CloudVision: DNN-based Visual Localization of Autonomous Robots using Prebuilt LiDAR Point Cloud

Evgeny Yudin, Pavel Karpyshev, Mikhail Kurenkov, Alena Savinykh, Andrei Potapov, Evgeny Kruzhkov, and Dzmitry Tsetserukou
ISR Laboratory, Skolkovo Institute of Science and Technology, Moscow, Russia
{Evgeny.Yudin, Pavel.Karpyshev, Mikhail.Kurenkov, Alena.Savinykh, Andrei.Potapov, Evgeny.Kruzhkov, D.Tsetserukou}@skoltech.ru

Abstract—In this study, we propose a novel visual localization approach to accurately estimate six degrees of freedom (6-DoF) poses of the robot within the 3D LiDAR map based on visual data from an RGB camera. The 3D map is obtained utilizing an advanced LiDAR-based simultaneous localization and mapping (SLAM) algorithm capable of collecting a precise sparse map. The features extracted from the camera images are compared with the points of the 3D map, and then the geometric optimization problem is being solved to achieve precise visual localization. Our approach allows employing a scout robot equipped with an expensive LiDAR only once — for mapping of the environment, and multiple operational robots with only RGB cameras onboard — for performing mission tasks, with the localization accuracy higher than common camera-based solutions. The proposed method was tested on the custom dataset collected in the Skolkovo Institute of Science and Technology (Skoltech). During the process of assessing the localization accuracy, we managed to achieve centimeter-level accuracy; the median translation error was as low as 1.3 cm. The precise positioning achieved with only cameras makes possible the usage of autonomous mobile robots to solve the most complex tasks that require high localization accuracy.

Index Terms—Autonomous robot, Visual localization, Mapping, Deep Learning, Sensors Fusion, LiDAR map

I. INTRODUCTION

A. Motivation

Mobile robotics is a rapidly expanding research field, with both the scientific community and industry pursuing modern and dependable solutions. The mobile robots market is valued at 1.61 billion dollars in 2021 and expected to grow to 22.15 billion dollars by 2030 [1], highlighting the industry’s interest in this area. The academic community’s interest is evident through a substantial increase in the number of papers, journals, and conferences covering all aspects and applications of autonomous machines. The focus is on providing reliable algorithms to ensure accurate, efficient, and safe operation of autonomous robots.

B. Problem Statement

Localization is a critical task in mobile robotics, requiring a robot to determine its position at any given moment for path planning and obstacle avoidance. This problem is typically addressed using algorithms that process data from cameras, LiDARs, or a combination of both sensors.

Cameras are the most widespread perception sensors due to their cheapness and ability to obtain dense information on the environment, including colors, depth, and shapes. However, visual localization algorithms show less accurate results than LiDAR-based solutions.

The main advantage of LiDARs, in turn, is obtaining highly reliable data on the distance to objects without preprocessing, regardless of changes in lighting and changing seasons. However, LiDAR data is vulnerable to change of weather conditions (snow, rain, mist).

For these reasons, currently there is no single approach to designing a sensor setup for the localization problem. On the one hand, LiDARs provide the highest accuracy, but are costly. On the other hand, cameras are affordable, but do not provide high accuracy.

C. Related Works

LiDAR-based simultaneous localization and mapping (SLAM) algorithms are able to achieve high localization accuracy and generate a high-quality sparse 3D map of the environment. According to the results on KITTI benchmark [2], the most accurate SLAM algorithms are based on Lidar Odometry and Mapping (LOAM) method [3], utilizing Li-
DAR as the main source of information. The core idea of this algorithm is the extraction of edges for geometrical surfaces depicted on the LiDAR scan.

Multiple papers are devoted to extending the LOAM approach to achieve better performance. For example, LeGO-LOAM [4] improves the original approach using ground partitioning, point cloud segmentation, and advanced computational optimization. Segmentation allows detecting noisy points that may represent unreliable features and filter them out.

Other approaches to LiDAR-based SLAM [5], [6] are based on Convolutional Neural Networks (CNNs) for point cloud processing. Although deep learning approaches show good results on open datasets in terms of odometry accuracy, they are unreliable in cases of drastic scene change, such as moving from indoor to outdoor scene. Thus, for robust operation, these approaches should be trained again on unseen scenes, which significantly reduces their applicability in real-world scenarios.

Sensor Fusion Methods, particularly camera and LiDAR data fusion, has found wide application for solving both localization and SLAM problems. Researchers [7], [8] propose highly accurate localization approaches based on the LiDAR-camera data fusion. However, all of them require the presence of both camera and LiDAR in the sensor setup of each robot.

Feng et al. [9] implemented a deep learning approach to compute the descriptors that allow direct matching of keypoints across an image and a 3D map. The proposed neural network is trained to match features on the specific 3D scene. Thus, this method is highly dependent on the training data and unable to achieve the reliable results on scenes not included in the training dataset.

Visual Localization methods can be divided into two subgroups based on their operation principle: image-based and structure-based. Image-based localization approaches rely solely on images, and do not require the storage of a 3D scene for localization in an explicit form. In particular, PoseNet [10] leverages a CNN, such as VGGNet or ResNet, to regress both camera position and orientation.

Image retrieval systems aim to find similar images to a query image (i.e., the current image obtained from the camera) among a dataset. State-of-the-art approaches aimed at solving the image retrieval problem mainly use trained global descriptors. For example, NetVLAD [11] is able to determine the most similar image to a query image with a high degree of invariance to changes in conditions (scale, illumination, etc.).

Structure-based methods calculate the pose of the camera in a reconstructed 3D map of the area. Most of them involve the Structure from Motion (SfM) [12] which is able to estimate the 3D structure of a scene from a set of 2D images. For localization, keypoints are extracted from the query image and matched with the 3D model using its descriptors. When the matches between the image and the 3D model are determined, the Perspective-n-Point problem is solved to calculate the position of the camera.

The most outstanding modern approach to solving visual localization problem is PixLoc [13], a scene-invariant neural network that allows to extract dense features from images. PixLoc learns to distinguish pieces of images, that are most suitable for localization, through end-to-end learning from pixels to pose, and demonstrates exceptional generalization to new scenes by separating model parameters and scene geometry.

Despite the fact that this approach shows most promising
results compared to its alternatives, accuracy may still not be enough in cases that require precise localization. We propose to create a visual localization pipeline based on LiDAR maps, investigating the impact of replacing Structure from Motion to a map obtained by the state-of-the-art LiDAR SLAM algorithm. In our hypothesis, the precise LiDAR point clouds would improve the localization accuracy and overall robustness of the pipeline.

D. Contribution

We propose and evaluate a novel approach to solving the problem of visual localization by matching camera data with a prebuilt 3D LiDAR map. The concept for this approach is to use two types of robots. The first one collects a 3D map and database images by means of camera and LiDAR, and the robots of the second type are equipped with only a camera and leverages the prebuilt LiDAR map for visual localization. The layout of the proposed concept is shown in Fig. 1.

In the scope of this research, we analyze the existing LiDAR-based mapping approaches and choose the optimal one for collecting a 3D map. We develop an algorithm that allows to collect a map with information about the visibility of each point based on the camera’s field of view. Next, we adapt the 3D map for the Pixloc pipeline, and conduct a series of experiments to determine the accuracy of indoor localization using the proposed approach.

II. METHOD OF VISUAL LOCALIZATION

Our pipeline can be divided into three main steps. The first step is to create a 3D structure of the environment using an advanced LiDAR-based SLAM algorithm. During the second step, the 3D structure segment is selected using image retrieval, which allows reducing the search area. In the last step, we use the dense feature extractor and classical geometric optimization to get the 6-DoF robot pose $\langle R, t \rangle$, where $R$ is a rotation matrix and $t$ is a translation vector of the robot. The full pipeline is shown in Fig. 2.

A. Obtaining 3D Map

The LOAM algorithm family is a leading performer in terms of odometry accuracy on the KITTI benchmark. To construct a 3D map, we utilized A-LOAM [14], an advanced implementation of LOAM that provides precise localization without requiring high accuracy LiDAR ranging and inertial measurements. The algorithm achieves high 3D map accuracy by reducing its update rate for proper matching and point cloud registration, while also operating in real time, making it ideal for use on a scout robot. Our approach extends A-LOAM with an algorithm for the projection of co-visible map points.

To estimate the position of the camera, it is necessary to find the map segment corresponding to the current query image and project the map points onto the image. The problem is that it needs to project only those map segment points that are visible to the camera, excluding the rest, e.g., located around the corner. To solve this problem, we modify the stage of obtaining a map in the A-LOAM algorithm: we assign an index of the database image to each point of the LiDAR scan corresponding to that image. The result of point cloud indexing is shown in Fig. 3. This approach allows to effectively filter the points that are not visible from the camera, knowing the common database images. The projection of map points on the image is shown in Fig. 4(c).

B. Image Retrieval

In our approach, to solve the image retrieval problem we use a CNN-based architecture called NetVLAD [11], that is trained in an end-to-end manner specially for the place recognition task. This network extracts global descriptors from images, and the comparison of such descriptors provides an identification of the most similar images. Thus, having a database of images, it is possible to quickly evaluate the place where the query image was made using NetVLAD. The high speed of such evaluation is possible due to the small size of descriptors, and the ability to calculate them for the database images preliminarily. Thereby, we can extract the most similar image from the database to the current query image and, using the previously described algorithm, define the co-visible map points. The described approach is used for efficient and fast filtering of a 3D scene, which significantly limits the search area. In our work, we use NetVLAD weights pretrained on the Pittsburgh (Pitts30k) dataset [15].

C. Pose Estimation

To estimate the robot pose, we utilize the algorithm proposed in Pixloc [13]. Using NetVLAD global descriptors, we find the image from the database that is the most similar to the current query image. After that, we use the Pixloc convolutional neural network to extract dense features from the database image. The network is invariant to the scene, and is able to work equally well in both indoor and outdoor scenes without retraining. In our work, we use the Pixloc network pretrained with MegaDepth dataset [16]. Next, the points of the filtered 3D structure are projected onto the retrieved database image, and 2D descriptors obtained from Pixloc are assigned to each 3D point. Thus, we get a point cloud with image descriptors. The next step is minimizing the difference in appearance between the query image and database image. Pixloc’s dense features are also extracted from the query image, and certain points with a descriptor are projected onto it. Thus, the position, from which the query image was captured, is estimated by minimizing the error using the Levenberg-Marquardt algorithm by aligning Pixloc features from the query image and Pixloc features and corresponding 3D points from the database image.

III. SYSTEM OVERVIEW

A. Hardware and Software Architecture

To collect the dataset, the HermesBot autonomous platform [17], depicted in Fig. 5, was utilized. The robot is equipped with all sensors required for a comprehensive
perception system. The sensor setup includes a Velodyne VLP-16 LiDAR, and two Intel RealSense D435 RGB-D cameras, mounted on the front and back sides of the robot. The LiDAR-to-camera transformation was initially estimated based on the robot’s CAD geometry and further improved via calibration algorithm [18].

B. Dataset Collection

To evaluate the proposed approach, we collected an indoor dataset of LiDAR and visual data at Skoltech campus to assess our approach. The dataset consists of two sequences recorded on similar closed trajectories, each spanning 380 meters and starting and ending at the same point. Sequence 1 includes LiDAR scans and camera data, and was obtained by the first scout robot. Sequence 2, obtained by the operation robot, includes visual data for localization algorithm testing and LiDAR data for obtaining ground truth. Sequence 1 was used to construct the 3D map and assemble the image database, while sequence 2 was utilized to evaluate the proposed method’s localization accuracy.

Sequence 1 involved recording 4966 LiDAR scans and visual data at 30 frames per second, with a resolution of 640x480 in RGB format. We selected 300 frames uniformly distributed along the entire sequence length from approximately 15000 camera frames per sequence to include in the database for NetVLAD image retrieval and approximate pose estimation. For sequence 2, the selected images were used as query images to evaluate the proposed method’s accuracy, with the number of images chosen as a balance between system speed and preliminary localization accuracy. Using 300 images allowed us to estimate the robot position with a 2-meter accuracy and fully assess our approach’s performance. We employed an advanced LiDAR-based SLAM algorithm to obtain the robot trajectory and ground truth poses. Each camera pose contained three linear and three angular coordinates for future 6-DoF pose estimation.

IV. EXPERIMENTS AND RESULT DISCUSSION

In order to evaluate the proposed approach, two sets of experiments were carried out. The aim of the first experiment was to demonstrate that the scale of the 3D map, obtained using the LiDAR-based SLAM approach, corresponds to real-world dimensions through localization accuracy estimation. The second set of experiments was aimed at assessing the localization accuracy of the proposed algorithm and comparing it with state-of-the-art visual approaches.

A. Mapping Evaluation

As A-LOAM was utilized as ground truth for camera positions in both database and query images in this study, it was essential to verify its accuracy. Given the lack of a loop closure algorithm in this method, we leveraged the close proximity of the starting and ending points of the trajectory to measure the distance between them.

The relative translation error in the experiment constituted 0.13%, while the entire length of the sequence was 380 m. The absolute translation error was 0.49 m.

The proposed approach estimates the robot pose using the most similar database image, therefore the error of the SLAM algorithm will be taken into account only in the distance relative to the positions from which the images were taken. It means that this translation error allows evaluating the accuracy of localization with high precision, since it accumulates during the entire sequence. In order to prove the consistency and uniformity of error accumulation, we ran the LiDAR-based SLAM on short sequences, physically measuring the distance traveled by the robot. Given that the average distance between the query image and the closest one database image found from the image retrieval is 2 meters, we conducted 10 launches of the robot on a trajectory.
Table I

| Method                        | Median translation error, cm | Median rotation error, deg. | Percentage of images at (5 cm, 2 deg.) | Number of points in millions |
|-------------------------------|------------------------------|-----------------------------|--------------------------------------|-----------------------------|
| hloc + SfM                    | 4.5                          | 0.29                        | 55.3                                 | 0.03                        |
| Pixloc + SfM                  | 3.1                          | 0.21                        | 64.7                                 | 0.03                        |
| Pixloc + Raw LiDAR data (ours)| 1.8                          | 0.13                        | 97.3                                 | 90                          |
| Pixloc + LiDAR Map (ours)     | 1.3                          | 0.09                        | 99.3                                 | 0.72                        |

Fig. 6. Percentage of query images within certain limits for translation and rotation error.

3 meters long and determined the absolute error, which averaged at 0.5 cm. This error makes it possible to estimate the accuracy of localization with centimeter accuracy.

B. Localization Evaluation

To assess the localization accuracy, we estimate the median translation and rotation error [10] of the robot position in the sequence 2 for 300 query images. Fig. 6 demonstrates the selected accuracy error thresholds and the percentage of poses estimated within this accuracy range. For the proposed approach validation, we compare its localization accuracy with the accuracy of state-of-the-art visual methods based on Structure from Motion, Pixloc and hloc [19], and with ablation modification of our method, in which the 3D map is replaced by simple raw LiDAR scans. The results of projecting 3D points on the image for different methods are shown in Fig. 4.

For Pixloc and hloc, the SfM models are reconstructed from 600 database images included in sequence 1. To build the model, we initially extract features from database images using Superpoint [20] and match them using SuperGlue [21]. To assign the positions of the cameras in the model, we use the previously obtained ground truth poses obtained by A-LOAM. Then, we triangulate the extracted points with respect to these poses using their match information.

The proposed approach achieves centimeter-level accuracy, with median translation and rotation errors of 0.13 cm and 0.09 deg, respectively, according to the experimental results in Fig. 6 and Table I. It significantly outperforms state-of-the-art visual localization approaches in terms of accuracy, at least 2.5 times. The authors attribute this superiority to the use of much denser maps generated by LiDAR SLAM, in contrast to SfM, which typically yields sparse point clouds due to unstable keypoint extraction and matching, making it more scene-dependent than LiDAR technology. (Fig. 4(b)).

The superiority over the raw LiDAR approach is due to the fact that the scans are projected onto the database image in lines (Fig. 4(a), and that negatively affects the geometric optimization of the pose. Although the raw data contains the accurate description of the environment, the proposed approach provides greater accuracy, and a significantly smaller number of points in the structure, due to a more uniform distribution of points in the 3D map.

V. CONCLUSIONS

We proposed a new approach for localizing autonomous robots using prebuilt LiDAR point cloud maps. This was achieved by modifying the existing visual localization algorithm Pixloc to use explicit point clouds instead of Structure from Motion (SfM).

To evaluate the proposed approach, we collected a dataset consisting of two trajectories in the same location, including both LiDAR and camera data. The proposed localization pipeline was applied to the second data sequence and compared with Pixloc and hloc state-of-the-art visual localization approaches. The experimental results have shown that the proposed pipeline significantly outperforms the existing camera-based methods in terms of accuracy. The median translation error was equal to 1.3 cm, which is approximately 2.5 times better than the best scoring Pixloc, that achieved a median translation error of 3.1 cm. The resulting pipeline is both accurate and scene-invariant due to the use of a neural network-based dense feature extractor.

The use of the proposed approach is able to significantly decrease the cost of a robot fleet compared with the one equipped with only LiDARs, and increase the accuracy and robustness compared to a camera-only fleet.

VI. DISCUSSION AND FUTURE WORK

In the future, we plan to perform a more extensive evaluation of the proposed approach. This includes testing on additional datasets [22], [23], in real-world conditions on various setups, for example, outdoor ground robots [24], indoor ground robots for hospitals [25], [26] and shopping rooms [27], and UAVs [28]–[31], as well as examine the
applicability of the approach to more sophisticated systems, e.g., modular two-wheeled rovers [32] and plant inspection robots [33].

ACKNOWLEDGEMENTS

The reported study was funded by CNRS and RFBR according to the research project No. 21-58-15006.

REFERENCES

[1] “Autonomous Mobile Robot Market by Type, by Application, by End-User - Global Opportunity Analysis and Industry Forecast 2022-2030,” 2022, [accessed 28-April-2022]. [Online]. Available: https://www.researchandmarkets.com/reports/5529480/autonomous-mobile-robot-market-by-type/by-

[2] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012, pp. 3354–3361.

[3] J. Zhang and S. Singh, “Loam: Lidar odometry and mapping in real-time,” in Robotic Science and Systems, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.

[4] T. Shan and B. Englot, “Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain,” in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018, pp. 4758–4765.

[5] H. Shi, G. Lin, H. Wang, T-Y. Hung, and Z. Wang, “Spsequencenet: Semantic segmentation network on 4d point clouds,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 4574–4583.

[6] D. Yin, Q. Zhang, J. Liu, X. Liang, Y. Wang, J. Maanpää, H. Ma, J. Hyppä, and R. Chen, “Cae-lo: Lidar odometry leveraging fully unsupervised convolutional auto-encoder for interest point detection and feature description,” arXiv preprint arXiv:2001.01354, 2020.

[7] S. Chen, B. Zhou, C. Jiang, W. Xue, and Q. Li, “A lidar/visual slam backend with loop closure detection and graph optimization,” Remote Sensing, vol. 13, no. 14, p. 2720, 2021.

[8] Y. Seo and C.-C. Chou, “A tight coupling of vision-lidar measures for an effective odometry,” in 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019, pp. 1118–1123.

[9] M. Feng, S. Hu, M. H. Ang, and G. H. Lee, “2d3d-matchnet: Learning to match keypoints across 2d image and 3d point cloud,” in 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 4790–4796.

[10] A. Kendall, M. Grimes, and R. Cipolla, “Posenet: A convolutional network for real-time 6-dof camera relocalization,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2938–2946.

[11] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “Netvlad: Cnn architecture for weakly supervised place recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 5297–5307.

[12] J. L. Schonberger and J.-M. Frahm, “Structure-from-motion revisited,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 4104–4113.

[13] P.-E. Sarlin, A. Unagar, M. Larsson, H. Germain, C. Toft, V. Larsson, M. Pollefeys, V. Lepetit, L. Hammarstrand, F. Kahn et al., “Back to the feature: Learning robust camera localization from pixels to pose,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3247–3257.

[14] HKUST-Aerial-Robotics, “HKUST-aerial-robotics/a-loam: Advanced implementation of loam,” [accessed 28-April-2022]. [Online]. Available: https://github.com/HKUST-Aerial-Robotics/a-LOAM

[15] A. Torii, J. Sivic, T. Pajdla, and M. Okutomi, “Visual place recognition with repetitive structures,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 883–890.

[16] Z. Li and N. Snavely, “Megadepth: Learning single-view depth prediction from internet photos,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2041–2050.

[17] S. Protasov, P. Karpyshhev, I. Kalinov, P. Kopanev, N. Mikhailovskiy, A. Sedunin, and D. Tsetsurukou, “Cnn-based omnidirectional object detection for hermesbot autonomous delivery robot with preliminary frame classification,” in 2021 26th IEEE International Conference on Advanced Robotics (ICAR). IEEE, 2021, pp. 517–522.

[18] D. Tsai, S. Worrall, M. Shan, A. Lohr, and E. Nebot, “Optimising the selection of samples for robust lidar camera calibration,” in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). IEEE, 2021, pp. 2631–2638.

[19] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, “From coarse to fine: Robust hierarchical localization at large scale,” in CVPR, 2019.

[20] D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superpoint: Self-supervised interest point detection and description,” in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 224–236.

[21] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, “SuperGlue: Learning feature matching with graph neural networks,” in CVPR, 2020.

[22] A. L. Pavlov, P. A. Karpyshhev, G. V. Ovchinnikov, I. V. Oselodets, and D. Tsetsurukou, “Kevisensor: lossless video dataset collected on russian winter roads with traffic sign annotations,” in 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 9597–9602.

[23] V. Ilin, I. Kalinov, P. Karpyshhev, and D. Tsetsurukou, “Deepscanner: a robotic system for automated 2d object dataset collection with annotations,” in 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, 2021, pp. 1–08.

[24] P. Karpyshhev, E. Kruzhkov, E. Yudin, A. Savinykh, A. Potapov, M. Kurenkov, A. Kolomeytsev, I. Kalinov, and D. Tsetsurukou, “Mucaslam: Cnn-based frame quality assessment for mobile robot with omnidirectional visual slam,” in 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE). IEEE, 2022, pp. 368–373.

[25] N. Mikhailovskiy, A. Sedunin, S. Perminov, I. Kalinov, and D. Tsetsurukou, “Ultrabot: Autonomous mobile robot for indoor uv-c disinfection with non-trivial shape of disinfection zone,” in 2021 18th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, 2021, pp. 1–7.

[26] S. Perminov, N. Mikhailovskiy, A. Sedunin, I. Okunevich, I. Kalinov, M. Kurenkov, and D. Tsetsurukou, “Ultrabot: Autonomous mobile robot for indoor uv-c disinfection,” in 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE). IEEE, 2021, pp. 2147–2152.

[27] A. Petrovsky, I. Kalinov, P. Karpyshhev, M. Kurenkov, V. Ramzhave, V. Ilin, and D. Tsetsurukou, “Customer behavior analytics using an autonomous robotics-based system,” in 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, 2020, pp. 327–332.

[28] I. Kalinov, A. Petrovsky, R. Agishev, P. Karpyshhev, and D. Tsetsurukou, “Importance-based control for soft uvw landing on a ground robot in heterogeneous robotic system,” in 2021 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2021, pp. 1653–1658.

[29] I. Kalinov, E. Safronov, R. Agishev, M. Kurenkov, and D. Tsetsurukou, “High-precision uvw localization system for landing on a mobile collaborative robot based on an ir marker pattern recognition,” in 2019 IEEE 9th Velocar Technology Conference (VTC2019-Spring). IEEE, 2019, pp. 1–6.

[30] I. Kalinov, A. Petrovsky, V. Ilin, E. Pristansky, M. Kurenkov, V. Ramzhave, I. Idrisov, and D. Tsetsurukou, “Warevision: Cnn barcode detection-based uav trajectory optimization for autonomous warehouse stocktaking,” IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 6647–6653, 2020.

[31] I. Kalinov, D. Trinitatova, and D. Tsetsurukou, “Warevr: Virtual reality interface for supervision of autonomous robotic system aimed at warehouse stocktaking,” in 2021 ieee international conference on systems, man, and cybernetics (smc). IEEE, 2021, pp. 2139–2145.

[32] A. Petrovsky, I. Kalinov, P. Karpyshhev, D. Tsetsurukou, A. Ivanov, and A. Golkar, “The two-wheeled robotic swamp concept for mars exploration,” Acta Astronautica, vol. 194, pp. 1–8, 2022.

[33] P. Karpyshhev, V. Ilin, I. Kalinov, A. Petrovsky, and D. Tsetsurukou, “Autonomous mobile robot for apple plant disease detection based on cnn and multi-spectral vision system,” in 2021 IEEE/SICE international symposium on system integration (SII). IEEE, 2021, pp. 157–162.