DOI: https://doi.org/10.52651/sam.a.2023.2.5-15

ANALYSIS OF STATE-OF-THE-ART SOFTWARE SOLUTIONS FOR UNSTRUCTURED OUTDOOR ENVIRONMENT PERCEPTION IN GOAL-ORIENTED NAVIGATION OF OFF-ROAD UNMANNED GROUND VEHICLES

Matej VARGOVČÍK, Peter PÁSZTÓ, Marian KĽÚČIK, Martin SMOĽÁK, Patrik ŠTEFKA, Jakub LENNER

Abstract: This paper presents a comprehensive analysis of state-of-the-art software solutions designed for the perception of unstructured outdoor environments, with a focus on their applicability in goal-oriented navigation for off-road unmanned ground vehicles (UGVs). The analysis encompasses the evaluation of various sensors for UGV navigation, comparison of methods for extracting perceptual information from sensor data, and the development of a perception system based on the findings. The proposed system leverages advancements in simultaneous localization and mapping (SLAM), visual odometry, and traversability segmentation using 3D LiDAR and visual systems. The analysis suggests a phased approach for the development of autonomous navigation, where the initial phase relies on complex SLAM and teleoperation assistance, and the subsequent phase introduces an autonomous mode utilizing semantical terrain representation.

Keywords: UGV; Perception; Autonomous; Navigation; SLAM; Off-road.

1 INTRODUCTION

In this study, we explore the possibilities of extending navigation algorithms currently employed and under development by our company, RoboTech Vision, to adapt them for off-road environments. The unique challenges presented in these environments, characterized by a lack of regular structures and clearly distinguishable roads, necessitate a comprehensive analysis.

Given the abundance of state-of-the-art perception systems available in technical science, it is required to delineate the types of perceptual information crucial for successful navigation towards specified goals and to specify input data from which the information will be extracted. Section 2 defines the terrains that the UGVs will operate on and the sensors suitable for navigating these terrains.

Section 3 compares various existing methods for extracting relevant information from sensor data, evaluating their effectiveness in off-road environments. The methods are selected based on their relevance for the particular tasks and environments and their ranking in appropriate benchmarks addressing similar environments and use-cases.

Finally, Section 4 outlines a perception system based on this analysis and its integration into a complex traversability map creation and navigation system.

2 SITUATION

2.1 Environment

The target environment for UGV operation is the territory of Slovakia, and the analysis will be thus conducted on datasets from forest and field terrain (Figure 1). The primary objective of autonomous motion is to navigate through unpaved roads and paths, taking into account potential challenges such as partial erosion or coverage by vegetation and leaves. It's important to note that the analysis excludes extreme situations, such as motion on rocks, in brushwoods, or in swamps. Additionally, we assume a simplified scenario without the presence of snow and extreme weather conditions.

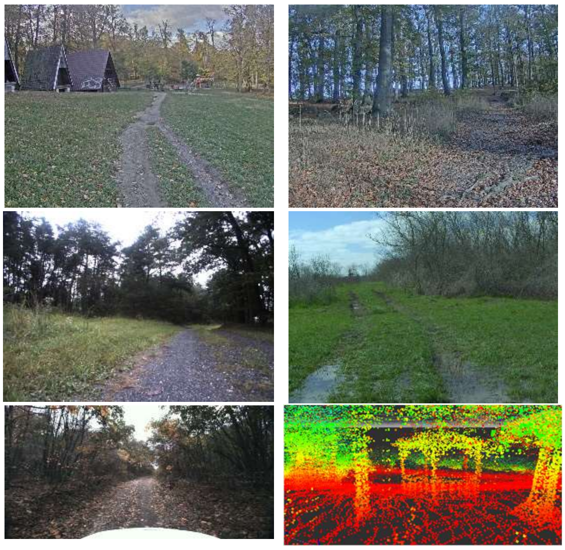

Fig. 1 Expected environment - camera images, bottom-right: LiDAR point-cloud
Source: author, datasets RUGD [1], RELLIS-3D [2], and Yamaha-seg [3].

2.2 Sensors

For navigation, UGV can use the following sensoric equipment:

- Odometry from incremental sensors in UGV chassis;
- IMU – accelerometer, gyroscope;

- Visual systems – RGB/IR cameras, depth cameras (TOF cameras, structured light cameras, stereo cameras);
- 3D LiDARs, in fog or smoke conditions extendible by MPR radars [4];
- Global satellite-based systems receivers (GPS, Glonass, Galileo).

Whereby two possible operation modes will be considered:

- Mode with full sensoric employment - appropriate for fast and reliable transfer;
- STEALTH mode - with employment of passive sensors only (no LiDARs, TOF cameras, and structured light cameras).

For the purpose of this work, we recorded a dataset with our RTV Crawler mobile robotic platform, equipped with locomotive odometry sensors (ams-OSRAM), IMU (Microstrain), a monocular front camera (Axis), a stereoscopic camera (Intel Realsense), a 16-channel 3D LiDAR (Velodyne), and a Glonass/GPS Receiver (NovAtel). The creation of this dataset aimed to include sensors of interest missing from RUGD [1] and Yamaha-seg [3] datasets or environments of interest missing from RELLIS-3D [2] dataset. Additionally, it allowed testing existing methods against new samples on which these methods were not originally developed. The sensor setup and dataset environments are illustrated in Figure 2.

Fig. 2 RTV Crawler mobile platform with sensoric setup in different environments during recording of our dataset
Source: author.

3 ANALYSIS OF PERCEPTION SYSTEM TASKS

Based on the sensors used and environmental conditions, we selected tasks to be performed by the perception system. The tasks will be described in the following sections, along with selected solutions, which will be tested on various datasets.

3.1 Simultaneous Localization and Mapping (SLAM) Using LiDAR

Localization is one of the basic requirements of autonomous UGVs, as information about the goal position relative to the UGV is essential during motion planning and execution. Additionally, localization is crucial for completing current UGV surroundings, which cannot be fully sensed from the current UGV pose, using a map created from previous poses.

Mapping involves storing sensor data (or processed information) in different poses of the UGV, which are mutually localized against each other. I.e. the essential and most complex element of mapping is correct and accurate localization. Therefore, in practice, mapping and localization are connected into one system - SLAM. When needed, mapping can be excluded, and localization is performed on a previously prepared static map.

Localization comprises two perceptional tasks: localization odometry and loop closure, which are combined using a pose graph.

3.1.1 Localization Odometry Using a 3D LiDAR

In contrast with odometry from incremental sensors in the UGV chassis (locomotive odometry), where motional information is retrieved from motor revolutions in a straightforward manner, localization odometry compares the sensorical description of the environment between consecutive poses of the UGV. UGV displacement is computed based on the relative motion of objects against the UGV. This type of odometry is resistant to skids and is usually significantly more accurate than locomotive odometry. Additionally, multiple types of odometry (LiDAR, visual, locomotive) can be combined to robustify the overall system, making it resistant to errors of individual odometries.

In the case of 3D LiDAR data (point-clouds), we used NDT registration [5] for pairing consecutive point-clouds. This registration, designed to match similar point-clouds, is suitable for the purpose of odometry due to its speed. It is designed to match similar point-clouds only, but despite this limitation, it is sufficient since similar consecutive captures are compared in the case of LiDAR odometry. Result of NDT odometry is illustrated in Figure 3.

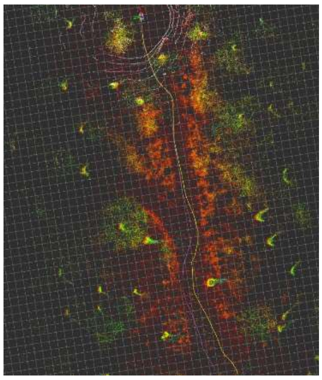

Fig. 3 NDT LiDAR odometry (yellow) compared to ground-truth UGV trajectory (purple) – locally it is highly accurate and consistent, although (like in the case of any odometry) its deviation increases gradually
Source: author.

3.1.2 Loop Closure

Due to the gradually increasing deviation of odometry, it is important to recognize previously mapped sites, when visited again, to interconnect the map. Site descriptors have been developed in computer science to optimize recognition and make it realizable in real-time.

In the analysis, we tested the applicability of ScanContext [6, 7] descriptor in off-road terrain, which proved suitable for such environment (Figure 4). For refining the mutual transformation between the recognized site and the actual pose, we used the GICP [8] registration algorithm, which is slower but more robust than NDT [5] registration used for LiDAR odometry (Figure 5).

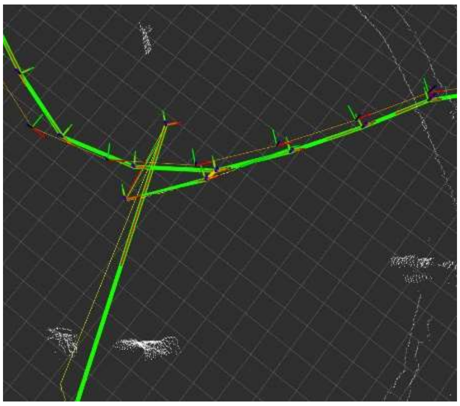

Fig. 4 A successful loop-closure marked by a short thick orange line between two nodes of a pose graph
Source: author.

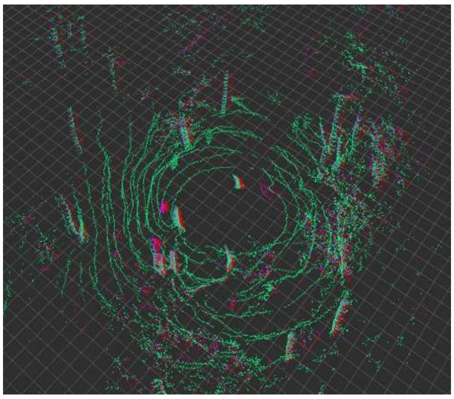

Fig. 5 Refining of the resulting transformation using GICP - current point-cloud transformed according to ScanContext recognition (red), point-cloud of the recognized site (green), point-cloud after transformation refinement (purple)
Source: author.

Final map shape is maintained by a pose graph [9], which uses odometry, loop-closure, GPS, and IMU measurements as input (Figure 6). Figure 7 depicts a final localization map created by placing LiDAR measurements on particular poses of the graph.

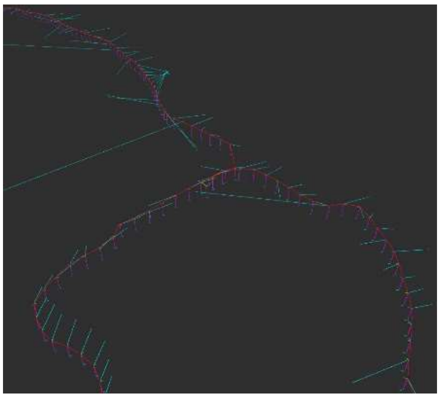

Fig. 6 Pose graph with odometry (red), loop-closure (orange), GPS (cyan), and IMU (purple) measurements
Source: author.

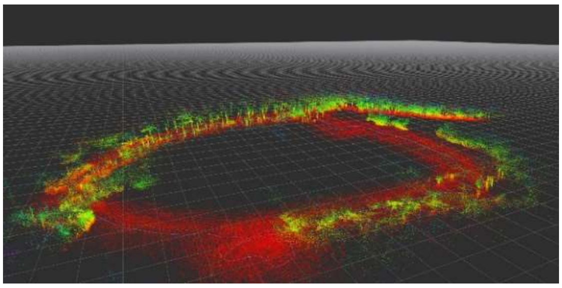

Fig. 7 Resulting localization map
Source: author.

In case that during development ScanContext appears as insufficient, it is additionally possible to use a machine-learning based method CVTNet [10], which shows significantly higher efficiency in benchmarks (although presumably requires higher maintenance due to requirement of dataset labeling and training).

3.2 Traversability Segmentation Using 3D LiDAR

The off-road environment includes a high amount of formations consisting of soft vegetation (e.g., grass clusters), which can appear as obstacles from a trivial view at point-cloud data, although they are actually traversable. For distinguishing such formations from real obstacles, it is appropriate to use modern neural-network methods with intuition-like attributes. We tested the operability of point-cloud neural networks against our datasets. DeepLabv3 ResNet101 [11] trained on TraversabilityClouds [12, 13] dataset proved successful (Figure 8). Good results were achieved also by SalsaNext [14] model trained on RELLIS-3D [2] dataset. However, it was observed that although the network had been trained for detailed classification (trees, soft vegetation, ground, puddle, human), in a different environment, it was able to correctly distinguish only two classes - trees and everything else (Figure 9).

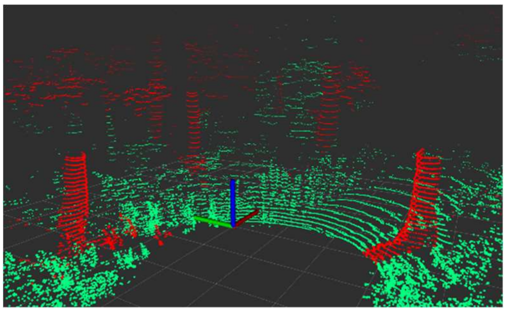

Fig. 8 Traversability estimation of a point-cloud using DeepLabv3 ResNet101/TraversabilityClouds
Source: author.

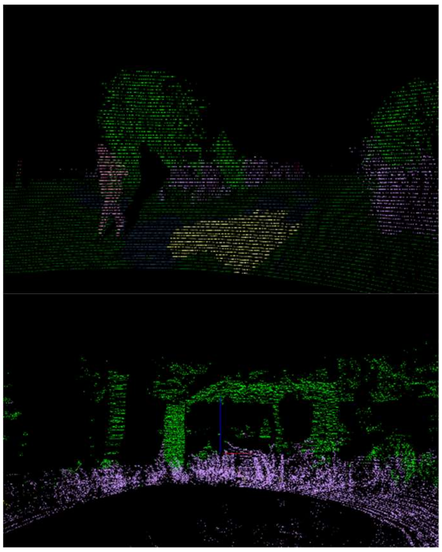

Fig. 9 SalsaNext - top: detailed segmentation in a familiar environment of RELLIS-3D dataset; bottom: segmentation of our data
Source: author.

3.3 Depth Estimation from Stereo-Pair of Images

In STEALTH mode, when active sensors measuring distances of points in the environment cannot be used, it is required to estimate the third dimension directly from camera images to project detected obstacles and terrain elements onto the traversability map. Depth estimation from a stereo-pair of images is a problem that has been targeted by computer vision science for a long time, resulting in a large number of different approaches, from traditional methods to modern ones with the use of machine learning.

Since a wider baseline between cameras of the stereoscopic system had proven as more appropriate for a complete comparable overview of depth estimation methods in off-road environments, we decided to prefer Nerian stereo camera data samples [15] and RELLIS-3D [2] dataset over our recorded dataset for this particular case.

From Figure 10, it is evident that strong randomly structured textures can induce minor problems for traditional methods that use mathematically defined similarity between elements of the left and right camera image, but these problems can be eliminated in postprocessing by despeckling resulting images. Methods with neural networks offer, thanks to their intuition-like attributes, higher consistency of output, although their results differ in quality as well.

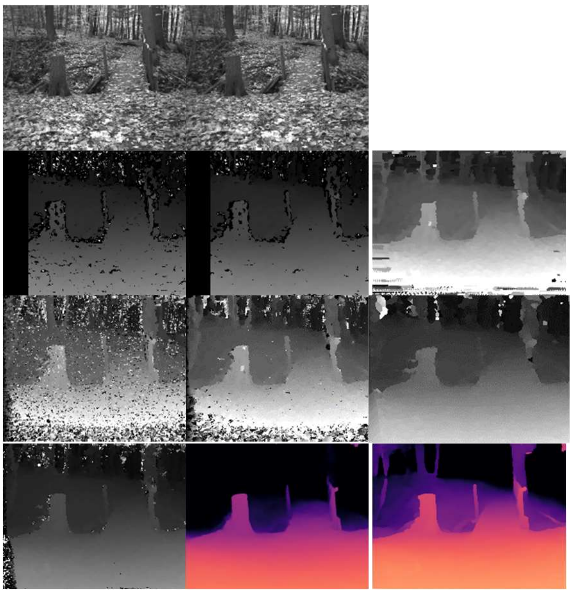

Fig. 10 Depth estimation. First row (left-to-right): left camera image, right camera image [15]; Second row: depth estimated by BM [16], SGBM [17], ASW [18] algorithms; Third row: AD-Census V1 [19], SAD [20], ZNCC [21]; Fourth row: SGM [22], ACVNet [23] trained on Sceneflow [24] dataset, HITNet [25] trained on Middlebury [26, 27, 28, 29, 30] dataset
Source: [15] (input data), author (processing).

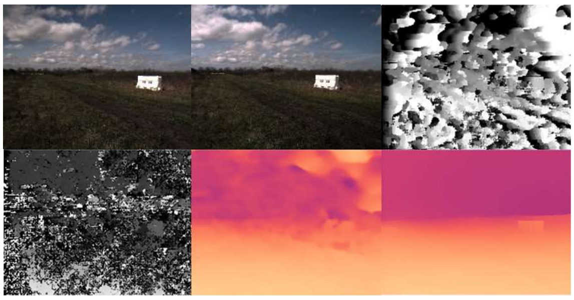

Fig. 11 Depth estimation from an uncalibrated stereo-pair of images. First row (left-to-right): left camera image, right camera image [2], depth estimated by ASW algorithm; Second row: SGM, ACVNet trained on KITTI [31, 32] dataset, HITNet trained on Middlebury dataset
Source: [2] (input data), author (processing).

To verify robustness of algorithms against calibration errors (which can occur e.g. due to thermal expansion) we tested the methods on poorly calibrated images as well (Figure 11). Among the analyzed methods, the only successful one was HITNet [25], which was able to eliminate the sky, deal with grass texture, and correctly differentiate objects in front of the cameras. Despite the compactness of the results, it is suggested to automatically recalibrate cameras [33] during the operation of the UGV. This ensures that the depth data accurately describe the spatial representation of the environment.

3.4 SLAM Using Visual Systém

The addition of visual SLAM into the perception system is desirable for two reasons:

- In STEALTH mode, UGV cannot take advantage of LiDARs; therefore, the use of LiDAR SLAM (Section 3.1) is not possible.
- Point-cloud data do not provide enough information value in a structurally monotonous environment to sustain fully operational localization (Figure 12).

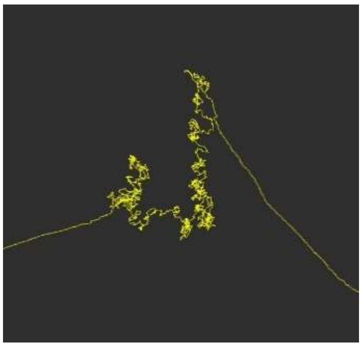

Fig. 12 Failure of LiDAR odometry in the middle of a grass field
Source: author.

3.4.1 Visual Odometry

In the analysis, we tested a basic odometry from RTAB-Map [34] system, using a stereo pair of cameras. The odometry uses BM [16] or SGBM [17] for depth estimation of camera pixels, which was sufficient for successful SLAM in an urban environment, however, did not prove as usable in off-road (Figure 13). The operability of this odometry can be improved using a more consistent depth estimation system (see Section 3.3).

Fig. 13 SLAM using RTAB-Map [30] basic odometry - top: success in urban environment, bottom: failure in grass field in Rellis-3D [2]
Source: author.

Another option is to use OpenVINS [35] visual odometry system (also supported by RTAB-Map), which uses one or more cameras as input and additionally uses IMU data for a complex fusion. It was approved by us (Figure 14) and also in tests provided by authors of ROOAD [36] off-road dataset.

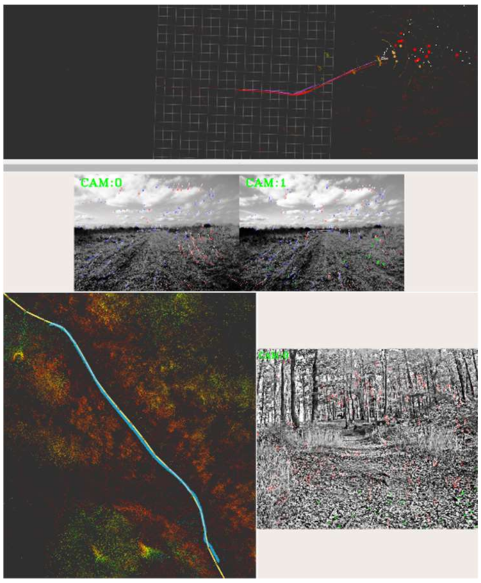

Fig. 14 Visual odometry using OpenVINS [35] - top: on Rellis-3D [2] dataset (red - ground-truth trajectory, purple - OpenVINS odometry); bottom: on our dataset (yellow - ground-truth trajectory, cyan - OpenVINS odometry)
Source: author.

3.4.2 Loop Closure Using Visual System

RTAB-Map offers a robust algorithm for loop closing [37], verified in urban environment by us (Figure 13) and in an off-road environment by the authors of the system (Figure 15).

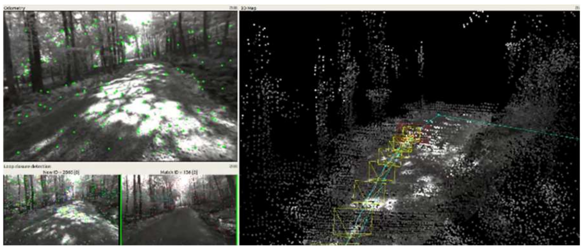

Fig. 15 Detection of a previously visited site and loop closure using RTAB-Map
Source: [38].

3.5 Traversability Segmentation Using Visual Systém

Semantic image segmentation is a common problem in computer vision science, and since it is a task requiring a certain degree of intuition, the best results are achieved by neural network models. In the analysis, we compared multiple models trained on different datasets with the purpose of segmentation in the off-road environment.

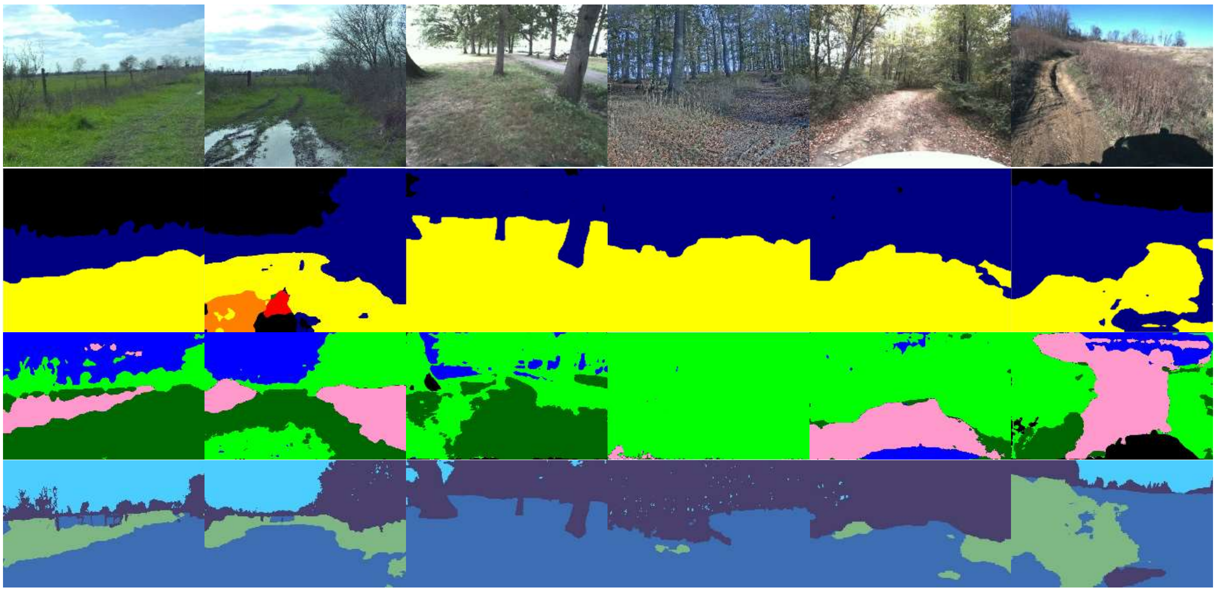

Fig. 16 Traversability segmentation of an image. First row: segmented images (from RELLIS-3D [2], Yamaha-seg [3], and our dataset). Second row: segmentation using GANav [39] model trained on RUGD [1] dataset. Third row: Unet [40] model trained within Traversability estimation [12, 13] project on RELLIS-3D [2] dataset. Fourth row: BiSeNet [41] model trained within OFFSEG [42] project on RUGD [1] dataset
Source: [2] (input data), [3] (input data), author (input data, processing).

From Figure 16, it is evident that off-road terrain segmentation is a difficult task even for state-of-the-art neural network models. The most successful are GANav [39] and OFFSEG [42] projects, whereby GANav shows slightly higher robustness (see the second image from the right), on the other hand, OFFSEG has higher sensitivity for detail. Both models were trained on RUGD [1] dataset, which proves to be more convenient for model training than RELLIS-3D [2] dataset because it is richer regarding environmental variety. Also, Yamaha-seg [3] dataset appears as a promising ground for model training due to the variety of field path images contained (see the failure of all networks on the rightmost image, which could be eliminated by further training on this dataset).

4 OUTLINE OF A COMPLEX PERCEPTION SYSTEM

Based on the performed analysis, we suggest splitting the development of autonomous navigation into two phases.

In the first phase, UGV will use complex SLAM from combined sensorical data (Section 3.1, Section 3.4), which is currently supported by operational and reliable systems. Since traversability segmentation of terrain, although promising and successful to a high degree, still brings risks and needs more development, in this phase, it will be used only for marginal tasks of refining a detailed trajectory during UGV motion. This refinement will be additionally enhanced by terrain roughness analysis systems, which are more straightforward (not included in the analysis since our company already has working solutions for that). During path planning, which requires more responsibility, the teleoperator will be asked for outlining an approximate path that will be preferentially followed by UGV. This lower level of autonomy can introduce multiple advantages:

- Attention is demanded only when asking for a next goal. This way, the teleoperator can concentrate on other tasks or operate multiple UGVs at once.
- During navigation, UGV is not dependent on reliable uninterrupted high bandwidth communication, since communication is required only occasionally, ideally in safe locations, where UGV can afford waiting for next instructions.
- In STEALTH mode transmission at the UGV side is reduced to a short broadcast of data about the situation around the new position, from the operator side it is required to transmit only information about the new goal and approximate path outline.
- Navigation in a familiar area (e.g., mapped terrain with a path marked in advance, return motion along the same path, repeated operation with identical paths as in previous operations) can be fully autonomous, since the UGV can follow known paths.

The second phase will introduce an autonomous mode, allowing the UGV to reach specified goals independently of the operator, even in an unfamiliar area, whereby semantical terrain representation will be used in a complete manner, tuned, and trained for general operation in the off-road environment.

Outline of such perception system, whose development is divided into two phases, is illustrated in Figure 17.

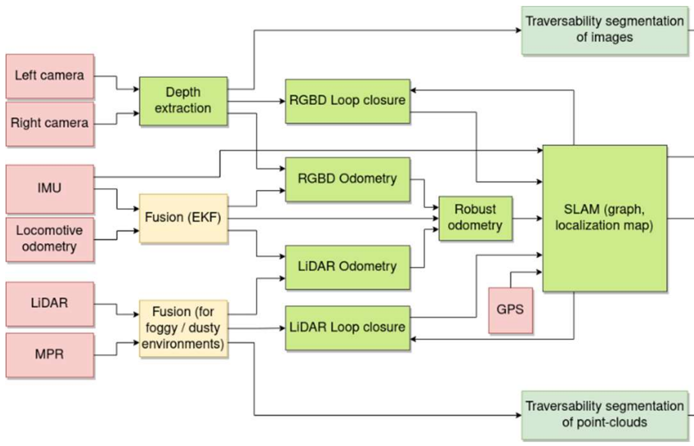

Fig. 17 Outline of the perception system and its application onto UGV navigation; red - inputs, green - perception system components analyzed in this work (pale green - second phase), yellow - other components of navigation system
Source: author.

.

5 CONCLUSION

This paper has presented a comprehensive overview of various perception tasks essential for resolving navigation challenges, alongside state-of-the-art methods tailored for off-road navigation. The selected solutions were tested against diverse environmental samples, extending beyond their original presentation by their authors, and included solutions primarily designed for different environments (urban, indoor) to assess their applicability in off-road settings.

The successful operation of a subset of the chosen methods for each predefined task across most tested environmental samples underscores the maturity of the current state of artificial intelligence in scientific research for deployment in autonomous or semi-autonomous off-road navigation. The paper has highlighted the advantages of such navigation over commonly used pure teleoperational methods for guiding robots in unexplored or partially unknown environments. Additionally, the paper outlined the concept of a navigation system with integrated perception components for the specified tasks.

However, off-road perception remains a relatively marginal focus in the scientific field. Despite the development of pioneering algorithms explicitly designed for this purpose and the collection of initial off-road datasets, this area lacks a high-quality platform providing multi-sensorial data for development, along with a standardized methodology for the assessment and comparison of off-road perception methods. The absence of such a platform, comparable to the KITTI [31] dataset for urban environments, hinders the development and selection of solutions that are not only operationally feasible but also most suitable for specific off-road navigation challenges. Therefore, future research, particularly before the second phase of development, should prioritize the creation of a methodological core and data collection for such a platform.

Acknowledgement

The analysis presented in this paper and the preparation of the article were conducted as part of Semi-Autonomous Reconnaissance Vehicle Project SEMOD-EL76/42-14/2023, supported by the Ministry of Defense of the Slovak Republic. We express our gratitude to the Ministry of Defense for their financial support and for making this research endeavor possible.

References

[1] WIGNESS, Maggie, EUM, Sungmin, ROGERS, John G, HAN, David and KWON, Heesung. A RUGD Dataset for Autonomous Navigation and Visual Perception in Unstructured Outdoor Environments. In *International Conference on Intelligent Robots and Systems (IROS)*. 2019. Available at: https://doi.org/10.1109/IROS40897.2019.8968283.

[2] JIANG, Peng, OSTEEN, Philip R., WIGNESS, Maggie B. and SARIPALLI, Srikanth. RELLIS-3D Dataset: Data, Benchmarks and Analysis. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. Online. 2020. P. 1110–1116. Available at: https://doi.org/10.1109/ICRA48506.2021.9561251.

[3] MATURANA, Daniel, CHOU, Po-wei, UENOYAMA, Masashi and SCHERER, Sebastian A. Real-Time Semantic Mapping for Autonomous Off-Road Navigation. In *International Symposium on Field and Service Robotics*. Online. 2017. Available at:
https://doi.org/10.1007/978-3-319-67361-5_22.

[4] FRITSCHE, Paul, KUEPPERS, Simon, BRIESE, Gunnar and WAGNER, Bernardo. Radar and LiDAR Sensorfusion

in Low Visibility Environments. In *International Conference on Informatics in Control, Automation and Robotics*. Online. 2016. Available at: https://api.semanticscholar.org/CorpusID:42431812.

[5] BIBER, P. and STRASSER, W. The normal distributions transform: a new approach to laser scan matching. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. 2003. p. 2743–2748. Available at: https://doi.org/10.1109/IROS.2003.1249285.

[6] KIM, Giseop, CHOI, Sunwook and KIM, Ayoung. Scan Context++: Structural Place Recognition Robust to Rotation and Lateral Variations in Urban Environments. *IEEE Transactions on Robotics*. 2021. Available at: https://doi.org/10.1109/TRO.2021.3116424.

[7] KIM, Giseop and KIM, Ayoung. Scan Context: Egocentric Spatial Descriptor for Place Recognition within 3D Point Cloud Map. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. October 2018. Available at: https://doi.org/10.1109/IROS.2018.8593953.

[8] KOIDE, Kenji, YOKOZUKA, Masashi, OISHI, Shuji and BANNO, Atsuhiko. *Voxelized GICP for Fast and Accurate 3D Point Cloud Registration*. 2020. EasyChair Preprint no. 2703. Available at:

https://doi.org/10.1109/ICRA48506.2021.9560835.

[9] KÜMMERLE, Rainer, GRISETTI, Giorgio, STRASDAT, Hauke, KONOLIGE, Kurt and BURGARD, Wolfram. G2o: A general framework for graph optimization. In *2011 IEEE International Conference on Robotics and Automation*. 2011. p. 3607–3613. Available at: https://doi.org/10.1109/ICRA.2011.5979949.

[10] MA, Junyi, XIONG, Guangming, XU, Jingyi and CHEN, Xieyuanli. CVTNet: A Cross-View Transformer Network for LiDAR-Based Place Recognition in Autonomous Driving Environments. *IEEE Transactions on Industrial Informatics*. 2023. Available at:
https://doi.org/10.1109/TII.2023.3313635.

[11] CHEN, Liang-Chieh, PAPANDREOU, George, SCHROFF, Florian and ADAM, Hartwig. Rethinking Atrous Convolution for Semantic Image Segmentation. *CoRR*. Online. 2017. Available at: http://arxiv.org/abs/1706.05587.

[12] AGISHEV, Ruslan, PETŘÍČEK, Tomáš and ZIMMERMANN, Karel. Trajectory Optimization Using Learned Robot-Terrain Interaction Model in Exploration of Large Subterranean Environments. In *IEEE Robotics and Automation Letters*. 2022. Vol. 7, no. 2p. 3365–3371. Available at: https://doi.org/10.1109/LRA. 2022.3147332.

[13] AGISHEV, Ruslan, ZIMMERMANN, Karel, PECKA, Martin and SVOBODA, Tomáš. MonoForce: Self-supervised learning of physics-aware grey-box model for predicting the robot-terrain interaction. *arXiv preprint arXiv*:2309.09007. 2023.

[14] CORTINHAL, Tiago, TZELEPIS, George and AKSOY, Eren Erdal. *SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving*. 2020. Available at: https://doi.org/10.1007/ 978-3-030-64559-5_16.

[15] Automatic Camera Re-Calibration for Robust Stereo Vision. Online. 2023. [Accessed 24 November 2023]. Available at: https://www.youtube.com/watch?v=2QGnOwfQKYo.

[16] GRAMMALIDIS, Nikos and STRINTZIS, Michael. G. Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences. In *IEEE Transactions on Circuits and Systems for Video Technology*. 1998. Vol. 8, no. 3p. 328–344.

[17] HIRSCHMULLER, Heiko. Stereo Processing by Semiglobal Matching and Mutual Information. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008. Vol. 30, no. 2p. 328–341. Available at: https://doi.org/10.1109/TPAMI.2007.1166.

[18] YOON, Kuk-Jin and KWEON, In So. Adaptive support-weight approach for correspondence search. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006. Vol. 28, no. 4p. 650–656. Available at:
https://doi.org/10.1109/TPAMI.2006.70.

[19] MEI, Xing, SUN, Xun, ZHOU, Mingcai, JIAO, Shaohui, WANG, Haitao and ZHANG, Xiaopeng. On building an accurate stereo matching system on graphics hardware. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 2011. p. 467–474. Available at: https://doi.org/10.1109/ICCVW.2011.6130280.

[20] HAMZAH, Rostam Affendi, RAHIM, Rosman Abd and NOH, Zarina Mohd. Sum of Absolute Differences algorithm in stereo correspondence problem for stereo matching in computer vision application. In *2010 3rd International Conference on Computer Science and Information Technology*. 2010. p. 652–657. Available at: https://doi.org/10.1109/ICCSIT.2010.5565062.

[21] CHANG, Qiong, ZHA, Aolong, WANG, Weimin, LIU, Xin, ONISHI, Masaki, LEI, Lei, ER, Meng Joo and MARUYAMA, Tsutomu. *Efficient stereo matching on embedded GPUs*

with zero-means cross correlation. 2022. Available at: https://doi.org/10.1016/j.sysarc.2021.102366

[22] HERNANDEZ-JUAREZ, D., CHACÓN, A., ESPINOSA, A., VÁZQUEZ, D., MOURE, J.C. and LÓPEZ, A.M. Embedded Real-time Stereo Estimation via Semi-global Matching on the GPU. In *Procedia Computer Science*. Online. 2016. Vol. 80, p. 143–153. Available at: https://doi.org/10.1016/j.procs.2016.05.305.

[23] XU, Gangwei, CHENG, Junda, GUO, Peng and YANG, Xin. *Attention Concatenation Volume for Accurate and Efficient Stereo Matching*. 2022. Available at: https://doi.org/10.1109/CVPR52688.2022.01264 .

[24] MAYER, N., ILG, E., HÄUSSER, P., FISCHER, P., CREMERS, D., DOSOVITSKIY, A. and BROX, T. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. Online. 2016. Available at: https://doi.org/10.1109/CVPR.2016.438

[25] TANKOVICH, Vladimir, HÄNE, Christian, ZHANG, Yinda, KOWDLE, Adarsh, FANELLO, Sean and BOUAZIZ, Sofien. *HITNet: Hierarchical Iterative Tile Refinement Network for Real-time Stereo Matching*. 2023.

[26] SCHARSTEIN, D., SZELISKI, R. and ZABIH, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. 2001. p. 131–140. Available at:

https://doi.org/10.1109/SMBV.2001.98877.

[27] SCHARSTEIN, D. and SZELISKI, R. High-accuracy stereo depth maps using structured light. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*. 2003. p. I-I. Available at: https://doi.org/10.1109/CVPR.2003.1211354.

[28] SCHARSTEIN, Daniel and PAL, Chris. Learning Conditional Random Fields for Stereo. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007. p. 1–8. Available at: https://doi.org/CVPR. 2007.383191.

[29] HIRSCHMULLER, Heiko and SCHARSTEIN, Daniel. Evaluation of Cost Functions for Stereo Matching. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007. p. 1–8. Available at: https://doi.org/10.1109/CVPR.2007.383248 .

[30] SCHARSTEIN, Daniel, HIRSCHMÜLLER, Heiko, KITAJIMA, York, KRATHWOHL, Greg, NEŠIĆ, Nera, WANG, Xi and WESTLING, Porter. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. In JIANG, Xiaoyi, HORNEGGER, Joachim and KOCH, Reinhard (eds..), *Pattern Recognition*. Cham: Springer International Publishing, 2014. p. 31–42. ISBN 978-3-319-11752-2. Available at: https://doi.org/10.1007/978-3-319-11752-2_3.

[31] GEIGER, Andreas, LENZ, Philip and URTASUN, Raquel. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012. Available at: https://doi.org/10.1109/CVPR.2012.6248074.

[32] MENZE, Moritz and GEIGER, Andreas. Object Scene Flow for Autonomous Vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. Available at: https://doi.org/10.1109/CVPR.2015.7298925.

[33] MARKO, Thomas and KUBINGER, Wilfried. Automatic Stereo Camera Calibration in Real-World Environments Without Defined Calibration Objects. In *Proceedings of the 29th DAAAM International Symposium*. 2018. p. 102–108. ISBN 9783902734204.

[34] LABBÉ, Mathieu and MICHAUD, François. Long-term online multi-session graph-based SPLAM with memory management. *Autonomous Robots*. Online. November 2017. Vol. 42, no. 6p. 1133. Available at: https://doi.org/10.1007/s10514-017-9682-5.

[35] GENEVA, Patrick, ECKENHOFF, Kevin, LEE, Woosik, YANG, Yulin and HUANG, Guoquan. OpenVINS: A Research Platform for Visual-Inertial Estimation. In *Proc. of the IEEE International Conference on Robotics and Automation*. 2020.

[36] CHUSTZ, George and SARIPALLI, Srikanth. *ROOAD: RELLIS Off-road Odometry Analysis Dataset*. 2021. Available at: https://doi.org/10.1109/IV51971.2022.9827133.

[37] LABBÉ, Mathieu and MICHAUD, François. Online global loop closure detection for large-scale multi-session graph-based SLAM. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014. p. 2661–2666. Available at: https://doi.org/10.1109/IROS.2014.6942926.

[38] RTAB-Map: Four Kilometers Walk in Forest (an uncut real-time visual SLAM demo). Online. 2023. [Accessed 24 November 2023]. Available at: https://www.youtube.com/watch?v=G-5jesjNfLc.

[39] GUAN, Tianrui, KOTHANDARAMAN, Divya, CHANDRA, Rohan, SATHYAMOORTHY, Adarsh Jagan, WEERAKOON, Kasun and MANOCHA, Dinesh. GA-Nav: Efficient Terrain Segmentation for Robot Navigation

in Unstructured Outdoor Environments. In *IEEE Robotics and Automation Letters*. 2022. Vol. 7, no. 3p. 8138–8145. Available at: https://doi.org/10.1109/LRA.2022.3187278.

[40] RONNEBERGER, Olaf, FISCHER, Philipp and BROX, Thomas. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. Available at: https://doi.org/10.1007/ 978-3-319-24574-4_28.

[41] YU, Changqian, WANG, Jingbo, PENG, Chao, GAO, Changxin, YU, Gang and SANG, Nong. *BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation*. 2018. Available at: https://doi.org/10.1007/978-3-030-01261-8_20.

[42] VISWANATH, Kasi, SINGH, Kartikeya, JIANG, Peng, SUJIT, P. B. and SARIPALLI, Srikanth. OFFSEG: A Semantic Segmentation Framework For Off-Road Driving. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. 2021. p. 354–359. Available at: http://10.1109/CASE49439.2021.9551643.

Eng. Matej **VARGOVČÍK**
RoboTech Vision s.r.o.
Červený kameň 61
900 89 Častá
Slovak Republic
E-mail: vargovcik@robotechvision.com

Eng. Peter **PÁSZTÓ**, PhD.
RoboTech Vision s.r.o.
Červený kameň 61
900 89 Častá
Slovak Republic
E-mail: paszto@robotechvision.com

Eng. Marian **KĽÚČIK**
RoboTech Vision s.r.o.
Červený kameň 61
900 89 Častá
Slovak Republic
E-mail: klucik@robotechvision.com

Eng. Martin **SMOĽÁK**
RoboTech Vision s.r.o.
Červený kameň 61
900 89 Častá
Slovak Republic
E-mail: smolak@robotechvision.com

Eng. Patrik **ŠTEFKA**
RoboTech Vision s.r.o.
Červený kameň 61
900 89 Častá
Slovak Republic
E-mail: stefka@robotechvision.com

Eng. Jakub **LENNER**
RoboTech Vision s.r.o.
Červený kameň 61
900 89 Častá
Slovak Republic
E-mail: lenner@robotechvision.com

Matej VARGOVČÍK studied robotics on Faculty of Electrical Engineering and Information Technology of STU in Bratislava. During his studies he was developing number of significant scientific works. He implemented his knowledge in RoboTech Vision Ltd. company in which he is currently working. He is responsible for development of autonomous navigation an localization algorithms. He also activelly collaborates with the scientific comunity in his field and is contributing his solutions on GitHub platform.

Peter PÁSZTÓ studied robotics on Faculty of Electrical Engineering and Information Technology of STU in Bratislava. He is one of the founders and CEOs of RoboTech Vision Ltd. Throughout his university studies he has focused on image processing algorithms and their application in the field of mobile robotics (navigation and localization). During PhD studies he was awarded together with his team (now already CEOs of company) at a scientific conference in Rijeka, Croatia for an image processing algorithm detecting obstacles in front of a mobile robot navigated only by a smartphone running on Android OS.

Marian KĽÚČIK studied robotics on Faculty of Electrical Engineering and Information Technology of STU in Bratislava. He is one of the founders and CEOs of RoboTech Vision Ltd. During his studies he focused on the development of a robot with a combined chassis. He was also working on development of genetic algorithms for navigation and localization of his robotic platform. His skills include programming in multiple programming languages, controlling various operating systems, control system development, mechanics, control software and electronics. He focuses on the development of communication layer software and the configuration of OS in his company.

Martin SMOĽÁK has been involved in mobile robotics since secondary school. He was also involved in telemedicine projects and developed smart cars for the police in the past. He studied robotics on Faculty of Electrical Engineering and Information Technology of STU in Bratislava. Within RoboTech Vision Ltd. (of which he is also on of founders and CEOs), he focuses mainly on the development of lower-level control software, hardware, mechanics, and navigation algorithms and mobile robotic platforms design.

Patrik ŠTEFKA studied robotics on Slovak University of Technology in Bratislava. Presently he is working in RoboTech Vision Ltd. company where he is responsible for development and integration of AI methods and neural network data processing into human interaction, navigation and localization algorithms.

Jakub LENNER studied robotics on Slovak University of Technology in Bratislava. Presently he is working in RoboTech Vision Ltd. company where he is cooperating on development of visual navigation and path-planning methods. He is also author of several scientific publications in the field of precise robot path-planning for docking algorithms using visual systems and image processing.