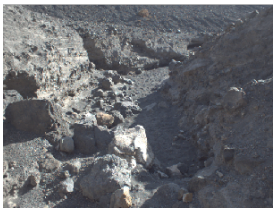
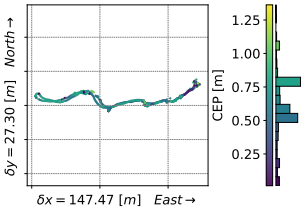
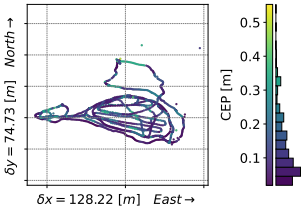
The S3LI "Vulcano" dataset is an extension of the S3LI dataset [5], recorded on a plateau near the summit of Mt. Etna, Sicily, and around the Cisternazza crater. The original S3LI dataset targeted a "moon-like" environment for its appearance, characterized by black and dark gray slopes of lava ash, with ambiguous terrain features and lack of outstanding structures. The overarching idea of both iterations is to provide multi-modal data, where a LiDAR modality is recorded from the perspective of a Solid-State LiDAR system, with a MEMS-actuated oscillating mirror, opposed to traditional spinning mirror LiDAR technologies. This, in conjunction with traditional visual and inertial sensing, drives the development of state estimation and SLAM algorithms, as well as place recognition for detection of loop closures, specifically in the context of potential space-qualifiable sensing technologies.

The S3LI Vulcano dataset provides complementary data to the original dataset improving the variety of data inputs and scenes. More specifically:

- we provide **RGB data** instead of monochrome data, which is particularly important in the context of **semantic segmentation** and classification of objects and terrains [29].
- the LiDAR modality exploits a second iteration of the original sensor, which increases the **minimum sensing range** to 1.5 meters, better overlapping the range of operation of depth from stereo.
- we record data that include **sand**, **rocks**, **water** and **vegetation**, driving towards a better balance between terrain classes and features.

Table 2 provides an organized overview of the seven sequences that the dataset provides. The `waterfront` sequence is recorded on a bay where dark basaltic rock creates a high promontory. In this sequence, dry grass mixes with rich brown terrains and dark rock, and frequent revisits of the same places allows to test place recognition algorithms across different viewpoints, as well as consistency of 3D mapping techniques. The `vegetation`, `capo_grillo` and `capo_grillo_2` observe a mixture of vegetation and dark soil, often looking at the terrain from various distances, where multi-modality can be well exploited. The sequences `straight_path`, `lava_tunnel` and `moon_lake` are all recorded near the summit of *Gran Cratere*, the summit of the volcano that gives the island its name. Here, complex volcanic features, such as an old tunnel from lava flow, or the basin of an old *caldera* with a constellation of scattered rocks and boulders, gives many options to the user to exploit

Table 2: Overview of sequences from the S3LI Vulcano Dataset

Seq. Name	Example Image	GPS Track [a]	Duration	Notes
`waterfront`			24:47s	Long sequence on a waterfront of basaltic rocks and vegetation. Plenty of opportunities for re-localization, and exploitation of multi-modality.
`vegetation`			3:59s	Simple sequence with loops under tree canopies and dry grass. Opportunities for structure-based loop closure detection, e.g. tree trunk detection and mapping.
`capo_grillo`			27:36s	Long sequence on a hiking path, covering diverse environments, from grassy to rocky. Some opportunities for loop closure but incomplete GNSS track.
`capo_grillo_2`			18:40s	Shorter sequence with similar character to `capo_grillo` focusing more on grassy areas but with far-range structures to investigate multi-modal mapping
`straight_path`			12:21s	Simple hiking sequence on a straight path with repetitive visual and structural features. Useful to investigate loop closure from opposing views.
`lava_tunnel`			8:47s	Unique sequence on an old lava tunnel, with walls up to 2 meters high. Offers opportunities to study loop closure techniques based on natural structures
`moon_lake`			22:28s	Long sequence walking in a flat "lake"-like basin of gray sand, inside a small crated of orange rocks. Numerous revisits and repeated traverses in alternating directions to evaluate loop closure and mapping.

[a] The trajectories plotted are filtered with respect to the *quality* metric, and plotted only for $\mathbf{Q} >= \mathbf{2}$ (floating solution or RTK Fix)

Table 3: List of components of the S3LI setup

Component	Rate	Characteristics
AVT G-319C (2x)	10 Hz	• 2 MPx images • Color images • Hardware triggered
XSens MTi-G 10	400 Hz	-
Blickfeld Cube 1	4.7 Hz	• Solid State LiDAR • 70(H)x30(V) FoV • Elliptical pattern
UBlox GNSS	5 Hz	• One GNSS Antenna • Offline D-GNSS
Intel NUC	-	• Intel i7 CPU • 16 Gb RAM

different terrain geometries.

Sensor Setup

The sensor setup, visible in Figure 1 and summarized in Table 3, comprises:

- a pair of hardware-triggered AVT Manta G-319C RGB cameras, capturing images at a fixed frame rate of 10 Hz
- a solid-state Blickfeld Cube-1 LiDAR, capturing scans at a rate of 4.7 Hz with a Field of View of 70 degrees horizontal and 30 degrees vertical, for a total of 17800 points per scan
- an XSens MTi-G 10 IMU connected via USB to the main pc
- a GNSS antenna connected to an UBlox module, and recording raw GNSS binary dumps to disk, for later offline processing of a differential solution

Sensors are time-synchronized with the clock of the main PC, an Intel NUC equipped with an i7 processor, using the PTP4l protocol, supported by the cameras and the LiDAR. IMU measurements are instead assigned timestamps at the time of arrival into the NUC and a value for the latency is later estimated within a Visual-Inertial Odometry framework. All sensor drivers, except for the UBlox device, are interfaced with the Robot Operating System (ROS) and the data is recorded and packaged using the *rosbag* serialization format.

Calibration and Post-Processing

Camera Calibration— Cameras are calibrated as a stereo setup using DLR's CalDe / Callab [2] [30]. Special calibration board with reference markers allow the camera to observe partial views of the checkerboard, with the important benefit for stereo camera setups to cover the entire image with corner points regardless of the overlap between views.

IMU-Camera Calibration— IMU-to-camera calibration is performed using Kalibr [31], an open-source multi-camera-IMU calibration toolkit. Kalibr uses custom calibration patterns relying on arrays of AprilTag markers and determines 6D transformations between sensors, as well as clock differences. We provide the best multi-camera-IMU calibration results obtained with respect to the original unprocessed data, therefore full resolution images as in Table 3.

In practise, however, the calibration from the Kalibr toolkit provides an initial approximation of the extrinsic parameters to either R-VIO2 [32], [33] or VINS-Fusion [34], [35], state of the art visual-inertial odometry algorithms that have the ability to estimate online, i.e., during motion, 1) the transformation from the IMU frame to the left camera frame, and 2) the time offset between the IMU timestamps and the left camera timestamps, assuming that there is time-varying latency in the loop. We experience that, when evaluating the performances of these algorithms, online optimization of the extrinsic parameters from cameras to IMU provide generally the best results and robustness.

[2] https://www.dlr.de/en/rm/research/publications-and-downloads/software

Table 4: Performance comparison of a selected sample of V-SLAM algorithms. RMSE is expressed in meters, as well as relative to the length of the ground-truth trajectory.

Name	RMSE [m]						
R-VIO2	2.29 0.28%	N/I	0.56 0.03%	0.39 0.04%	0.92 0.14%	1.02 0.14%	1.65 0.12%
VINS-Fusion[a]	35.5 8.73%	2.07 0.28%	F	F	N/I	N/I	1.62 0.13%
SqrtVINS	N/I	N/I	N/I	N/I	N/I	1.04 0.15%	N/I
	`waterfront`	`vegetation`	`capo_grillo`[b]	`capo_grillo_2`[b]	`straight_path`	`lava_tunnel`	`moon_lake`

[a] evaluated using the *global_fusion* mode, i.e., with loop closure.
[b] RMSE is evaluated using a RANSAC implementation of the Horn's algorithm, due to the significant quantity of noisy D-GNSS data points, see Tab. 2

Generation of Differential GNSS Solution for Ground Truth

Although a GNSS track was recorded from the S3LI setup with a single antenna, we provide a differential solution generated offline using the RTKLIB[3]. GNSS data from the UBlox module was recorded as a sequence of NMEA sentences. These include all the necessary information to compute a differential solution with a second reference static antenna on the ground. As the base antenna, the closest one to the Vulcano island and part of the EUREF Permanent GNSS Network is the one in Noto, Sicily. After converting data to the RINEX format, a differential solution was computed for all sequences and saved in the `.pos` format. Details on how to use, obtain or generate data are included in the dataset repository. Table 2 reports the D-GNSS tracks for all sequences, including points with quality level Q$\leq$2, where 1 is full fix (10^{-2} to 10^{-1} meters accuracy) and 2 is floating solution (10^{-1} to 10^{0} meters accuracy). Each D-GNSS track is color-coded with a Circular Error Projection (CEP) estimate based on the standard deviations of the position solution in the directions of latitude and longitude.

4. Example Applications

Compared to traditional datasets for benchmarking and evaluation of visual [36] or multi-modal SLAM techniques, where the scenes are often short in trajectory length and duration, and the scene is usually set in a controlled manner, operating rovers to achieve scientific or infrastructural inspection tasks often implies the absence of bounds on trajectory length and map size. The S3LI Vulcano dataset, in addition to the previous iteration on Mt. Etna, provides sequences with strong

[3] rtklib.com

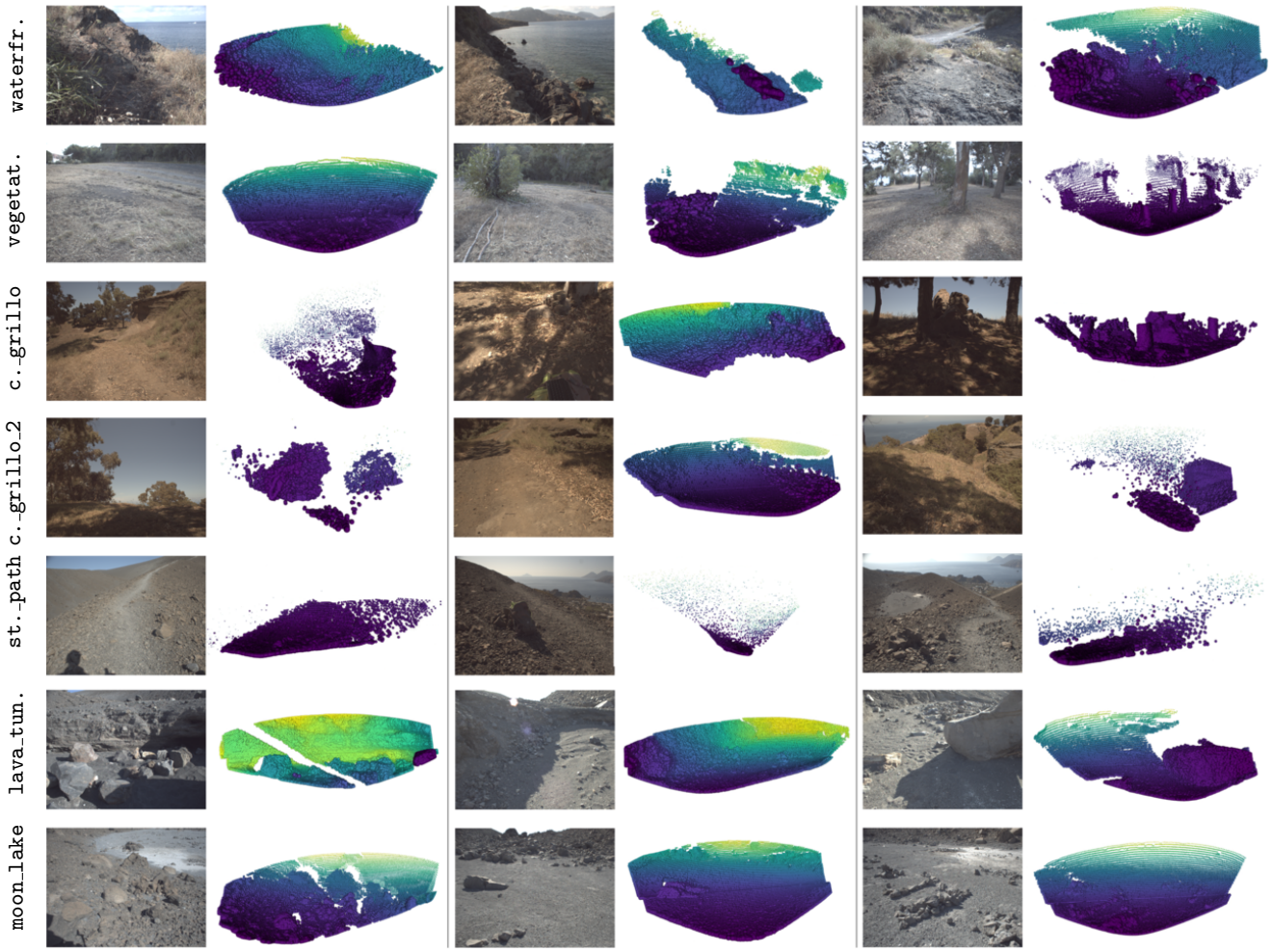


Figure 2: Graphical examples of associated RGB images and LiDAR scans from the dataset sequences. For each sequence, 3 samples of corresponding images and scans are provided

resemblance to actual traversing or large-scale mapping operation of entirely unstructured natural environments. Traditional *full* visual-SLAM techniques, such as ORB-SLAM [37], which require maintenance of feature maps without particular partitioning or marginalization strategies, often fail to guarantee asymptotic performance metrics that are appropriate for real-time operation requirements. In Table 4 is a performance evaluation given as an example application case for visual, or visual-inertial SLAM techniques. The real-time camera pose estimates are aligned using Horn's algorithm, fixing the initial position and determining only the rotation for the best alignment. The total RMSE (Root Mean Square Error) for all the position residuals, as well as the RMSE normalized on the trajectory length, provide an insight on how well an algorithm fits the real camera movement, although it has to be considered that large trajectory errors for odometry measurements can depends on single failure events along the path, and are not indicative of constant drift.

Table 4 reports an simple evaluation of Visual-Inertial SLAM algorithms in various configurations. For these tests, RGB images are downsampled by a factor 4 to facilitate real-time performances in stereo rectification and depth estimation. The first, R-VIO2 [33] is a monocular Visual-Inertial Odometry (VIO) algorithm. It uses a single camera image stream and IMU measurements, and requires static initialization, therefore the algorithm must be started without exciting the IMU axes in order to provide a first bias estimation. The second, VINS-Fusion [35], is a VIO algorithm, used in stereo configuration, with a pose-graph optimization stage, that performs appearance-based loop closure detection and place recognition with BRIEF features to correct the accumulated drift. Errors are expressed as the Root Mean Square Errors between estimated camera poses and corresponding ground truth positions, after aligning the trajectories using the Horn's quaternion method with a fixed null translation. The table reports with the nomenclature "N/I" the case where the algorithm fails to initialize and therefore does not provide any pose output. Otherwise, if the algorithm fails along the trajectory, estimating less than 30% of the camera poses, the table denotes the result as "F". As expected, providing unstructured natural scenes rich local information, pose estimation is a task that can be successfully accomplished by modern algorithms, while the challenges lays on relocalization due to their ambiguity.

Multi-modal Place Recognition

In order to foster research on multi-modal place recognition, we provide as part of the toolkit, functionalities to generate geo-localised pairs of images and LiDAR scans with 6D

ground reference (position w.r.t. an arbitrary origin, and orientation w.r.t. an ENU reference frame). These can be used to train, fine-tune and test models for uni- or multi-modal place recognition. That is, their capabilities to recall previously visited scenes. The user can then implement its own custom mechanism to assign ground truth labels, given the global localization information, as well as depth distribution statistics for the pointclouds. Figure 3 shows the graphical output of the toolkit, allowing the user to carefully inspect the generated data and 6D ground truth poses of data point pairs. In the background, the pipeline stores all information as *pickle* files, to be loaded later in a second time from the user for place recognition tasks.

5. Conclusions

With this paper we release **S3LI Vulcano**, an extension to the original S3LI dataset [5]. The dataset presents additional data, recorded in a selection of severely unstructured natural environments, presenting basaltic rocks, sand, vegetation and water, from the Vulcano island, part of the Aeolian islands in Sicily, Italy. The dataset contains RGB images, pointclouds from an solid-state LiDAR, inertial and GNSS data, in form of ROS bagfiles. We demonstrate an example use case as for the development and testing of visual or multi-modal SLAM algorithms. Finally, we release a software toolkit that generates samples with global ground truth to train and test multi-modal place recognition algorithms.

Acknowledgements

This work was supported by the Helmholtz Association project iFOODis (contract number KA2-HSC-06).

References

[1] R. Giubilato, C. Le Gentil, M. Vayugundla, M. J. Schuster, T. Vidal-Calleja, and R. Triebel, "Gpgm-slam: a robust slam system for unstructured planetary environments with gaussian process gradient maps," *Field Robotics*, vol. 2, pp. 1721–1753, 2022.

[2] D. Geromichalos, M. Azkarate, E. Tsardoulias, L. Gerdes, L. Petrou, and C. Perez Del Pulgar, "Slam for autonomous planetary rovers with global localization," *Journal of Field Robotics*, vol. 37, no. 5, pp. 830–847, 2020.

[3] P. Yin, J. Jiao, S. Zhao, L. Xu, G. Huang, H. Choset, S. Scherer, and J. Han, "General place recognition survey: Towards real-world autonomy," *IEEE Transactions on Robotics*, 2025.

[4] L. Meyer, M. Smíšek, A. Fontan Villacampa, L. Oliva Maza, D. Medina, M. J. Schuster, F. Steidle, M. Vayugundla, M. G. Müller, B. Rebele, A. Wedler, and R. Triebel, "The madmax data set for visual-inertial rover navigation on mars," *Journal of Field Robotics*, vol. 38, no. 6, pp. 833–853, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22016

[5] R. Giubilato, W. Stürzl, A. Wedler, and R. Triebel, "Challenges of slam in extremely unstructured environments: The dlr planetary stereo, solid-state lidar, inertial dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8721–8728, 2022.

[6] C. H. Tong, D. Gingras, K. Larose, T. D. Barfoot, and Érick Dupuis, "The canadian planetary emulation terrain 3d mapping dataset," *The International Journal of Robotics Research*, vol. 32, no. 4, pp. 389–395, 2013. [Online]. Available: https://doi.org/10.1177/0278364913478897

[7] Z. Wang, C. Yao, Y. Ge, G. Shi, N. Yang, Z. Zhu, K. Dong, H. Wei, Z. Jia, and J. Wu, "Are we ready for planetary exploration robots? the tail-plus dataset for slam in granular environments," in *IEEE ICRA Workshop on Field Robotics*, 2024. [Online]. Available: https://arxiv.org/abs/2404.13600

[8] S. Lacroix, A. De Maio, Q. Labourey, E. P. Mendes, P. Narvor, V. Bissonette, C. Bazerque, F. Souvannavong, R. Viards, and M. Azkarate, "The erfoud dataset: a comprehensive multi-camera and lidar data collection for planetary exploration," in *15th Symposium on Advanced Space Technologies in Robotics and Automation*, 2020.

[9] R. A. Hewitt, E. Boukas, M. Azkarate, M. Pagnamenta, J. A. Marshall, A. Gasteratos, and G. Visentin, "The katwijk beach planetary rover dataset," *The International Journal of Robotics Research*, vol. 37, no. 1, pp. 3–12, 2018.

[10] P.-Y. Lajoie, K. Soma, H. M. Bong, A. Lemieux-Bourque, R. Zhang, V. S. Varadharajan, and G. Beltrame, "Multi-robot decentralized collaborative slam in planetary analogue environments: Dataset, challenges, and lessons learned," *IEEE Transactions on Field Robotics*, 2025.

[11] F. Schmidt, J. Daubermann, M. Mitschke, C. Blessing, S. Meyer, M. Enzweiler, and A. Valada, "Rover: A multi-season dataset for visual slam," *IEEE Transactions on Robotics*, 2025.

[12] J. Knights, K. Vidanapathirana, M. Ramezani, S. Sridharan, C. Fookes, and P. Moghadam, "Wild-places: A large-scale dataset for lidar place recognition in unstructured natural environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 11 322–11 328.

[13] Y. Liu, Y. Fu, M. Qin, Y. Xu, B. Xu, F. Chen, B. Goossens, P. Z. Sun, H. Yu, C. Liu *et al.*, "Botanicgarden: A high-quality dataset for robot navigation in unstructured natural environments," *IEEE Robotics and Automation Letters*, vol. 9, no. 3, pp. 2798–2805, 2024.

[14] J. Kim, H. Kim, S. Jeong, Y. Shin, and Y. Cho, "Diter++: Diverse terrain and multi-modal dataset for multi-robot slam in multi-session environments," *arXiv preprint arXiv:2412.05839*, 2024.

[15] F. Cañadas-Aránega, J. L. Blanco-Claraco, J. C. Moreno, and F. Rodriguez, "The greenbot dataset: Multimodal mobile robotic dataset for a typical mediterranean greenhouse," *arXiv preprint arXiv:2402.00438*, 2024.

[16] P. Jiang, K. Viswanath, A. Nagariya, G. Chustz, M. Wigness, P. Osteen, T. Overbye, C. Ellis, L. Quang, and S. Saripalli, "Go: The great outdoors multimodal dataset," *arXiv preprint arXiv:2501.19274*, 2025.

[17] M. Sewtz, Y. Fanger, X. Luo, T. Bodenmüller, and R. Triebel, "Indoormcd: A benchmark for low-cost multi-camera slam in indoor environments," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1707–1714, 2023.

[18] J. Yin, Q. Li, Y. He, Z. Sui, X. Ji, W. Wan,

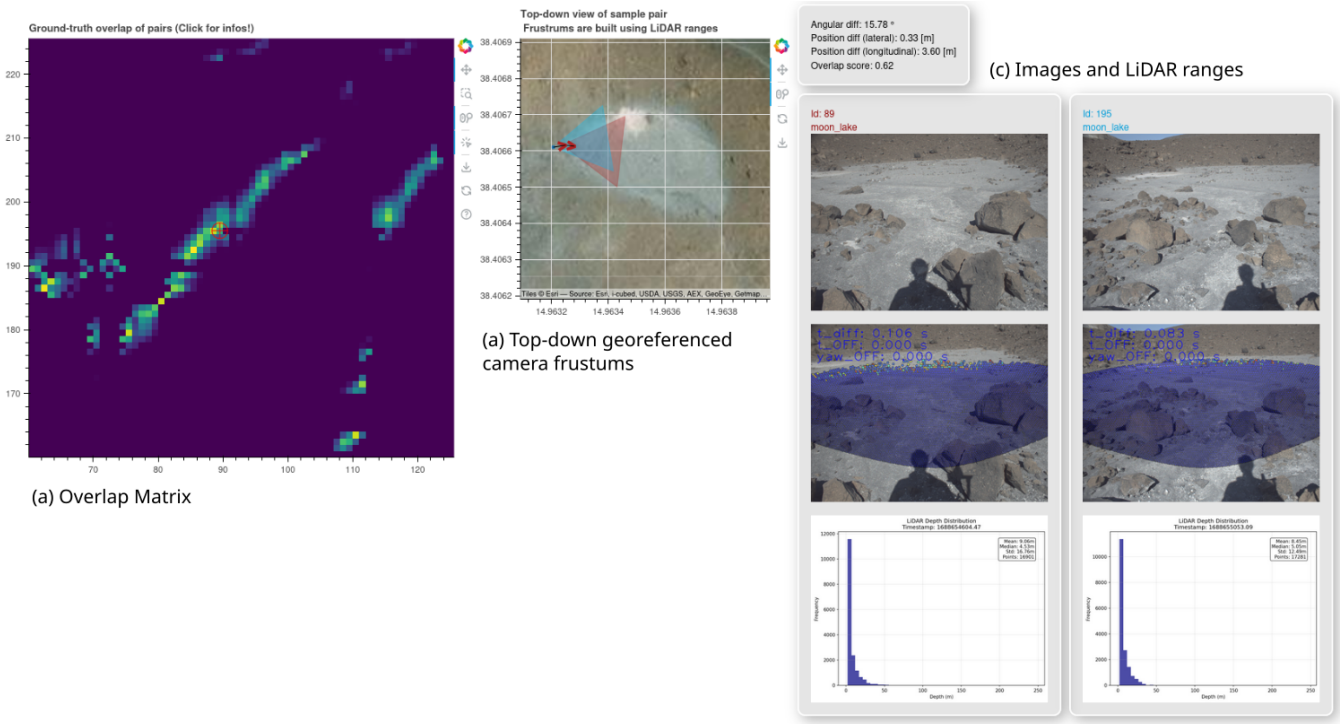


Figure 3: Example View from the place recognition toolkit. On the left, a clickable overlap matrix shows, with colors from blue to yellow, the amount of overlap between two views. When clicked, the corresponding position and FoV (Field of View) of the two images is shown on the map in the center, while images, pointcloud overlaps and point depth histograms are shown on the right.

and Z. Liu, "M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots," *arXiv preprint arXiv:2112.13659*, 2021. [Online]. Available: https://arxiv.org/abs/2112.13659

[19] D. Feng, Y. Qi, S. Zhong, Z. Chen, Q. Chen, H. Chen, J. Wu, and J. Ma, "S3e: A multi-robot multimodal dataset for collaborative slam," *IEEE Robotics and Automation Letters*, 2024.

[20] P. Kaveti, A. Gupta, D. Giaya, M. Karp, C. Keil, J. Nir, Z. Zhang, and H. Singh, "Challenges of indoor slam: A multi-modal multi-floor dataset for slam evaluation," in *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2023, pp. 1–8.

[21] S. Zhao, Y. Gao, T. Wu, D. Singh, R. Jiang, H. Sun, M. Sarawata, Y. Qiu, W. Zhang, and S. Scherer, "Subt-mrs dataset: Pushing slam towards all-weather environments," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. [Online]. Available: https://arxiv.org/abs/2307.07607

[22] J. Sturm, S. Magnenat, N. Engelhard, F. Pomerleau, F. Colas, D. Cremers, R. Siegwart, and W. Burgard, "Towards a benchmark for rgb-d slam evaluation," in *Rgb-d workshop on advanced reasoning with depth cameras at robotics: Science and systems conf.(rss)*, 2011.

[23] M. Helmberger, K. Morin, B. Berner, N. Kumar, G. Cioffi, and D. Scaramuzza, "The hilti slam challenge dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7518–7525, 2022.

[24] T. Schops, T. Sattler, and M. Pollefeys, "Bad slam: Bundle adjusted direct rgb-d slam," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 134–144.

[25] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The international journal of robotics research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[26] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 6433–6438.

[27] P. Wenzel, R. Wang, N. Yang, Q. Cheng, Q. Khan, L. Von Stumberg, N. Zeller, and D. Cremers, "4seasons: A cross-season dataset for multi-weather slam in autonomous driving," in *DAGM German Conference on Pattern Recognition*. Springer, 2020, pp. 404–417.

[28] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *ieee transactions on robotics*, vol. 32, no. 1, pp. 1–19, 2015.

[29] M. G. Müller, M. Durner, W. Boerdijk, H. Blum, A. Gawel, W. Stürzl, R. Siegwart, and R. Triebel, "Uncertainty estimation for planetary robotic terrain segmentation," in *2023 IEEE aerospace conference*. IEEE, 2023, pp. 1–8.

[30] K. H. Strobl, W. Sepp, S. Fuchs, C. Paredes, M. Smisek, and K. Arbter. DLR CalDe and DLR CalLab. http://www.robotic.dlr.de/callab. Institute of Robotics and Mechatronics, German Aerospace Center (DLR).

Oberpfaffenhofen, Germany. [Online]. Available: http://www.robotic.dlr.de/callab/

[31] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 4304–4311.

[32] Z. Huai and G. Huang, "Square-root robocentric visual-inertial odometry with online spatiotemporal calibration," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9961–9968, 2022.

[33] ——, "A consistent parallel estimation framework for visual-inertial slam," *IEEE Transactions on Robotics*, 2024.

[34] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE transactions on robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[35] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," *arXiv preprint arXiv:1901.03642*, 2019.

[36] A. Fontan, T. Fischer, J. Civera, and M. Milford, "Vslam-lab: A comprehensive framework for visual slam methods and datasets," *arXiv preprint arXiv:2504.04457*, 2025.

[37] 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.

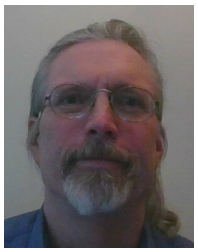
BIOGRAPHY

Riccardo Giubilato is a researcher at the Institute of Robotics and Mechatronics, German Aerospace Center (DLR), in the Dept. of Perception and Cognition since 2019. He leads the Intelligent Exploration and Mapping group, focusing on development of localization, semantic and long-term mapping, and multi-agent exploration approaches to facilitate high-level understanding and exploitation of known or unknown environments. He holds a Ph.D. in Space Sciences, Technologies and Measurements from CISAS "Giuseppe Colombo", and Master's and Bachelor's degree in Aerospace Engineering from the University of Padova, Italy.

Marcus Gerhard Müller is a researcher in the department of "Perception and Cognition" at the German Aerospace Center (DLR) since 2016 and Ph.D. student at ETH Zurich. He is the leader of the MAV Exploration Team at the Institute of Robotics and Mechatronics (DLR-RM), where he is working on autonomous navigation algorithms for MAVs. Before joining DLR he conducted research at the Jet Propulsion Laboratory (JPL) of NASA in Pasadena, USA, where he worked on visual inertial navigation for MAVs and on radar signal processing. Marcus received his Master's and Bachelor's degree in Electrical Engineering from the University of Siegen, Germany

Marco Sewtz is the lead for software and interfaces of the new lunar exploration rover at the German Aerospace Center (DLR) and responsible for the development of AI technology on robotic space systems. He received his B.Eng. at the University of Applied Sciences of Munich, his M.Sc. at the Technical University of Munich (TUM) and his PhD at the Karlsruhe Institute of Technology (KIT). Before his current role, he worked as an electrical designer for high-performance processing modules for space hardware at Airbus Defence and Space

Laura Alejandra Encinar Gonzalez Laura Alejandra Encinar Gonzalez received her BSc in Mathematics and Computer Science from UPM in 2023 and her MSc in ICT Innovation (Autonomous Systems, AI focus) from Aalto University and KTH in 2025. She worked as a student at MIT Media Lab, focusing on multi-objective optimization for spatial design, and at the German Aerospace Center (DLR), where she developed multimodal systems for perception and localization. Her background combines AI, computer vision, and robotics, with a strong interest in real-time autonomous systems.



John Folkesson John Folkesson received his PhD degree from KTH Royal Institute of Technology, Sweden, in 2006. He was a postdoc at MIT until 2008 and then was a research Engineer at MIT until 2010. In 2010 he returned to KTH as a Research Engineer. He is currently a full Professor of Robotics at KTH. Since returning to KTH in 2010, Dr. John Folkesson has been instrumental in establishing underwater robotics within the Robotics, Perception, and Learning (RPL) division of the EECS School. As a Principle Investigator in the Swedish Maritime Robotics Centre (SMaRC), he focuses on developing planning, perception, and navigation software to enhance autonomy in various scenarios.

Rudolph Triebel received his PhD in 2007 from the University of Freiburg in Germany. From 2007 to 2011, he was a postdoctoral researcher at ETH Zurich, working on machine learning algorithms for robot perception within several EU funded projects. From 2011 to 2013, he worked in the Mobile Robotics Group at the University of Oxford, developing unsupervised and online learning techniques for detection and classification in mobile robotics and autonomous driving. From 2013 to 2023, Rudolph was a lecturer at the Technical University of Munich (TUM), teaching master-level courses in Machine Learning for Computer Vision. In 2015, he became the leader of the Department of Perception and Cognition at the Robotics Institute of the German Aerospace Center (DLR), and in 2023, he was appointed as an university professor at the Karlsruhe Institute of Technology (KIT) in "Intelligent Robot Perception".