

Simultaneous Localization and Mapping

Subjects: Robotics

Contributor: Haihan Zhang, Chun Xie, Hisatoshi Toriya, Hidehiko Shishido, Itaru Kitahara

Simultaneous Localization and Mapping (SLAM) forms the foundation of vehicle localization in autonomous driving. Utilizing high-precision 3D scene maps as prior information in vehicle localization greatly assists in the navigation of autonomous vehicles within large-scale 3D scene models.

Keywords: SLAM ; 3D reconstruction ; point cloud completion

1. Introduction

In recent years, autonomous driving technology has developed rapidly, paralleling the widespread adoption of advanced driver assistance systems (ADAS). Various sensors, such as radar and cameras, have been widely integrated into vehicles to enhance driving safety and facilitate autonomous navigation ^{[1][2]}. As vehicles mainly operate in urban-scale outdoor environments, it is crucial to accurately and dynamically locate the vehicle's position in the global environment. SLAM technology enables vehicles and robots to locate and map in the environment through visual sensors ^[3] or other perception methods, such as lasers ^{[4][5]}. Traditional SLAM algorithms calculate coordinates based on the initial pose of the first frame, disregarding the world coordinate system in large-scale environments. Assisted localization using prior maps can integrate the initial pose of the first frame in traditional localization algorithms into the world coordinate system of large-scale environments. Two primary objectives are associated with the use of high-precision maps in SLAM systems. Firstly, it provides a world coordinate system for localization, facilitating its extension to applications such as navigation. It also ensures the accurate establishment of system relationships in regions with weak Global Navigation Satellite System (GNSS) signals. Secondly, by leveraging the prior information provided by high-precision maps, the optimization of localization problems caused by scale and feature-matching errors can be achieved. To implement navigation in Global Navigation Satellite Systems (GNSS) Challenge environments, route planning, and danger prediction applications in large scene maps, using known large scene maps as prior information to optimize vehicle localization has emerged as a key solution in SLAM-related research.

Many studies have explored using on-board sensors such as binocular cameras ^[6], radar ^[7], etc., to drive on the road in advance and reconstruct high-precision maps to obtain prior information for providing global information for vehicle localization. However, constructing high-precision maps ^[8] requires expensive hardware requirements, specific driving path planning, manual map repairs, etc., which can result in significant production costs. Furthermore, it is difficult to incorporate comprehensive city information, such as high-rise buildings, when creating city-scale 3D scenes. These limitations in cost and functionality have hindered the ability of vehicles to perform autonomous driving tasks that require global information, such as precise navigation in city maps or full-scale city model reconstruction. Aerial visual sensors are fused with ground visual sensors to address these SLAM challenges and adapt to city-scale 3D scenes for vehicle localization. This research provides a new perspective for understanding large-scale 3D scenes by integrating aerial and ground viewpoints.

The aerial visual sensor consists of a monocular camera mounted on UAVs, providing an aerial view of the city. This aerial view is used to reconstruct a prior 3D large scene model at the business district scale. The ground visual sensor includes two pairs of stereo cameras mounted on the vehicle, capturing the ground view. This ground view is used for vehicle localization within the large scene. However, significant differences exist in the visual information obtained from these distinct views ^[9]. The aerial view primarily covers upper-level information, such as building rooftops but lacks accurate side-level information due to factors like occlusion and pixel size limitations. Conversely, the ground view mainly contains information about building walls and lacks upper-level information that could be linked to the aerial view. Consequently, establishing a relationship between aerial and ground views is challenging. A point cloud completion algorithm based on the building's geometric structure is proposed to establish associations between these views. The system achieves accurate vehicle localization in large-scale scenes. The localization error using a low-precision map as prior information is 3.91 m, which is close to the error of 3.27 m achieved when using a high-precision map, thereby attaining a comparable

level of performance. A generic timeline of the entire process is shown in **Figure 1**. The algorithm is evaluated using a 3D city model in a CG simulator as the ground truth.

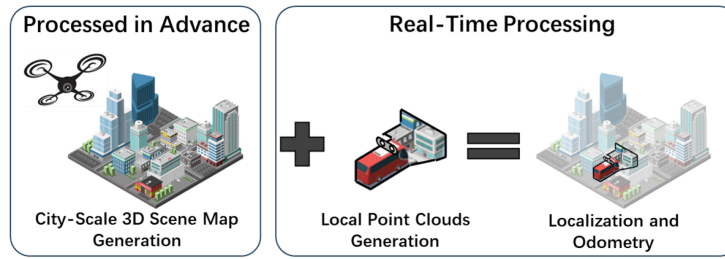


Figure 1. A generic timeline of the entire process.

2. Three-Dimensional Reconstruction with UAVs

Reconstructing a 3D large scene model using UAVs is the most efficient method in 3D urban modeling. Numerous approaches have been proposed to address 3D city map reconstruction challenges. Bashar et al. proposed 3D city modeling based on Ultra High Definition (UHD) video [10], using UHD video acquired from UAVs for 3D reconstruction and comparing the reconstructed point cloud density with HD video results. While UHD video improves point cloud quality, the orthoscopic view from UAVs causes some information loss due to the occlusion problem with 2D figures. In 2021, Chen et al. proposed a 3D reconstruction system for large scenes using multiple UAVs, optimizing the trajectory drift problem in multi-drone tracking [11]. Increasing the number of captured images can partially reduce missing information resulting from occlusion issues. However, high-precision 3D map reconstruction with multiple UAVs entails increased labor and time costs during operation. Omid et al. employed UAV radio measurements for 3D city reconstruction [12]. Although radio measurements provide higher ground reconstruction accuracy than visual sensors, they cannot avoid missing information due to occlusions such as building roofs in aerial views. Furthermore, the reconstructed model lacks RGB information, which is unfavorable for the potential rendering or visualization of the 3D city model. In 2022, Huang et al. employed the topological structure of building rooftops to complete the missing wall surfaces in drone-based 3D reconstruction [13]. However, this approach utilized a higher-precision radar for reconstruction and incorporated both interior and exterior wall information during the completion process. For vehicles, the interior wall surfaces of buildings act as noise and can cause drift in the localization process. In the system, a point cloud completion algorithm based on architectural geometry is proposed to address the issue of missing building walls caused by occlusion in the drone's frontal view. This method selectively preserves only the exterior wall information, which is more relevant to the building, while retaining the RGB information from the visual sensor. Simultaneously, it resolves potential occlusion problems.

3. Prior Map-Based SLAM

SLAM is a widely used technology in vehicle localization and mapping. Applying SLAM to large scenes involves calculating the relative pose relationship between the camera coordinate system, the world coordinate system, and the image coordinate system. Most traditional visual SLAMs match feature points in each image frame to track the camera, only establishing the pose relationship between frames but not localizing the camera coordinates in large scenes. Research on pose optimization in SLAM using prior information generally relies on in-vehicle sensors [14][15] or high-precision maps [16]. On the one hand, in-vehicle sensors only provide perspective information limited to the area surrounding the vehicle (e.g., roads or bridges) [17], which is hard to obtain the whole city map. During the mapping process, moving on the road for map creation also carries significant restrictions in terms of freedom to compare with the UAV movements. Furthermore, to ensure that the features of the prior map can match the features obtained during the localization process, the same sensor is generally used for mapping and localization. This imposes many restrictions specific to the sensor for the SLAM system. On the other hand, creating high-precision maps for city-scale models is quite expensive as they require costly equipment, such as Light Detection and Ranging (LiDAR), and sometimes manual repairs are needed [18]. A method is proposed for assisting vehicle positioning by obtaining low-precision prior large scene maps using inexpensive drones. The method uses drones with higher degrees of freedom to obtain prior maps and achieves a localization accuracy close to that of using high-precision maps while complementing the city model reconstructed by the UAVs. This research addresses the cost challenges related to high-precision map creation and also expands the possibilities for using 3D prior maps for localization.

4. Aerial-to-Ground SLAM

Substantial works are related to utilizing aerial perspectives to aid in ground-based positioning. For instance, Kummerle et al. [19] achieved impressive results by assisting radar-based positioning with global information from aerial images. In 2022, Shi et al. [20] used satellite images to assist in ground-based vehicle localization, helping the vehicles acquire accurate global information. These methods involve extracting features from 2D aerial images and matching them with ground-based information. Compared to 2D images, the feature set from 3D information obtained after urban reconstruction is more abundant. However, due to the immense volume of 3D information, its use in assisting localization can impact the real-time performance of the system. Moreover, there is a significant discrepancy between the 3D information computed from aerial data and that from ground data, making the match between the aerial and ground perspectives even more challenging to achieve. This research proposes a method that uses the 3D information from drone reconstruction to assist in vehicle localization. Furthermore, by segmenting the global point cloud map, the data volume for vehicle localization is reduced to ensure the system's real-time performance.

References

1. Ibragimov, I.Z.; Afanasyev, I.M. Comparison of ROS-based visual SLAM methods in homogeneous indoor environment. In Proceedings of the 2017 14th Workshop on Positioning, Navigation and Communications (WPNC), Bremen, Germany, 25–26 October 2017; pp. 1–6.
2. Blochliger, F.; Fehr, M.; Dymczyk, M.; Schneider, T.; Siegwart, R. Topomap: Topological Mapping and Navigation Based on Visual SLAM Maps. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 3818–3825.
3. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* 2015, 31, 1147–1163.
4. Shan, T.; Englot, B. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 4758–4765.
5. Zhao, S.; Zhang, H.; Wang, P.; Nogueira, L.; Scherer, S. Super odometry: IMU-centric LiDAR-visual-inertial estimator for challenging environments. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 8729–8736.
6. Cui, L.; Wen, F. A monocular ORB-SLAM in dynamic environments. *J. Phys. Conf. Ser.* 2019, 1168, 052037.
7. Wang, Q.; Zhang, J.; Liu, Y.; Zhang, X. High-Precision and Fast LiDAR Odometry and Mapping Algorithm. *J. Adv. Comput. Intell. Inform.* 2022, 26, 206–216.
8. Liu, R.; Wang, J.; Zhang, B. High definition map for automated driving: Overview and analysis. *J. Navig.* 2020, 73, 324–341.
9. Fruh, C.; Zakhori, A. Constructing 3D city models by merging aerial and ground views. *IEEE Comput. Graph. Appl.* 2003, 23, 52–61.
10. Alsadik, B.; Khalaf, Y.H. Potential Use of Drone Ultra-High-Definition Videos for Detailed 3D City Modeling. *ISPRS Int. J. Geo-Inf.* 2022, 11, 34.
11. Chen, F.; Lu, Y.; Cai, B.; Xie, X. Multi-Drone Collaborative Trajectory Optimization for Large-Scale Aerial 3D Scanning. In Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Bari, Italy, 4–8 October 2021; pp. 121–126.
12. Esrafilian, O.; Gesbert, D. 3D city map reconstruction from UAV-based radio measurements. In Proceedings of the GLOBECOM 2017—2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
13. Huang, J.; Stoter, J.; Peters, R.; Nan, L. City3D: Large-scale building reconstruction from airborne LiDAR point clouds. *Remote Sens.* 2022, 14, 2254.
14. Sasiadek, J.Z.; Ahmed, A. Multi-Sensor Fusion for Navigation of Ground Vehicles. In Proceedings of the 2022 26th International Conference on Methods and Models in Automation and Robotics (MMAR), Międzyzdroje, Poland, 22–25 August 2022; pp. 128–133.
15. Liu, Y.; Yang, D.; Li, J.; Gu, Y.; Pi, J.; Zhang, X. Stereo Visual-Inertial SLAM with Points and Lines. *IEEE Access* 2018, 6, 69381–69392.

16. Liu, J.; Xiao, J.; Cao, H.; Deng, J. The Status and Challenges of High Precision Map for Automated Driving. In Proceedings of the China Satellite Navigation Conference (CSNC), Beijing, China, 22–25 May 2019; Sun, J., Yang, C., Yang, Y., Eds.; Springer: Singapore, 2019; pp. 266–276.
17. Guizilini, V.; Ambrus, R.; Pillai, S.; Raventos, A.; Gaidon, A. 3D packing for self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2485–2494.
18. Lin, L.; Liu, Y.; Hu, Y.; Yan, X.; Xie, K.; Huang, H. Capturing, Reconstructing, and Simulating: The UrbanScene3D Dataset. In Proceedings of the Computer Vision—ECCV 2022, Tel Aviv, Israel, 23–27 October 2022; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer: Cham, Switzerland, 2022; pp. 93–109.
19. Kümmerle, R.; Steder, B.; Dornhege, C.; Kleiner, A.; Grisetti, G.; Burgard, W. Large scale graph-based SLAM using aerial images as prior information. *Auton. Robot.* 2009, 30, 25–39.
20. Shi, Y.; Li, H. Beyond cross-view image retrieval: Highly accurate vehicle localization using satellite image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17010–17020.

Retrieved from <https://encyclopedia.pub/entry/history/show/109096>