



Occupancy grid mapping for rover navigation based on semantic segmentation

Sebastiano Chiodini^{1,2}, Marco Pertile^{1,2}, Stefano Debei^{1,2}

¹ CISAS "Giuseppe Colombo" - Università degli Studi di Padova, Via Venezia 15, 35131 Padova, Italy

² Dipartimento di Ingegneria Industriale - Università degli Studi di Padova, Via Gradenigo 6/a, 35131 Padova, Italy

ABSTRACT

Obstacle mapping is a fundamental building block of the autonomous navigation pipeline of many robotic platforms such as planetary rovers. Nowadays, occupancy grid mapping is a widely used tool for obstacle perception. It foresees the representation of the environment in evenly spaced cells, whose posterior probability of being occupied is updated based on range sensors measurement. In more classic approaches, the cells are updated to occupied at the point where the ray emitted by the range sensor encounters an obstacle, such as a wall. The main limitation of this kind of methods is that they are not able to identify planar obstacles, such as slippery, sandy, or rocky soils. In this work, we use the measurements of a stereo camera combined with a pixel labeling technique based on Convolution Neural Networks to identify the presence of rocky obstacles in planetary environment. Once identified, the obstacles are converted into a scan-like model. The estimation of the relative pose between successive frames is carried out using ORB-SLAM algorithm. The final step consists of updating the occupancy grid map using the Bayes' update Rule. To evaluate the metrological performances of the proposed method images from the Martian analogous dataset, the ESA Katwijk Beach Planetary Rover Dataset have been used. The evaluation has been performed by comparing the generated occupancy map with a manually segmented ortomosaic map, obtained by drones' survey of the area used as reference.

Section: RESEARCH PAPER

Keywords: Convolutional neural network; SLAM; semantic segmentation; robot vision; Martian environment

Citation: Sebastiano Chiodini, Marco Pertile, Stefano Debei, Occupancy grid mapping for rover navigation based on semantic segmentation, Acta IMEKO, vol. 10, no. 4, article 25, **Citation:** Sebastiano Chiodini, Marco Pertile, Stefano Debei, Occupancy grid mapping for rover navigation based on semantic segmentation, Acta IMEKO, vol. 10, no. 4, article 25, December 2021, identifier: IMEKO-ACTA-10 (2021)-04-25

Section Editor: Roberto Montanini, Università di Messina and Alfredo Cigada, Politecnico di Milano, Italy

Received July 26, 2021; **In final form** December 6, 2021; **Published** December 2021

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Project BIRD181070 funded by the program BIRD 2018 sponsored by the University of Padova.

Corresponding author: Sebastiano Chiodini, e-mail: sebastiano.chiodini@unipd.it

1. INTRODUCTION

The problem of perception of the environment is fundamental for safe robot navigation. Planetary rovers' exploration has some peculiarities that we do not find in other autonomous robotics applications, (1) there is no GPS system that can help with the localization process, (2) terrain assessment deals with an unstructured environment that is characterised by sharp rocks (large or small), sand, and bedrocks, which are often confused with the noise of a stereo point cloud. A review of learning-based perception and navigation methods for rescue robotics, planetary exploration, and agricultural robotics can be found in [1]. It is not only safe navigation that benefits from the autonomous navigation capabilities of a rover, but also its scientific output, as an example, the distance travelled for each

sol by the NASA MSL rover has increased from a few meters to 100 m [2]. The NASA MER rover disposed of the GESTALT (Grid-based Estimation of Surface Traversability Applied to Local Terrain) system [3], which is one of the first autonomous terrain assessment systems for planetary rovers. This system was able to detect geometric hazards such as rock, ditches, and cliffs by processing the 3D point clouds generated by the rover stereo-images; it looked mainly at geometric characteristics such as steps, slopes, and terrain roughness. An alternative method for estimating the traversability of a terrain is presented in [4], where the unevenness of the terrain is analysed by means of the power spectral density (PSD) of the surface profile as measured by a stereo camera. In [5] an evaluation system for the traversability of rough terrain for a rover based on aerial UAV survey is presented.

Looking at the context of indoor navigation, many methods involve the construction of occupancy grid maps that are updated via the Bayesian occupancy filter with the range measurements coming from LiDAR sensors. As an example, [6] presents a corridor detection method based on 2D lidar and occupancy filtering.

Thanks to semantic segmentation, which is the process of dividing an image into digital objects, it is possible to add to the map information that otherwise cannot be deduced from the three-dimensional model alone, such as if the terrain is sandy or slippery. Compared to other methods, such as classification, semantic segmentation allows for pixel-by-pixel labelling of the image, see Figure 1. This enables a more detailed interpretation of the surrounding environment.

Cadena et al. [7] says that Simultaneous Localization and Mapping is now entering its third stage, characterised by robust perception. This robustness requires the realization of robust performance, high-level understanding, resource awareness, and task-driven perception. Semantics is an important tool in the pursuit of improved robustness, intuitive visualization, and efficient human-robot-environment interaction. The term Semantic SLAM can be used to identify methods that comprise either semantic-based robustness/accuracy enhancements or semantic mapping. For details, see [8], which presents a detailed review of recent advances in semantic simultaneous localization and mapping, considering multiscaled map representation, object simultaneous localization and mapping system, and deep neural network-based SLAM.

DeepLab networks make a significant contribution to semantic segmentation, see, e.g., [9], [10], which introduce the concept of “atrous convolution” in CNN models and exhibit excellent accuracy in semantic segmentation, see [8]. In the literature, it is possible to find many works on semantic mapping in the urban environment. As an example [11] generates a dense 3D reconstruction with associated semantic labelling from stereo camera images, [12] and [13] propose a multimodal sensor-based semantic 3D mapping system using a 3D LiDAR combined with a camera. In the frame of maritime navigation, the authors of [14] present a Water Obstacle Detection network based on Image Segmentation (WODIS) for autonomous surface vehicles. Gan et al. [15] describe a method that provides a unified probabilistic model for both occupancy and semantic probabilities, producing a Bayesian continuous 3D semantic occupancy map from noisy point clouds. Wang et al. [16] describe a joint method of a priori convolutional neural networks at superpixel level and soft restricted context transfer. In [17], the authors describe a method to build a complete semantic map based on merging the segmentation results from street and satellite view images. In [18] satellite images are segmented using a SegNet network. Li et al. [19] present a fast 3D semantic mapping system based on monocular vision by fusion of localization, mapping, and scene parsing. The method is based on an improved version of DeepLab-v3+ [9], [10]. [20] presents an approach to generate an outdoor large-scale 3D dense semantic map based on binocular stereo vision and the SegNet deep learning architecture. Paz et al. [21] fuses image and pre-built point cloud map information to perform automatic and accurate labelling of static landmarks such as roads, sidewalks, crosswalks, and lanes. In this work, semantic segmentation is also performed on 2D images using the DeepLab-v3+ network [9], [10].

However, these techniques have rarely been applied to the case of planetary exploration and environments with a low number of features. This paper proposes a terrain assessment

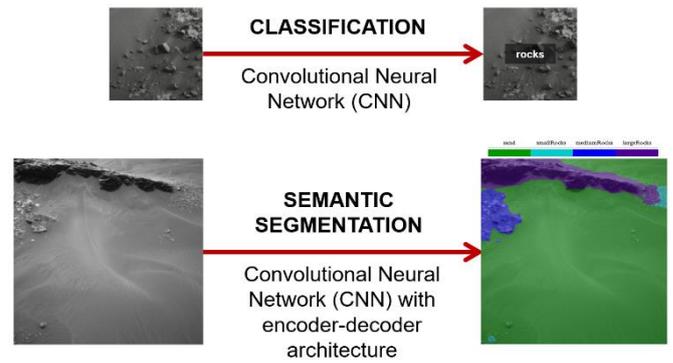


Figure 1. Compared to other methods, such as classification, semantic segmentation allows for a pixel-by-pixel labelling of the image. This enables the creation of more detailed maps.

method for Martian environment based on semantic mapping. The method takes as input a set of stereo-images which are pixel-wise labelled with the state-of-the-art Convolutional Neural Network labeller DeepLabv3+ [9]. Thanks to the stereo reconstruction of the scene, it is possible to associate the labels to the 3D points; therefore, obstacles are represented through a scan-like model [22]. All scans are combined with each other in an occupancy grid considering the trajectory travelled by the rover. The trajectory was calculated using the ORB-SLAM algorithm [23].

To evaluate the metrological performances of the proposed method, images from the public available ESA Katwijk Beach Planetary Rover Dataset [24] are used. The dataset was created for the validation of localization and navigation algorithms in Martian-like environment, and it provides trajectory and map ground truth. The method evaluation has been performed using a manually segmented ortomosaic map, taken by a drone, as reference, and the trajectory reconstruction performances have been evaluated by comparison with a differential GPS (DGPS).

The major contributions of this work are:

- the application of CNN to identify obstacles characteristic of the planetary environment;
- the generation of laser-scan like point cloud representing the obstacle;
- the application of Bayes filtering to obtain global obstacle maps;
- comparison of the generated map with the ground truth for method validation.

The paper is divided as follows: Section 2 describes the proposed terrain assessment method. In Section 3 the algorithm performance on a Martian analogous environment has been evaluated, and in Section 4 the concluding remarks are reported.

2. METHOD

The proposed method takes as input a rectified stereo image (I_l, I_r), which is used to (1) identify the obstacle with a CNN labeller (DeepLabv3+), (2) estimate the scene point cloud, and (3) estimate the rover trajectory. Afterward, the point cloud is fused with the labels to obtain a scan-like representation of the obstacles. The scans associated with the rover poses that compose the trajectory are merged into an occupancy grid map using Bayesian filtering. A diagram summarizing the various steps is shown in Figure 2.

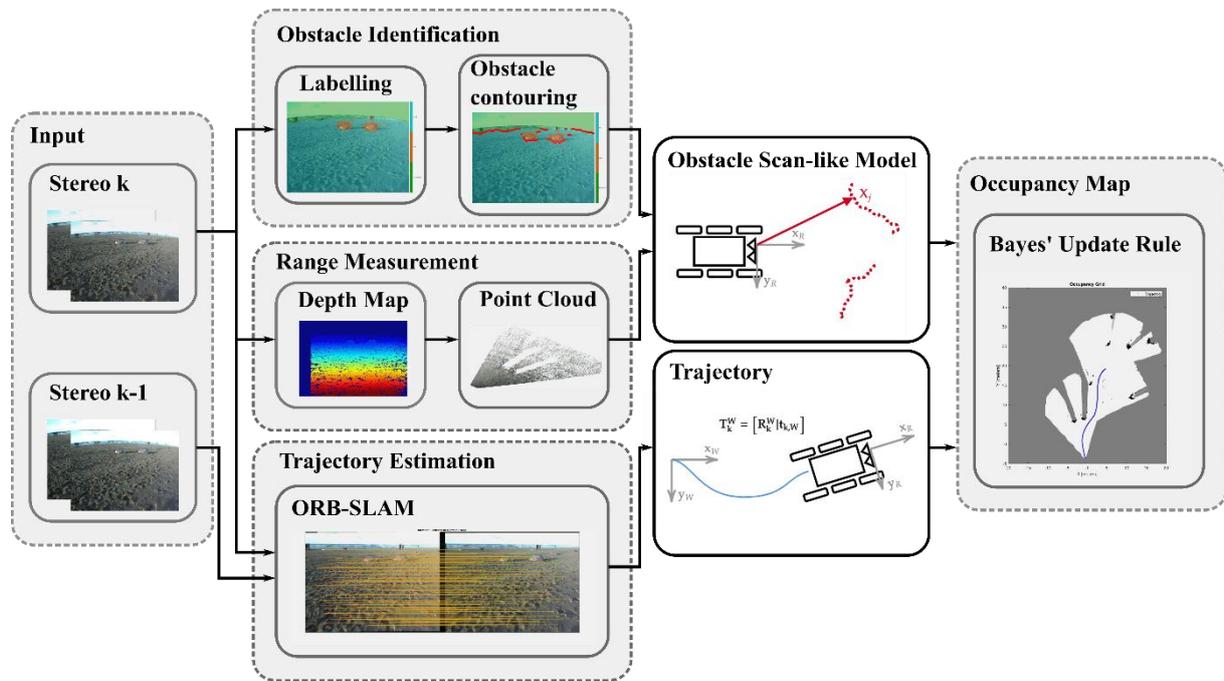


Figure 2. Scheme overview of the proposed terrain assessment algorithm. The algorithm takes as input a rectified stereo image, which is used to (1) identify the obstacle with a CNN labeller (DeepLabv3+), (2) estimate the scene point cloud, and (3) estimate the rover trajectory. Afterward, the point cloud is fused with the labels to obtain a scan-like representation of the obstacles. The scans associated with the rover poses that compose the trajectory are merged into an occupancy grid map using Bayesian filtering.

2.1. Obstacle Identification

The purpose of the obstacle identification step is labelling each pixel of the image (u_i, v_i) into obstacle or free area. Image labelling is performed using a Convolutional Neural Network (CNN) technique, as CNNs have demonstrated a significant improvement in semantic segmentation tasks. Moreover, compared to other methods, such as classification, semantic segmentation enables a more detailed interpretation of the surrounding environment. Among the state-of-the-art methods (FCN [25], SegNet [18] and U-net [26]), DeepLabv3+ [9] has been chosen because (1) it has been demonstrated to outperform similar methods in several applications, (2) it is a pre-trained network. The latter feature is necessary to reduce the training time. Network pretraining has been performed on ImageNet, which is a dataset containing more than one million images divided into 1000 classes. Pre-training allows training the deeper layer of the network with the most characteristic feature. Instead, the final part of the training phase is application dependent and is used to train the outer layer of the network.

For the context of this application, three classes have been chosen: *sand*, *rocks*, and *background*. Fehler! Verweisquelle

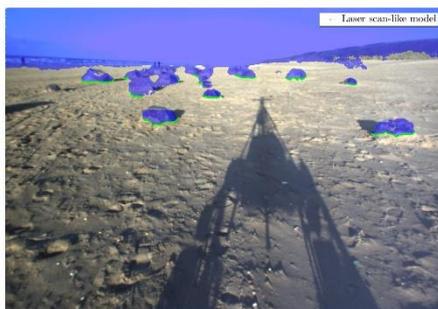


Figure 3. The 3D obstacle map is broken down to a laser scan-like model.

konnte nicht gefunden werden.2 shows the obstacle detection step applied to the images of the dataset used for testing. We have taken advantage of a pre-trained network to reduce the training images number. Data augmentation techniques have been applied to improve the training accuracy and robustness to image variation, in particular, cropping, mirroring, and resizing. The number of images used for training is 400: 50 original images and 350 obtained with data augmentation; more details are given in [27], [28], and [29].

Since it may happen that the boundaries of the identified classes may present imperfections or smudges, morphological operations have been applied to filter the identified regions. The morphological operations that have been applied are the following: binary erosion, binary dilation, and removal of a small region [30].

For collision avoidance purposes, the main interest is limited to the potential point of collision with the obstacle. For this reason, for each identified region, only the lower limit is considered, which is the one that corresponds to the intersection between the ray that starts from the camera centre and the obstacle. This contour, combined with the range measurements obtained with the method described in Section 2.2, allows us to have a scan-like representation of the obstacle (see Figure 3).

2.2. Range Measurement

First, images are stereo-rectified using camera intrinsic and extrinsic parameters. Then, the Semi-Global Block Matching algorithm [31] is used to compute the disparity map. Semi-Global Block Matching algorithm computes the scene disparity by comparing the sum of absolute differences (SAD) for each block of pixels in the image and forces a similar disparity on neighbouring blocks. By knowing the disparity map and the intrinsic parameters of the camera it is possible to estimate the depth of the scene and compute the related point cloud relative to the rover frame $(X_j = [X_j, Y_j, Z_j])$, where $j = 1, \dots, n$ with n

total number of pixels with an associated disparity. The 3D points coordinates in the left camera frame are given by:

$$\begin{aligned} X_j &= \frac{u_l Z_j}{f} \\ Y_j &= \frac{v_l Z_j}{f} \\ Z_j &= \frac{bf}{u_l - u_r}, \end{aligned} \quad (1)$$

where (u_l, v_l) and (u_r, v_r) are respectively the pixel coordinates in the left and right image, b is the camera baseline and f is the camera focal length. Each 3D point of the point cloud (\mathbf{X}_j) is associate with a pixel of the image (u_l, v_l) . Only 3D points that correspond to the lower boundary of the obstacles are used to update the occupancy grid map; see Section 2.3.

2.3. Relative Trajectory Reconstruction

Sections 2.1 and 2.2 show how to determine the position of the obstacles with reference to the rover frame. However, to build a global map of obstacles and filter the measurements belonging to different rover poses, the position of the obstacles position \mathbf{X}_i needs to be defined with respect to an absolute reference system \mathbf{X}_i^W , the W world reference frame. In this work the frame W corresponds to the first stereo-image frame rotated by the pan and tilt angles of the PTU (Pan and Tilt Unit).

To perform the trajectory reconstruction step, only stereo camera visual input has been used in combination with the state-of-the-art visual SLAM algorithm, ORB-SLAM2 [23]. For the purposes of this work, the camera pose issued by the tracking step was used, this pose belongs to SE(3) (the Special Euclidean Group in three dimension). Assuming that the rovers are mainly moving in a planar environment, the SE(3) pose has been transformed into the corresponding SE(2) pose $\mathbf{T}_k^W = [\mathbf{R}_k^W | \mathbf{t}_{k,W}]$, which for completeness is reported in the following relation:

$$\mathbf{T}_k^W = \begin{bmatrix} \cos(\psi_k^W) & -\sin(\psi_k^W) & t_{x(k,W)} \\ \sin(\psi_k^W) & \cos(\psi_k^W) & t_{y(k,W)} \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

where $\psi_{k,W}$ is the absolute heading angle and $t_{x(k,W)}$ $t_{y(k,W)}$ is the rover absolute position.

Finally, obstacle scan line points are transformed from the rover frame to the W frame using the following equation:

$$\mathbf{X}_i^W = \mathbf{R}_k^W \mathbf{X}_i + \mathbf{t}_{k,W}. \quad (3)$$

2.4. Sensor Fusion

The last step is the update of the occupational grid. The probability of each grid cell being occupied is updated following the standard Bayes update rule [29] [32] and using the obstacle scan \mathbf{X}_i^W . In the experimental part, a grid resolution of 0.2×0.2 m has been considered. Assuming that the cells of the grid map are independent from each other, and given a series of rock observations $\mathbf{z}_{1:j}$, the probability belief of a single cell to be occupied by an obstacle or not $p(m_{x,y} | \mathbf{z}_{1:j})$, is reported in Equation (4).

$$p(m_{x,y} | \mathbf{z}_{1:j}) = \frac{p(m_{x,y} | \mathbf{z}_j) p(\mathbf{z}_j) p(m_{x,y} | \mathbf{z}_{1:j-1})}{p(m_{x,y}) p(\mathbf{z}_j | \mathbf{z}_{1:j-1})}, \quad (4)$$

where $p(m_{x,y} | \mathbf{z}_j)$ is the inverse measurement model of the depth data retrieved from the stereo-camera, $p(m_{x,y} | \mathbf{z}_{1:j-1})$ is the depth measurements of the previous rover poses, and $p(m_{x,y})$ is the prior map. To avoid difficult-to-calculate probabilities, we use the binary Bayes filter in log-odds form:

$$l_j = l_{j-1} + \log \frac{p(m_{x,y} | \mathbf{z}_j)}{1-p(m_{x,y} | \mathbf{z}_j)} + \frac{p(m_{x,y})}{1-p(m_{x,y})}, \quad (5)$$

where $l_j = \log \frac{p(m_{x,y} | \mathbf{z}_{1:j})}{1-p(m_{x,y} | \mathbf{z}_{1:j})}$. The log odd form of the inverse measurement model $\log \frac{p(m_{x,y} | \mathbf{z}_j)}{1-p(m_{x,y} | \mathbf{z}_j)}$ an occupancy value l_{occ} assigns to all cells within the 3D labelled points \mathbf{X}_j . In experiments, the occupied threshold is $l_{occ} = 0.65$, the free threshold is $l_{free} = 0.2$. The grid is initialized without prior knowledge of the map $p(m_{x,y}) = 0.5$.

3. EXPERIMENTAL RESULTS

The metrological performances of the proposed method have been evaluated using the ESA Katwijk Beach Planetary Rover Dataset [24], which provides images analogous to Mars. Of the overall dataset, the LocCam stereo images have been used for semantic maps generation, site ortomosaic combined with differential GPS has been used as ground truth. The characteristics of the sensors used in the experimental part are summarized in Table 1. The DGPS has been used to register the generated maps to the ortomosaic taken by the drone, which has been used as a map reference. The ground truth of the map has been obtained by manually labelling the ortomosaic drone, and the ground truth of the trajectory is given directly by the DGPS.

Figure 4 shows the occupancy grid map generated with the proposed method, the occupancy grid is superimposed on the drone image used for performance evaluation. The metrics normally used to evaluate object detection algorithms were used: the accuracy, the Intersection over Union (IoU) and the F1 score:

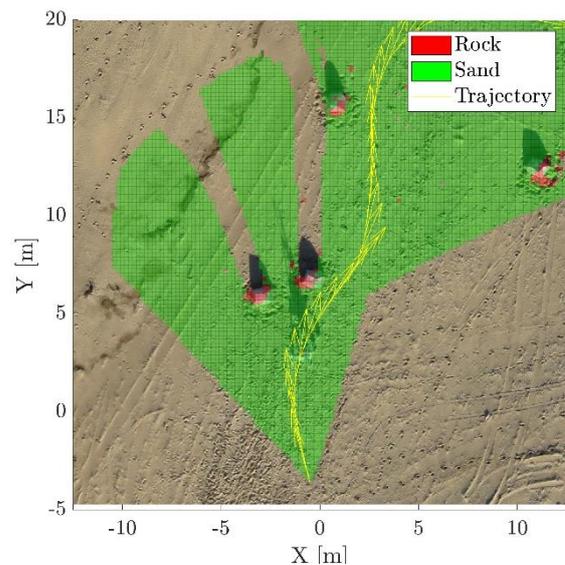


Figure 4. Occupancy grid map generated with the proposed method. Free cells are shown in green, obstacles-occupied cells in red, and yellow arrows show the trajectory of the rover. The occupancy grid is superimposed on the drone image used for performance evaluation.

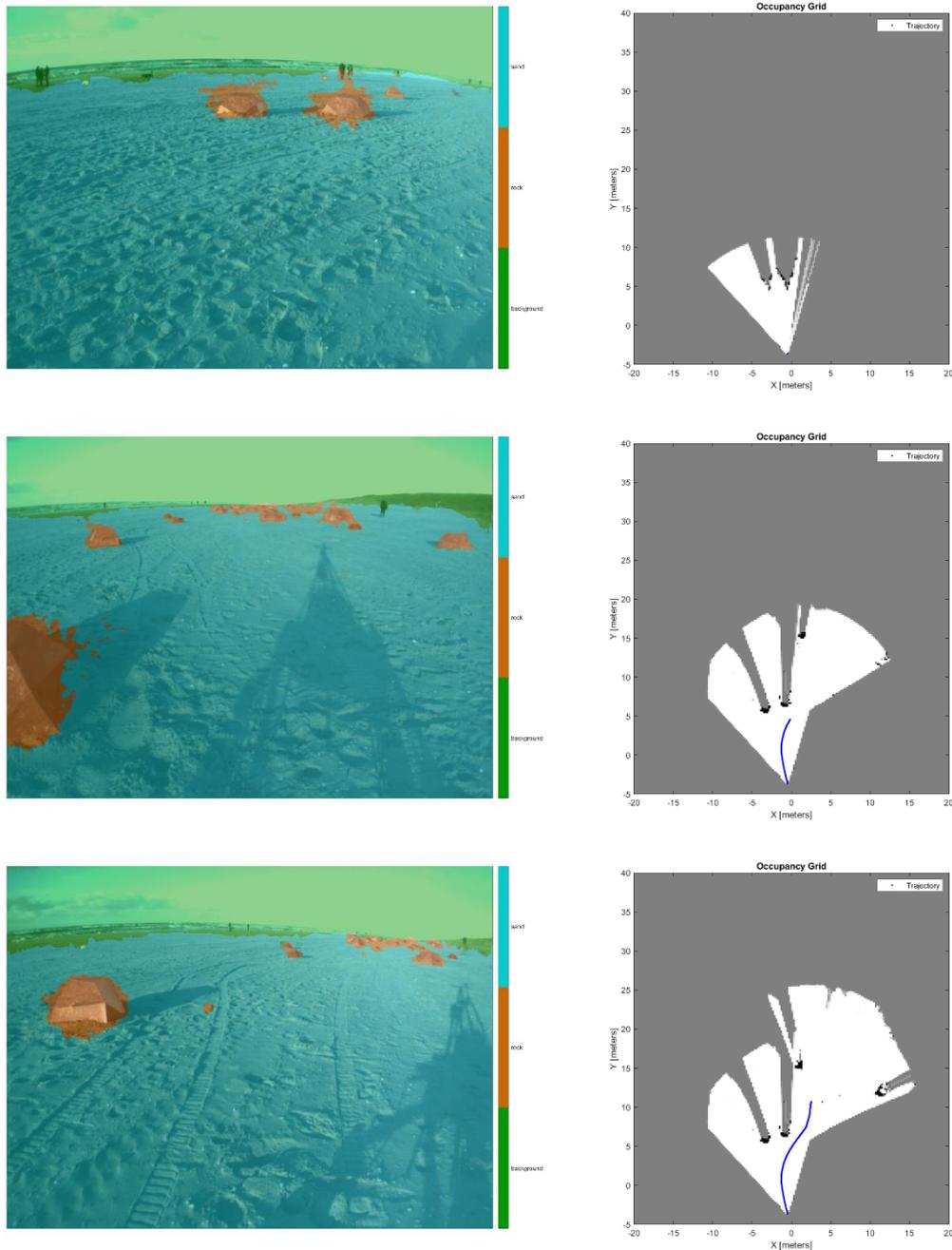


Figure 5. Left column: labelled images with rock, sand, and background classes. Right column: occupancy grid map and trajectory estimated with the proposed method. From top to bottom successive images of the sequence.

Table 1. Sensors characteristics.

Sensor	Description	Data Logged
LocCam	PointGery Bumblebee2 (BB2-08S2C-38) 12 cm baseline stereo camera.	1024 × 768 images
RTK GPS	Trimble BD 970 Receiver with Zephyr Model2 Antenna (rover) Trimble BX 982 Receiver with Zephyr Geodetic Antenna (base station).	Latitude, Longitude, and Altitude expressed on WGS84 ellipsoid

Table 2. Global mapping performances in terms of accuracy and IoU metrics.

Labeller	Trajectory	Accuracy	IoU
DeepLabv3+	ORB-SLAM2	0.987	0.282

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (6)$$

$$IoU = \frac{TP}{TP + FP + FN}, \quad (7)$$

where TP represent the True Positives, TN the True Negatives, FP the False Positives, and FN the False Negatives. In the case of an ideal classifier $FP = FN = 0$, thus the accuracy metric would be equal to 1. IoU is the ratio of correctly classified cells over the sum of the total number of cells labelled and cells classified. Table 2 summarises the performances (Accuracy and IoU) of the proposed method.

The left column of Figure 5 shows the images labelled with rock, sand and background classes, and the right column shows

the occupancy grid map and trajectory estimated with the proposed method. As it is possible to observe in the first row of Figure 5 the two rocky obstacles are a few meters away from the camera and not easily recognizable, however, they are correctly labelled by the DeepLabv3+ labeller. The effectiveness of the labeller can also be observed in the middle and bottom row of Figure 5, although the shadows of the rover are present in both rock and sand, it does not affect the performance of the label.

4. CONCLUSIONS

In this paper, a hazard mapping method for planetary rovers' navigation is presented. The method is based on occupancy grid mapping. The posterior probability of the grid cells is updated using the inverse measurement model of a scan-like obstacle detector. Objects are identified as obstacles by means of the DeepLabv3+ deep neural network. The estimation of the relative pose between successive frames is carried out using the state-of-the-art visual SLAM method, ORB-SLAM. The proposed method shows the ability to produce accurate occupancy grid maps with associated label up to a dozen meters from the camera and when the rover shadow is present in the image field of view. Finally, the method has been tested on a public available dataset of a Martian analogous environment.

REFERENCES

- [1] D. C. Guastella, G. Muscato, Learning-Based Methods of Perception and Navigation for Ground Vehicles in Unstructured Environments: A Review. *Sensors* 2021, 21, 73. DOI: [10.3390/s21010073](https://doi.org/10.3390/s21010073)
- [2] M. Bajracharya, M. W. Maimone, D. Helmick, *Autonomy for Mars Rovers: Past, Present, and Future*, Computer, 41 (12) (2008), pp. 44-50. DOI: [10.1109/MC.2008.479](https://doi.org/10.1109/MC.2008.479)
- [3] M. W. Maimone, P. C. Leger, J. J. Biesiadecki, Overview of the Mars exploration rovers autonomous mobility and vision Capabilities, Proc. Of the IEEE International Conference on Robotics and Automation (ICRA) space robotics workshop, April 2007.
- [4] G. Reina, A. Leanza, A. Milella, A. Messina, Mind the ground: a power spectral density-based estimator for all-terrain rovers. *Measurement*, 151 (2020), 107136. DOI: [10.1016/j.measurement.2019.107136](https://doi.org/10.1016/j.measurement.2019.107136)
- [5] D. C. Guastella, L. Cantelli, D. Longo, C. D. Melita, G. Muscato, Coverage path planning for a flock of aerial vehicles to support autonomous rovers through traversability analysis, *Acta IMEKO* 8 (2019) 4, pp. 9-12. DOI: [10.21014/acta_imeko.v8i4.680](https://doi.org/10.21014/acta_imeko.v8i4.680)
- [6] L.-H. Chen, C. Peng, A Robust 2D-SLAM Technology With Environmental Variation Adaptability, *IEEE Sensors Journal* 19 (23) (2019), pp. 11475-11491. DOI: [10.1109/JSEN.2019.2931368](https://doi.org/10.1109/JSEN.2019.2931368)
- [7] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, John J. Leonard, Past, present, and future of simultaneous localization and mapping: toward the robust-perception age, *IEEE Trans. Robot.* 32(6) (2016), pp 1309–1332. DOI: [10.1109/TRO.2016.2624754](https://doi.org/10.1109/TRO.2016.2624754)
- [8] L. Xia, J. Cui, R. Shen, X. Xu, Y. Gao, X. Li, A survey of image semantics-based visual simultaneous localization and mapping: Application-oriented solutions to autonomous navigation of mobile robots, *Int. J. of Advanced Robotic Systems* (2020), pp. 1–17. DOI: [10.1177/1729881420919185](https://doi.org/10.1177/1729881420919185)
- [9] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Proc. of the European conference on computer vision (ECCV), 2018, pp. 801-818. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, Springer, Cham, vol 11211. DOI: [10.1007/978-3-030-01234-2_49](https://doi.org/10.1007/978-3-030-01234-2_49)
- [10] L.-C. Chen, G. Papandreou, I. Kokkinos, Kevin Murphy, Alan L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 40(4) (2017), pp 834–848. DOI: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184)
- [11] S. Sengupta, E. Greveson, A. Shahrokni, P. H. Torr, Urban 3d semantic modelling using stereo vision, Proc. 2013 IEEE International Conference on robotics and Automation. IEEE, 2013, pp. 580-585. DOI: [10.1109/ICRA.2013.6630632](https://doi.org/10.1109/ICRA.2013.6630632)
- [12] J. Jeong, T.S. Yoon, J.B. Park, Multimodal sensor-based semantic 3d mapping for a large-scale environment, *Expert Systems with Applications*, 105 (2018), pp. 1–10. DOI: [10.1016/j.eswa.2018.03.051](https://doi.org/10.1016/j.eswa.2018.03.051)
- [13] Z. Qiu, Y. Zhuang, F. Yan, H. Hu, W. Wang, RGB-DI Images and Full Convolution Neural Network-Based Outdoor Scene Understanding for Mobile Robots, *IEEE Transactions on Instrumentation and Measurement*, 68(1) (2019), pp. 27-37. DOI: [10.1109/TIM.2018.2834085](https://doi.org/10.1109/TIM.2018.2834085)
- [14] X. Chen, Y. Liu, K. Achuthan, WODIS: Water Obstacle Detection Network Based on Image Segmentation for Autonomous Surface Vehicles in Maritime Environments, *IEEE Transactions on Instrumentation and Measurement*, 70 (2021), pp. 1-13, 7503213. DOI: [10.1109/TIM.2021.3092070](https://doi.org/10.1109/TIM.2021.3092070)
- [15] L. Gan, R. Zhang, J.W. Grizzle, R.M. Eustice, M. Ghaffari, Bayesian Spatial Kernel Smoothing for Scalable Dense Semantic Mapping, *IEEE Robotics and Automation Letters*, 5(2) (2020). DOI: [10.1109/LRA.2020.2965390](https://doi.org/10.1109/LRA.2020.2965390)
- [16] Q. Wang, J. Gao, Y. Yuan, A Joint Convolutional Neural Networks and Context Transfer for Street Scenes Labeling, *IEEE Trans. on Intelligent Transportation Systems*, 19(5) (2018), pp. 1457-1470. DOI: [10.1109/ITITS.2017.2726546](https://doi.org/10.1109/ITITS.2017.2726546)
- [17] V. Balaska, L. Bampis, I. Kansizoglou, A. Gasteratos, Enhancing satellite semantic maps with ground-level imagery, *Robotics and Autonomous Systems*, 139 (2021) 103760. DOI: [10.1016/j.robot.2021.103760](https://doi.org/10.1016/j.robot.2021.103760)
- [18] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder–decoder architecture for image segmentation, *IEEE Trans. Pattern Analysis Mach. Intell.* 39 (12) (2017), pp. 2481–2495. DOI: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615)
- [19] X. Li, D. Wang, H. Ao, R. Belaroussi, D. Gruyer, Fast 3D Semantic Mapping in Road Scenes, *Appl. Sci.* 9(4) (2019), pp. 631. DOI: [10.3390/app9040631](https://doi.org/10.3390/app9040631)
- [20] Y. Yang, F. Qiu, H. Li, L. Zhang, M.-L. Wang, M.-Y. Fu, Large-scale 3D Semantic Mapping Using Stereo Vision, *International Journal of Automation and Computing*, 15(2) (2018), pp. 194-206. DOI: [10.1007/s11633-018-1118-y](https://doi.org/10.1007/s11633-018-1118-y)
- [21] D. Paz, H. Zhang, Q. Li, H. Xiang, H.I. Christensen, Probabilistic Semantic Mapping for Urban Autonomous Driving Applications, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) October 25-29, 2020. DOI: [10.1109/IROS45743.2020.9341738](https://doi.org/10.1109/IROS45743.2020.9341738)
- [22] F. Andert, Drawing stereo disparity images into occupancy grids: Measurement model and fast implementation, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2009, pp. 5191–5197. DOI: [10.1109/IROS.2009.5354638](https://doi.org/10.1109/IROS.2009.5354638)
- [23] Raúl Mur-Artal, Juan D. Tardós, ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. *IEEE Transactions on Robotics*, 33 (5) (2017), pp. 1255-1262. DOI: [10.1109/TRO.2017.2705103](https://doi.org/10.1109/TRO.2017.2705103)

- [24] R. A. Hewitt, E. Boukas, M. Azkarate, M. Pagnamenta, J. A. Marshall, A. Gasteratos, G. Visentin, The katwijk beachplanetary rover dataset, *The International Journal of Robotics Research*, 37(1) (2018), pp. 3-12.
DOI: [10.1177/0278364917737153](https://doi.org/10.1177/0278364917737153)
- [25] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.
- [26] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Proc. International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234-241.
DOI: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [27] S. Chiodini, L. Torresin, M. Pertile, S. Debei, Evaluation of 3D CNN Semantic Mapping for Rover Navigation, *Proc. IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 2020, pp. 32-36.
DOI: [10.1109/MetroAeroSpace48742.2020.9160157](https://doi.org/10.1109/MetroAeroSpace48742.2020.9160157)
- [28] L. Torresin, Sviluppo ed applicazione di reti neurali per segmentazione semantica a supporto della navigazione di rover marziani, Master's Thesis Report, 2019.
- [29] S. Chiodini, M. Pertile S. Debei, Semantic Mapping for Rover Navigation, *Proc. IV Forum Nazionale delle Misure*, 2020.
- [30] R. Van den Boomgard, R. van Balen, Methods for Fast Morphological Image Transforms Using Bitmapped Images, *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, 54(3) (1992), pp. 254-258.
DOI: [10.1016/1049-9652\(92\)90055-3](https://doi.org/10.1016/1049-9652(92)90055-3)
- [31] H. Hirschmuller, Accurate and efficient stereo processing by semi-global matching and mutual information, *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 05)*, 2005, pp. 807-814.
DOI: [10.1109/CVPR.2005.56](https://doi.org/10.1109/CVPR.2005.56)
- [32] S. Thrun, Probabilistic robotics, *Communications of the ACM*, 45(3) , 2002, pp. 52-57.
DOI: [10.1145/504729.504754](https://doi.org/10.1145/504729.504754)