Research and Analysis of Robot SLAM Technology in Complex Scenarios

Jie Gao

{ gaojienora@dlmu.edu.cn }

Faculty of Information Science and Technology, Dalian Maritime University, Dalian 116026, China

Abstract. With the advancement of robotics technology, more and more robots are being deployed in various application scenarios. They are no longer confined to basic indoor environments but have extended to more complex outdoor environments. Therefore, the demand for highly reliable SLAM technology is increasing. This article examines three challenging scenarios: texture loss in low-light images, high dynamic interference, and signal degradation in steep terrain. First, the paper analyzes the reasons why the SLAM system is prone to failure under these conditions. For each scenario, two solutions are proposed. Through method comparison and analysis, we put forward improvement suggestions to break through the technical bottlenecks, such as enhancing the feature extraction ability through computation, integrating point and line features to achieve trajectory tracking, processing moving objects using semantic information, and fusing data from multiple sensors. By conducting a comparative analysis of three scenarios, the study identified three common challenges: insufficient environmental adaptability of the SLAM system, inadequate handling capacity for moving objects, and efficiency bottlenecks. Finally, three research directions were proposed: Firstly, SLAM should shift from constructing static maps to understanding dynamic scene changes; Secondly, more efficient methods need to be developed to adapt to various types of devices and environments; Finally, the system should possess the ability to continuously learn from new environments.

Keywords: SLAM, Complex Scenarios, Dynamic Scene Understanding.

1 Introduction

Nowadays, robotics and artificial intelligence are developing at an extremely rapid pace. Intelligent robots are moving from laboratories to complex real-world environments, for instance, in the field of autonomous driving, security, rescue, and even military applications. Simultaneous Localization and Mapping (SLAM) technology, as the core support for robots, endows them with the ability of autonomous navigation and environmental perception. The performance of SLAM is a direct factor determining the operational level of the robot. However, traditional SLAM is only applicable to ideal environments that are basically stationary, well-lit,

ICMEEA 2026, April 24-26, Singapore, Singapore

DOI 10.4108/eai.24-4-2026.2364850

and have rich visual details. When dealing with complex and extremely dynamic scenarios, the accuracy and reliability of this system will significantly decrease.

Recently, many researchers have studied these problems. For example, some use deep learning features or combine point and line features to help robots see better in dark or plain places. Others use semantic segmentation and geometry rules to find and remove moving objects. Also, some use multiple sensors tightly together and optimization techniques to improve position accuracy when there is no GPS. But, most existing review papers group research by technology type (like visual SLAM or laser SLAM) or by sensor type. They do not systematically focus on and compare solutions for the three main failure situations mentioned above.

This paper aims to review and summarize SLAM research for three complex situations: Texture Loss in Low-Light Images, Images with Severe Motion Blur, and Signal Degradation in Highly Steep Areas. This paper builds a problem analysis framework from the view of "environment-caused SLAM failure." It clearly explains the basic challenges for these three situations. Also, it deeply compares the newest technologies for different situations. Based on this, it summarizes the limits of current methods and points out promising future research directions.

2 Complex Scene Definition and SLAM Challenge Analysis

Texture Loss in Low-Light Image situations means places with bad light or very little visual detail. This makes visual sensors unable to find stable features. The main traits are very bad lighting and few textures. This makes the image quality poor, and old manual feature methods do not work. Common examples are night patrols and underground repair. These often cause the robot to lose its track and fail to locate itself.

A high dynamic interference environment usually refers to a scenario where there are many moving objects, resulting in distorted positioning and mapping. In this scenario, there are a large number of densely moving objects with complex trajectories, which are very likely to cause "false constraints" for the robot. Its application scenarios mainly include navigation robots in shopping malls, as well as robots used in logistics transportation. In these scenarios, the moving objects may be wrongly classified as stationary objects, which leads to the contamination of the map and the inability to avoid obstacles in time.

The signal degradation scenario in high and steep areas refers to an outdoor environment with complex terrain and no GPS signal coverage. This poses a huge challenge for robots to recognize their surrounding environment. The characteristics of such environments include signal absence, limited sensors, and usually extreme features. Robots are usually employed in mountain rescue operations or underground exploration scenarios. If positioning is solely based on the Inertial Measurement Unit (IMU), significant errors will accumulate over time, ultimately leading to the loss of global positioning consistency.

3 Key Technologies for SLAM in Complex Scenarios

3.1 Texture Loss in Low-Light Image Scenarios

Deep Learning Feature-Based SLAM. This method is no longer the same as the traditional manual feature extraction method; instead, it uses deep neural networks to learn features. Stable tracking was achieved in the TUM RGB-D fr3 sequence, and its performance in the UMA-VI low-light sequence was significantly superior to that of ORB-SLAM3. After applying the KNN matching technique, the absolute trajectory error (ATE) was reduced from 0.107 meters to 0.052 meters, a reduction of over 50% [1]. However, this method still has the problems of high computational cost and a lack of semantic information. Moreover, its generalization ability in new scenarios also needs to be further improved [2].

Point-Line Feature Fusion SLAM. This method can simultaneously obtain point features and line features. Using short line segments for fusion enhances the tracking ability of the robot. In low-texture areas, it supplements the geometric information through line features [3]. When tested in some highly complex EuRoC sequences, the accuracy of this method has been improved by 20% - 50% compared to ORB-SLAM3, and the iteration speed has also increased by approximately 20% [4]. However, in some unstructured scenarios and in outdoor environments with dense vegetation, due to the scarcity of line features, the advantages of this method cannot be demonstrated [5].

3.2 High Dynamic Interference Scenarios

Fusion of Optical Flow and Mask R-CNN for SLAM. This method mainly combines optical flow with the Mask-RCNN network to construct a 3D semantic map that is not affected by moving objects [6]. According to the TUM motion dataset, this method has reduced the average time error (ATE) by approximately 90%, and has also enabled the root mean square error (RMSE) to be decreased to the centimeter level. In addition, the processing time for each frame is approximately in milliseconds, which can meet the real-time requirements [6]. However, this method is only applicable to predefined types of movement, such as “people”. When encountering unknown moving objects, its processing capability is weak and it may miss the detection of some moving objects [7].

Dynamic Object Detection and Rejection Based on Semantic Segmentation. This method employs the YOLOv8-seg semantic segmentation network to identify moving objects, and then uses depth information and polar coordinate line geometric rules to filter and identify undesirable features. This resulted in a reduction of the root mean square error from 0.752 meters to 0.013 meters in the fr3/walking_xyz sequence, and an increase of the average time error by over 95% [8]. However, high-precision segmentation networks require significant GPU computing power support, and temporarily stationary objects may be wrongly classified as static targets, thereby affecting the accuracy and precision of positioning [9].

3.3 Signal Degradation in Highly Steep Areas Scenarios

Handheld LiDAR SLAM. This method optimizes end-to-end using a portable LiDAR system (e.g., DJI Livox Mid-360) to construct a globally consistent point cloud map [10]. It employs a SLAM front-end framework integrated with laser radar-IMU (e.g., FAST-LIO2) [11], which can directly register point clouds without the need for feature extraction, thereby enhancing the accuracy and robustness of the system. In the vicinity of the trajectory, the point cloud density can reach 8000 points per square meter, and it still maintains 1500 points per square meter within a 10-meter range. Its ability to capture details is extremely outstanding [10]. Whether in the east-west direction, north-south direction, or height direction, the root mean square error (MAE) of the absolute coordinate accuracy is less than 0.2 meters, meeting the centimeter-level accuracy requirements for most surveying tasks [10]. This method's hardware is highly portable and has a low cost. However, there are still limitations: the accuracy and consistency of the map are highly dependent on the computer performance, especially in the global optimization and closed-loop detection stages [12]. The current methods mainly employ techniques such as statistical filtering (SORF) to deal with noise, but they are unable to actively model and reconstruct moving targets [10].

Multi-source Perception Tightly Coupled Graph Optimization for SLAM. In this method, the front-end couples the lidar and the IMU, and the back-end constructs an attitude map consisting of odometry, closed-loop detection, GNSS global positioning and co-scan constraints. The global consistency problem is solved using graph optimization. This method (similar to DALI-SLAM) reduced the front-end root mean square error by 43.3% on the DALI dataset, and the back-end optimization further decreased the RMSE by 25.2% to 52.4%. This method is applicable to rugged terrain and urban canyons, but its drawback lies in remaining within the assumption of a "static world" and having a relatively weak ability to handle moving objects [13].

4 Comparative Analysis and Evolution Across Technology Pathways

4.1 Technical Comparison Analysis

Evolution in the realm of perceptual robustness. Regarding the problem of texture information loss in low-light images, the development path of SLAM technology has clearly shown a transition from "geometric enhancement" to "learning-based understanding". The feature fusion method of points and lines is the final step in improving the traditional geometric method. By introducing line features to enhance geometric constraints, it effectively compensates for the deficiency of point features in low-texture areas, and improves the positioning accuracy by approximately 20% to 50% [14]. However, this type of method is highly dependent on the structural regularity of the environment, and its performance is still insufficient in unstructured environments such as natural scenes. Deep learning-based SLAM represents a fundamental leap forward [15]. It completely overcomes the limitations of manually designed features by directly learning high-level abstract features from data. This enables stable tracking performance even under extreme conditions, such as intense light changes or completely textureless areas. However, this approach also demands significantly higher computational resources.

Breakthroughs at the level of environmental assumptions. A highly dynamic interference environment has broken the static assumption. The key technical focus lies in the identification and processing of objects. Some known types, such as pedestrians or vehicles, can be removed through semantic segmentation, and the average error time (ATE) has been reduced by 95%. However, this method cannot handle unknown types and may mistakenly classify temporarily stationary objects as static. Combining optical flow and Mask R-CNN technologies can achieve instance segmentation and restoration, making the processing procedure more complete [16,17]. Both of these methods require a balance between computational cost and generalization ability. Compared with geometric paths, it does not rely on semantic labels and uses multi-view geometric consistency to detect dynamic points. Although it has strong generalization ability, it may fail in areas with sparse textures. These two contrasts demonstrate the balance achieved by dynamic SLAM between the breadth of semantic understanding and the universality of geometric constraints.

Integration at the system architecture level. Facing the extreme uncertainty of wild, steep, and untrusted places, technology improves through system-level optimization. This means moving from using one sensor to deeply combining multiple information sources. Handheld LiDAR SLAM shows the great ability of high-precision LiDAR in static places. It makes dense point cloud maps with centimeter accuracy and no ghosting. Its core technology is end-to-end state estimation optimization, but it still assumes the world is mostly static, treating moving objects as noise to be removed. Multi-source perception, tightly coupled graph optimization SLAM, is a more complete and systematic solution. It goes beyond one sensor, deeply combining different information from odometers, LiDAR, vision, and GNSS (if available) at the state estimation level using a graph optimization framework [18]. This design enables the system to cope with various challenges, such as the degradation of LiDAR features in unstructured scenarios and the GNSS failure caused by the urban canyon effect. At the same time, the complexity of the system also imposes higher requirements on the algorithm efficiency and the accuracy of the sensors.

In complex scenarios, a single SLAM technology cannot achieve the best results. From the perspective of development history, the future trend should be to integrate the advantages of all methods. For instance, within a tightly-coupled multi-source framework, integrating deep learning features and the ability to understand moving objects, while also integrating with a lightweight design, significantly enhances the overall performance of the system [19].

4.2 Common Limitations and Core Challenges

In cases of texture loss in low-light images, the perception of the environment will significantly decrease. In extremely dark or nearly texture-less areas, traditional learning features and visual features will fail, resulting in tracking loss. At the same time, the number of point and line features has significantly decreased, which is insufficient to provide sufficient constraints for the SLAM system.

In the high dynamic interference scenario, the existing technology has the problem of overly relying on predefined categories. The existing methods use semantic segmentation to identify “people” or other known moving objects, but they have difficulty handling objects of unknown types or those with changed shapes. Furthermore, when encountering stationary objects, the system might mistakenly identify them as stationary and thus cause map contamination [20].

In the scenarios with steep terrain, the current technology still has two drawbacks. First, they are not good at processing and analyzing moving objects. The existing methods usually assume a “static world” and treat moving objects as noise that needs to be filtered out, rather than conducting perception modeling for them. In the wild environment, there may be wild animals that could pose a threat to the robot’s SLAM system. Second, when the sensor characteristics are weak, the stability of SLAM is insufficient. In environments without obvious geometric features, such as dense forests or sandy areas, laser radar and cameras will lose their ability to identify features, eventually leading to positioning drift or even complete failure.

4.3 Future-Oriented Technology Evolution Pathways

In low-light conditions with texture loss, there are two methods that are insufficiently optimized. Firstly, by adding event cameras, the imaging capability under extreme lighting conditions is enhanced. This type of camera can sample based on changes in brightness. Ordinary cameras cannot capture images due to overexposure or underexposure, but this camera can provide information about the motion contours. Secondly, utilize dense or semantic information. By employing deep learning methods such as dense depth prediction or semantic segmentation, and leveraging scene semantic consistency to assist robots in position estimation, the reliance on sparse geometric features can be reduced.

For Images with Severe Motion Blur situations, the problem can be fixed in two ways: One way is to do motion detection without needing categories. This combines geometry and motion information to remove various moving objects using multi-view geometry consistency or optical flow consistency, without semantic labels. The other way is to add time-based motion modeling. This uses video instance segmentation and multi-object tracking to give objects identity IDs and guess their motion state. This stops temporarily, but still objects from making the map dirty.

For Signal Degradation in Highly-Steep Areas situations, there are two good solutions: First, build a dynamic perception SLAM system. This system should find, track, and clearly show moving objects. It can copy how self-driving cars handle obstacles to support safe navigation. Second, deeply mix multiple sensors with prior information. Tightly combine data from vision, lidar, and others. Use IMU to make up for the main sensors’ limits, while using prior maps as weak global constraints to reduce error drift [21].

5 Limitations

This article has three limitations in its research. Firstly, the technical verification lacks both breadth and depth. The proposed solutions are based on the validation of public datasets such as TUM and do not undergo systematic evaluation and testing on real robot platforms. As a result, the actual application performance varies significantly. Secondly, the understanding of dynamic scenarios is still at an elementary stage, and the proposed solutions are still based on the assumption of a static world. Thirdly, the ability of generalization and universality have not been fully addressed. For instance, the weak-light model performs poorly in generalization under normal lighting conditions [22].

6 Recommendations for Future Development

The future research trends in SLAM will develop in three aspects: Firstly, it will move towards enhancing the understanding of dynamic scenes. Change from “removing moving objects” to “modeling the changing scene”. This means combining event cameras with LiDAR, adding time-based dynamic modeling, and using large models to build semantic prior maps. This improves the ability to handle dynamic and unknown targets. Second, move to lightweight and adaptive methods. Make lightweight models and efficient algorithms for the front end. Design adaptive graph optimization strategies for the back end. Balance accuracy with speed through hardware-algorithm co-design to fit limited device resources. Third, explore open-world and lifelong learning abilities. Rely less on single data sources by using multimodal fusion. Design incremental learning mechanisms for self-updating. Break closed-loop perception through natural language interaction. Build a “lifelong SLAM” system to improve generalization.

7 Conclusion

This paper has provided a systematic review of SLAM technologies developed for three challenging scenarios: Texture Loss in Low-Light Images, Images with Severe Motion Blur, and Signal Degradation in Highly Steep Areas. Based on the detailed analysis and comparison mentioned above, the following key conclusions can be drawn. No single technology can perfectly solve all the problems encountered in complex environments. Every method for any scenario has its own advantages and disadvantages. For instance, deep learning features perform exceptionally well in low-texture scenarios but require substantial computing resources; the point-line fusion technique can enhance accuracy in structured environments, but its effect is not as significant in unstructured environments; semantic methods can efficiently identify and eliminate known dynamic objects, but they cannot handle unknown categories. The technology of multi-sensor fusion can enhance the accuracy of positioning, but it also makes the system more complex. The current development trend of SLAM is moving from simplicity to complexity, from using a single sensor to integrating multiple sensors for fusion. SLAM technology still faces three challenging problems in real-world environments: assuming a “static scene”, its adaptability to complex environments is limited, and it is restricted by the availability of computing resources on platforms with limited capabilities. The key focus for future development should lie in the understanding of dynamic scenarios, the optimization of computing resources, and the continuous learning ability of robots in an open environment. This article’s comparison of various scenario methods provides a reference for the future development of SLAM systems in the fields of autonomous driving, emergency response, and service robots.

References

[1] Peng, Q., Zhao, X., Dang, R., Xiang, Z.: RWT-SLAM: Robust visual SLAM for weakly textured environments. In: 2024 IEEE Intelligent Vehicles Symposium (IV), pp. 913–920. IEEE, Piscataway (2024).

[2] Xu, K., Hao, Y., Yuan, S., Wang, C., Xie, L.: AirSLAM: An efficient and illumination-robust point-line visual SLAM system. IEEE Transactions on Robotics 41, 1674–1692 (2025).

[3] Liu, Y., Feng, Z., Zhang, H., Dong, W.: Post-integration-based point-line feature visual SLAM in low-texture environments. Scientific Reports 15(14606), 1–13 (2025).
[4] He, J., Li, M., Wang, Y., Wang, H.: PLE-SLAM: A visual-inertial SLAM based on point-line features and efficient IMU initialization. IEEE Sensors Journal 25(4), 6801–6811 (2025).
[5] Jiang, X., Yu, F., Song, S., Wang, H., Cheng, W., Zhu, W.: On visual SLAM algorithm based on point-line features in weak texture environment. In: 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), pp. 286–290. IEEE, Piscataway (2023).
[6] Zhang, K. et al.: A semantic visual SLAM based on improved mask R-CNN in dynamic environment. Robotica 42(10), 3570–3591 (2024).
[7] Tao, C., Jin, Y., Cao, F., Zhang, Z., Li, C., Gao, H.: 3D semantic VSLAM of indoor environment based on mask scoring RCNN. Discrete Dynamics in Nature and Society 2020, Article ID 5916205, 14 pages (2020).
[8] Ji, Q., Zhang, Z., Chen, Y., Zheng, E.: DRV-SLAM: An adaptive real-time semantic visual SLAM based on instance segmentation toward dynamic environments. IEEE Access 12, 43827–43837 (2024).
[9] Peng, Y., Xu, R., Xu, Y., Wu, Y.: A visual SLAM method based on semantic segmentation network in dynamic scenes. In: Proc. Unknown Conference (2023).
[10] Gao, H., Kong, G., Huang, X., Fan, H., Zhong, R.: RobotSLAM: A lightweight low-cost 3D LiDAR SLAM handheld device. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-G-2025, 497–503 (2025).
[11] Guadagnino, T., Mersch, B., Gupta, S., Vizzo, I., Grisetti, G., Stachniss, C.: KISS-SLAM: A simple, robust, and accurate 3D LiDAR SLAM system with enhanced generalization capabilities. IEEE Robotics and Automation Letters 10(4), 1–8 (2025).
[12] Liu, Z. et al.: Voxel-SLAM: A complete, accurate, and versatile LiDAR-inertial SLAM system. arXiv preprint (2024).
[13] Wu, W., Chen, C., Yang, B., Zou, X., Liang, F., Xu, Y., He, X.: DALI-SLAM: Degeneracy-aware LiDAR-inertial SLAM with novel distortion correction and accurate multi-constraint pose graph optimization. ISPRS Journal of Photogrammetry and Remote Sensing 221, 92–108 (2025).
[14] Qin, Y.X., Zhou, W.J., Liu, J.L., Chen, Y., Wang, X.G., Chen, L.L.: Integrating voxel mapping with deep network-based point-line feature fusion for robust SLAM. In: PLOS ONE, 21(1), e0337917 (2026).
[15] Kazi, K., Kalhoro, A.N., Memon, F., Memon, A.R., Iqbal, M.: Beyond Handcrafted Features: A Deep Learning Framework for Optical Flow and SLAM. In: J. Imaging, 11, 155 (2025).
[16] Taghavi, S., Razavi, A.K., Fotouhi, F.: REDO-SLAM: Robust Efficient Dynamic Optical Flow-based SLAM with Deep Reinforcement Learning. In: 2024 18th International Conference on Control, Automation, Robotics and Vision (ICARCV), pp. 1086–1093. IEEE, Singapore (2024).
[17] Hao, H., Ling, Y., Wang, Y., Yu, F.: Visual SLAM Based on Semantic Segmentation and Optical Flow in Dynamic Scenes. In: IEEE Access, 13, 177797–177808 (2025).
[18] Cai, H., Jin, S., Lou, B., Hu, E.: H-GLIVO: A Hierarchically Tightly-Coupled SLAM Framework for GNSS, LiDAR, Inertial, and Visual Odometry. In: 2026 38th Chinese Control and Decision Conference (CCDC), pp. 7321–7328. IEEE, Nanjing (2026).
[19] Yang, Z.: SLAM and Autonomous Navigation Optimization for Unstructured Environments: A Simulation Study Based on End-to-End Multimodal Fusion Algorithms. In: Applied and Computational Engineering, 242, 40–47 (2026).
[20] Bescos, B., Fácil, J.M., Civera, J., Neira, J.: DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes. In: IEEE Robotics and Automation Letters, 3(4), 4076–4083 (2018).
[21] He, W., Li, R., Liu, T., Yu, Y.: LiDAR-based SLAM pose estimation via GNSS graph optimization algorithm. Measurement Science and Technology 35(9), 096304 (2024).
[22] Yang, G., Liu, P., Ma, W.: Enhanced RGB-D SLAM through orthogonal plane constraints and point-line-plane collaborative optimization. PLOS ONE 20(9), e0330839 (2025).