GPGM-SLAM: Towards a Robust SLAM System for Unstructured Planetary Environments with Gaussian Process Gradient Maps

Riccardo Giubilato[1†], Cedric Le Gentil[2†], Mallikarjuna Vayugundla[1†], Teresa Vidal-Calleja[2], Rudolph Triebel[1,3]

***Abstract*— Simultaneous Localization and Mapping (SLAM) in unstructured planetary environments is a challenging task for mobile robots due to the appearance and structure of the environment. In urban and man-made scenarios, individual objects (e.g. cars, trees or buildings) are easily discernible and the visual appearance is likely to provide unique cues for the purpose of localization. Contrarily, planetary scenarios are often characterized by repetitive structures and ambiguous terrain features. To provide robust place recognition abilities in the context of submap-based stereo visual SLAM, we propose to utilize the gradient of elevation maps generated by Gaussian Processes (GPs). Visual features computed on GP Gradient Maps (GPGMaps) provide means for efficient place recognition, through encoding in Bag-of-Words vectors, and for SE(2) alignment to establish loop closure constraints in a pose graph. We evaluate the proposed SLAM system on relevant Moon-like environments through real data captured on Mt. Etna, Sicily.**

I. Introduction

A key task towards robotic autonomy is the ability to localize in previously unknown environments. Visual Simultaneous Localization and Mapping (SLAM) exploits cameras for ego-motion estimation, mapping and place recognition. Although visual SLAM [1] techniques can be considered mature enough in man-made environments for the purpose of autonomous driving or augmented reality, unstructured environments, such as on planetary surfaces, challenge the task of place recognition due to visual aliasing and lack of unique visual or structural cues. Thus, by traditional means (i.e. through similarity detection in the visual or structural domain), detection of loop closures and relocalization or multi-session mapping are hindered. In [2] we propose to establish loop closures by matching image-like Gaussian Process Gradient Maps (GPGMaps) obtained from submap point clouds with the use of Gaussian Processes (GPs), showing significant performance improvements over traditional methods based on local structure similarity. In this paper we investigate further extensions of this work towards a full-featured SLAM system that can be run onboard a planetary-like rover. Specifically, we develop further our previous work in the following aspects:

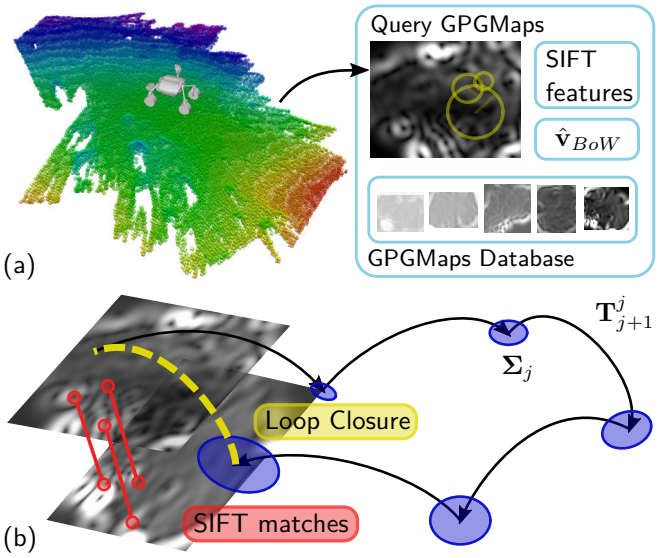

Fig. 1. Overview of the GPGM-SLAM pipeline: (a) GPGMaps are generated using 3D submaps from aggregated stereo clouds. Visual features, computed on the gradient images, are used to create BoW vectors and detect loop closures from a database. (b) Validated GPGMap matches are used to establish loop closure constraints in a pose graph, where submap origins (blue ellipses) are joined by VIO (Visual-Inertial Odometry) pose constraints.

- we implement SKI (Structured Kernel Interpolation) GP regression [3] to help mitigate the computational complexity of GPs.
- we implement a loop closure detector for GPGMaps based on traditional visual features and Bag-of-Words.
- we integrate the new loop closure detector with the match validation scheme from [2] in a submap-based graph SLAM for mobile robots equipped with stereo vision systems [4]. Fig. 1 shows an overview of the proposed pipeline and Fig. 6 shows a visualization of a SLAM session.

II. The SLAM Architecture

In this section we introduce briefly the SLAM system running onboard the Lightweight Rover Unit (LRU), a planetary-like rover developed at the DLR Institute of Robotics and Mechatronics (see Fig. 4). A pose estimation front-end is based on Visual-inertial odometry, which provides locally accurate ego-motion to accumulate stereo point clouds in submaps. Submaps are created by either enforcing a maxi-

† *The authors assert equal contribution and joint first authorship.
[1]German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Weßling, Germany {firstname.lastname@dlr.de}
[2]Centre for Autonomous Systems at the Faculty of Engineering and IT, University of Technology Sydney {cedric.legentil@student.uts.edu.au, teresa.vidalcalleja@uts.edu.au}
[3]Technical University of Munich (TUM), Department of Computer Science {rudolph.triebel@in.tum.de}

TABLE I

COMPARISON OF THE COMPUTATION TIME TO GENERATE IMAGE-LIKE GPGMAPS WITH THE EXACT GP INFERENCE USED IN [2] AND THE PROPOSED METHOD BASED ON SKI [3] (AVERAGE OVER 17 SUBMAPS OF THE ETNA DATASET WITH AN INFERENCE RESOLUTION OF 0.01 m)

Downsampling	Exact GP	SKI
3000 points	592.3 s	19.1 s
None (avg. 38k points)	N.A.	24.9 s

TABLE II

AREA UNDER CURVE (AUC) OF THE TESTED DESCRIPTORS FOR VOCABULARES BUILT FROM 64 AND 128 CLUSTERS

n_cl	3D-SHOT	3D-CSHOT	SURF	KAZE	SIFT
64	(0.13)	(0.15)	0.71	0.43	0.81
128			0.78	0.60	**0.87**

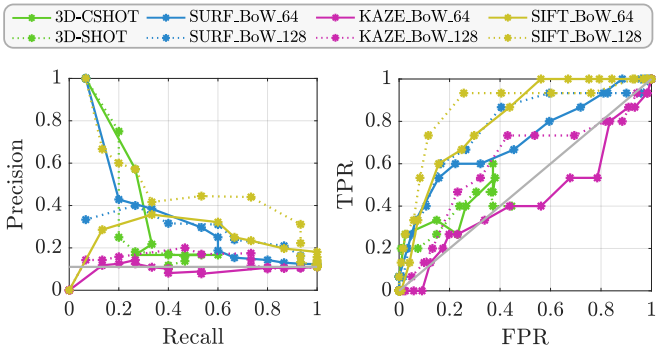

Fig. 2. PR and ROC curves showing the performances of the Loop Closure detector based on BoW vector similarity. The curves are generated by varying a threshold on the BoW score to label a submap pair matching or non matching. This test is performed on GPGMaps from a dataset captured on Mt. Etna and for a selection of traditional visual features. Results are compared with 3D feature matching + RANSAC on the original submap point clouds.

mum path length or a maximum covariance growth from the submap origin. Local reference frames denote the origins of each submap and are constrained sequentially by inter-pose constraints, from VIO, as well as inter-robot detections and loop closures [5]. The optimization back-end is based on the incremental smoother iSAM2 [6]. For more details and performance analyses we refer to [4].

III. GAUSSIAN PROCESS GRADIENT SUBMAPS

Our previous work [2] introduces the concept of GPGMap to address the challenge of noisy and sparse geometric data in the context of place recognition in unstructured planetary environments. GPGMaps represent the gradient of the terrain elevation and are probabilistically inferred from 3D point clouds of the system's environment using GPs and linear operators [7] for direct prediction of the elevation's derivatives. While GP models allow for data-driven accurate interpolation, they suffer from a cubic $\mathcal{O}(n^3)$ computational complexity for the first inference and linear $\mathcal{O}(n)$ for each additional prediction, with n the number of sample data.

To mitigate this constraint, [2] resorts to naive index-based downsampling of the input point clouds. In this work, we propose to use the SKI scheme [3] to directly infer the elevation derivatives using linear operators [7] with elevation measurements solely. The key concept of SKI consists in the use of inducing points that form a grid across the input space, and the efficient interpolation of the data samples' covariance between these points. Our implementation is built on the fast matrix-vector-multiplication computations presented in [8], resulting in the efficient estimation of the kernel's derivatives. The linear time $\mathcal{O}(n)$ inference (and $\mathcal{O}(1)$ for additional predictions) provide a tremendous advantage compared to [2] (cf. Table III).

IV. LOOP CLOSURE WITH GPGMAPS

A. Loop Closure Detection

In our previous work [2], SURF features were extracted on GPGMaps in order to match them in a RANSAC scheme to validate submap matches. Here we evaluate the performances of a loop closure detector, based on visual features and BoW, in selecting valid candidate matches across a database of previously generated GPGMaps. As shown in Fig. 1, features extracted in every generated GPGMap are extracted and encoded, through a vocabulary trained before-hand, in a BoW vector $\mathbf{v}_i$. Similar GPGMaps are queried from a database based on their distance in the BoW vectors space. An evaluation of commonly used visual feature, SIFT, SURF and KAZE is also presented here. In this test, the vocabulary is a plain kd-tree without word weighting (e.g. TF-IDF) using both 64 or 128 clusters and it is computed using GPGMaps built within a variety of datasets recorded on Mt. Etna. The scoring of BoW vectors is based on cosine similarity: $s(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{v}_i \cdot \mathbf{v}_j / |\mathbf{v}_i||\mathbf{v}_j|$. Fig. 2 reports the performances of this loop closure detector through a Precision-Recall curve and a Receiver-Operating-Curve. The curves are built by varying a threshold on the cosine similarity score to label a GPGMap pair as matching or non-matching. True matches are labeled from the ground truth given the Intersection-over-Union (IoU) between submap bounding boxes, see [2] for more details. Results show that the choice of SIFT descriptors leads to better overall performances, as the Area Under Curve (AUC) is higher than in the other configurations. Other descriptors, such as KAZE, are instead not able to classify correct matches and degenerate to random guesses.

B. Loop Closure Validation

The loop closure detection stage presents potential candidate matches based on the distance between BoW vectors. Potential matches are validated to obtain a transformation between the corresponding submap point clouds, which is finally used to establish loop closure constraints in the factor graph. See Fig. 1 and Fig. 3 for a graphical overview of the overall loop closure detection scheme an validation process respectively. The first step is to match SIFT features across GPGMap_1 and GPGMap_2, selected as candidate matching pair. Given the ambiguous and repetitive appearance of gradient images, we evaluate both a brute-force matching approach as well as the ratio metric from [9]. The RANSAC approach presented in [2] is employed to determine an SE(2)

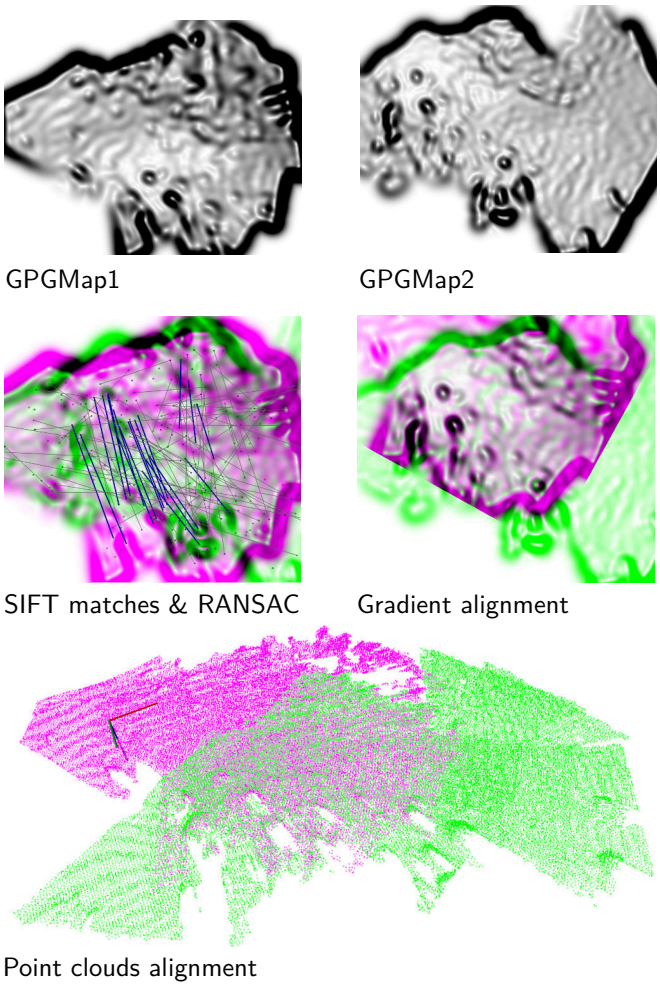

Fig. 3. Overview of the match validation process for candidate matches after loop closure detection: SIFT features are extracted on the generated GPGMap_1 and GPGMap_2 discarding keypoints on locations with high covariance. SIFT features are matched and the RANSAC scheme from [2] is applied to select a fitting affine alignment between the two images. The affine transformation is finally expressed in the local reference system of the submap point clouds and a loop closure constraint is added to the graph.

transformation between the two images without optimizing for scale, which is fixed given the image resolution in meters per pixels defined beforehand as a parameter for generating the GPGMap. The RANSAC model also leverages the gradient and covariance to compute an additional error metric based on a weighted difference of gradients aligned using the affine transformation determined during every iteration with sufficient inliers. The affine transformation returned by the RANSAC test is applied to the original submap point clouds after correcting the translation part for the GPGMap resolution and offset between the image and submap origins. The transformation between point clouds is finally refined using the ICP algorithm and establishes an inter-submap constraint in the pose graph, followed by an iSAM2 update.

V. Experiments on a Moon-like Environment

We test our pipeline on datasets captured on Mt. Etna, a Moon-like environment, in the context of the ROBEX

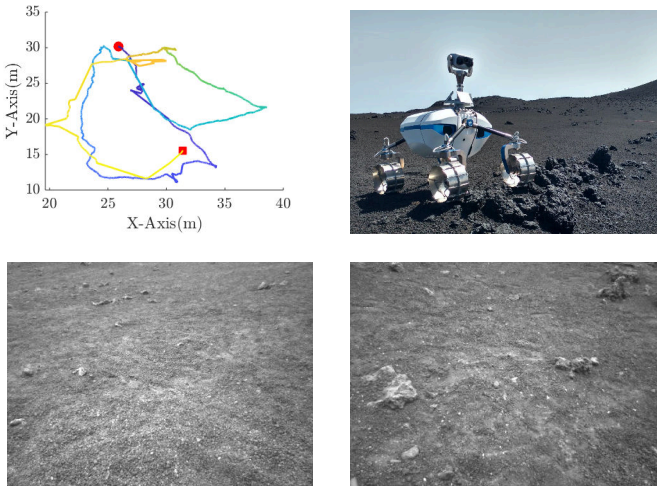

Fig. 4. From top left and clockwise: D-GPS track of the test environment (figure from [2]), the Lightweight Rover Unit (LRU) on Mt. Etna (Credits to DLR, CC-BY 3.0), two example views from the left camera within the test dataset.

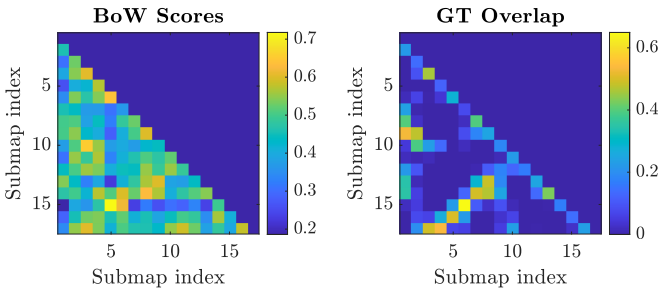

Fig. 5. BoW scores between submaps from the an Etna sequence (left) compared with the spatial overlap (right). BoW scores are expressed as cosine similarity of BoW vectors, overlap is expressed as the Intersection over Union (IoU) of submap bounding boxes. Note how the hotspots in the BoW scores correspond to regions of high overlap, e.g. the diagonal from (16, 3) to (13, 8) or the scores from (8, 0) to (10, 3).

(ROBotic Exploration in eXtreme environments) demonstration mission in 2017 [10]. The volcanic landscape offers a very complex scenario for place recognition tasks. The repetitiveness and ambiguity of the visual appearance challenges the capability of modern visual localization pipelines to recollect similar images from previouly visited places. The viewpoint of the LRU camera, looking predominantly downward towards the ground for obstacle avoidance and accurate mapping, also does not help in finding unique structures and features at longer ranges. We refer to [2] for an introduction to the four sequences used also in the context of this work.

GPGM-SLAM builds upon our submap-based SLAM system [4] replacing its loop closure detection module, which is purely based on point cloud registration. In the former loop closure detector 3D C-SHOT features are computed on keypoints extracted from high curvature regions [11] and valid rigid rototranslation models are selected through Hough3D clustering. This approach is especially challenged in the context of outdoor planetary-like environments such

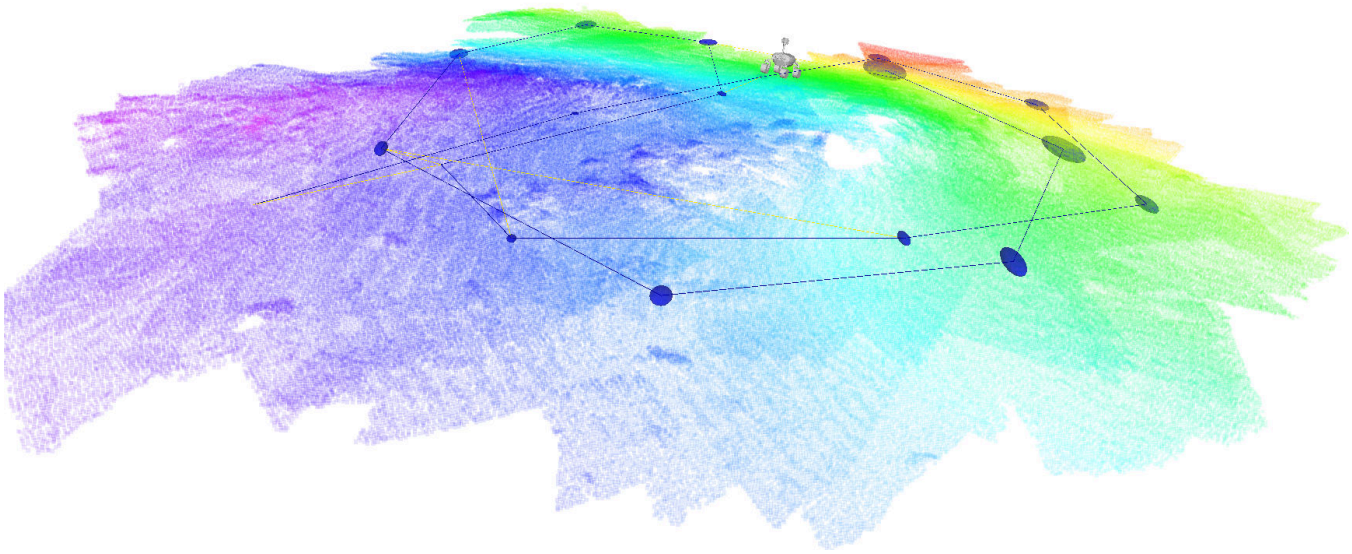

Fig. 6. Visualization of a GPGM-SLAM session on the test Etna sequence. Blue ellipses are origins of submaps with covariance, blue lines connect consequent submaps through VIO constraints and yellow lines are validated loop closures.

as the volcanic landscapes on Mt. Etna. As visible in Fig. 3, in fact, matching submaps present minimal structure and the noise from dense stereo makes it difficult to compute repeatable keypoints and local refence systems.

We compare qualitatively the performance of GPGM-SLAM with the original pipeline from [4], [11] observing the number of established loop closures. This is not only an indicator of pose estimation accuracy but also suggests a higher likelihood of establishing loop closures in challenging scenarios, where other approaches might fail. Fig. 4 shows the D-GPS track of the LRU rover on the test scenario as well as example camera views and pictures of the volcanic environment. With reference to [2], the test scenario is Etna1, while all datasets from Etna1 to Etna4 have been used to build the vocabulary of SIFT features for the Bag-of-Words framework. Within the test dataset, the LRU rover drives autonomously through a set of waypoints and maps an area of approximately 440 m^2 in 25 minutes, revisiting previous locations in two parts of the trajectory, see Fig. 5. Many overlapping submaps, however, do not contain enough structure to be matched and half of them are built while travelling from opposite directions. Nevertheless, GPGM-SLAM establishes between 3 and 4 correct loop closure constraints while our previous map matching strategy establishes between 1 and 2 loop closures. The variability is due to the RANSAC components coping with high fractions of outliers. Note that false loop closures are present in Fig. 6 as a result of brute-force matching between SIFT descriptors without applying fixed distance thresholds. However, thanks to the Cauchy loss function applied to the factors in the non-linear graph [4], their effect is nullified.

VI. Conclusions

In this short paper we presented an overview of GPGM-SLAM, a submap-based SLAM system based on Gaussian Process Gradient Maps targeted at stereo vision systems. The main benefit of employing Gaussian Process inference on elevation is to cope with noise and incompleteness in the submap point clouds, allowing to robustly establish loop closures where visual-only or structure-only SLAM might fail due to changing viewpoints and visual aliasing. SKI GP regression allowed us both to generate GPGMaps at a higher resolution than in our previous work [2] and to run the system online on recorded data. As future developments of this pipeline we plan to further speed-up the generation of GPGMap and optimize the BoW candidate selection scheme to be able to cope with limited vocabulary size and unseen visual words. Furthermore, we plan to thoroughly test and compare the performances of GPGM-SLAM with respect to the state of the art in visual SLAM on several datasets from Moon-like and Mars-like environments.

Acknowledgment

This work was supported by the UA-DAAD 2018 Australia-Germany Joint Research Cooperation Scheme, project COSMA (contract number 57446007) and the project ARCHES (contract number ZT-0033).

References

[1] C. Cadena, *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2] C. Le Gentil, *et al.*, "Gaussian process gradient maps for loop-closure detection in unstructured planetary environments," 2020.

[3] A. Wilson *et al.*, "Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP)," *Proceedings - International Conference on Machine Learning*, 2015.

[4] M. J. Schuster, *et al.*, "Distributed stereo vision-based 6d localization and mapping for multi-robot teams," *Journal of Field Robotics*, vol. 36, no. 2, pp. 305–332, 2019.

[5] M. J. Schuster, *et al.*, "Multi-robot 6d graph slam connecting decoupled local reference filters," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 5093–5100.

[6] M. Kaess, *et al.*, "isam2: Incremental smoothing and mapping using the bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.
[7] S. Särkkä, "Linear operators and stochastic partial differential equations in Gaussian process regression," in *International Conference on Artificial Neural Networks (ICANN)*, 2011, pp. 151–158.
[8] D. Eriksson, *et al.*, "Scaling gaussian process regression with derivatives," in *Advances in Neural Information Processing Systems*, 2018, pp. 6867–6877.
[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[10] A. Wedler, *et al.*, "First results of the ROBEX analogue mission campaign: Robotic deployment of seismic networks for future lunar missions," in *68th International Astronautical Congress (IAC)*, 2017.
[11] R. Giubilato, *et al.*, "Relocalization with submaps: Multi-session mapping for planetary rovers equipped with stereo cameras," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 580–587, 2020.