

Around View Monitor-Based Visual Simultaneous Localization and Mapping

Subjects: **Others**

Contributor: Yangwoo Lee , Minsoo Kim , Joonwoo Ahn , Jaeheung Park

Accurately estimating the pose of a vehicle is important for autonomous parking. The study of around view monitor (AVM)-based visual Simultaneous Localization and Mapping (SLAM) has gained attention due to its affordability, commercial availability, and suitability for parking scenarios characterized by rapid rotations and back-and-forth movements of the vehicle.

visual Simultaneous Localization and Mapping

autonomous parking

AVM distortion error

deep learning

1. Introduction

One of the important technologies for the commercialization of fully autonomous parking is to accurately estimate the pose of the vehicle. Simultaneous Localization and Mapping (SLAM) technology has been developed for localizing the vehicle [1]. Recently, visual SLAM [2], which utilizes cameras as the main sensor, has demonstrated significant advancements in autonomous vehicles, wherein a front camera is commonly employed to estimate the vehicle pose by detecting changes in feature locations or intensity variations of pixels within the image [3]. However, visual SLAM based on a front camera is less suitable for parking scenarios due to the narrow field of view (FOV) of the sensor and a motion bias problem [4] in which the performance of SLAM in forward and backward motion is different.

Instead of a front camera, an around view monitor (AVM) has been also investigated for its application to autonomous parking in visual SLAM. AVM provides a bird's eye view image using cameras facing in four different directions. Many studies [5][6][7] have applied AVM-based visual SLAM to parking scenarios by taking advantage of wide FOV and no motion bias. These studies have used road-marking information as semantic features to avoid the deformation caused by Inverse Perspective Mapping (IPM). However, inaccurate pose estimation occurs due to AVM distortion error caused by uneven ground and inaccurate camera calibration.

Despite the fact that numerous approaches [8][9][10][11][12] have been explored to achieve accurate bird's eye view images, they still exhibit several limitations, including limited quantitative comparison between features observed in AVM and features in real-world environments and substantial efforts for collecting data. Alternatively, in [5][6], they avoided the influence of distortion errors by utilizing an additional Inertial Measurement Unit (IMU) sensor based on a pre-built map or leveraging an externally provided High Definition (HD) vector map. The approach of [7] attempted

to create an accurate map in real-time using the sliding window fusion technique without additional information, but it exhibited insufficient accuracy in pose estimation for autonomous parking.

2. Visual SLAM

2.1. Front-Camera-Based Visual SLAM

In the field of visual SLAM, there are various methods using a front camera as the main sensor. These can be broadly categorized into direct and feature-based methods. Direct methods use the raw image, while feature-based methods use only features in the image. Representative direct methods of visual SLAM include Direct Sparse Odometry (DSO), which estimates camera pose by minimizing the pixel intensity difference between two images [13]. ORB-SLAM2 [14] is a well-known feature-based method that estimates camera pose by minimizing the distance difference between matched feature points. Another method, Semi-direct Visual Odometry (SVO), combines the advantages of both direct and feature-based methods [15]. However, these methods are not suitable for autonomous parking. A front camera with a narrow FOV can have challenges detecting changes in the surrounding environment of a vehicle during rapid rotation. Moreover, depth estimation in these methods causes a difference in SLAM performance between forward and backward motion [4].

2.2. AVM-Based Visual SLAM

Various SLAM studies have proposed utilizing AVM as the main sensor for autonomous parking. Autonomous valet parking-SLAM (AVP-SLAM) [5] estimates the pose of the vehicle by matching current road-marking features with a road-marking map created in the preceding process. AVP-Loc [6] matches current road-marking features to an HD vector map of a parking lot for vehicle localization. In [7], the edge information of both free space and road markings are utilized as input features to estimate the pose of a vehicle using Iterative Closest Point (ICP) [16]. However, the AVM distortion error was not taken into account in these studies. In [5], the inherent inaccuracies caused by the distortion error in the pre-built map necessitated an additional IMU to enhance the localization accuracy. The method of [6] employed an externally provided HD vector map to avoid distortion errors. Although the approach in [7] aimed to create an accurate map in real-time using the sliding window fusion technique without additional information, the construction of maps that include distortion errors negatively affects the localization performance of autonomous vehicles.

3. AVM Image Enhancement Techniques

3.1. AVM Image Modification Using Automatic Calibration

Several methods have been proposed for the automatic calibration of AVM images, utilizing extracted feature shapes to achieve accurate bird's eye views. In [8], a method was proposed to estimate the extrinsic parameters of four AVM cameras using point patterns. The method proposed in [9] aims to calibrate an AVM image by matching the gradient and position of the lane observed from the front and rear cameras with that seen from the left and right

cameras. On the other hand, the method proposed in [10] focuses on calibrating the AVM image to ensure the detected parking lines are parallel or perpendicular to the vehicle. However, since these methods only perform relative comparisons between the images of each camera and do not quantitatively compare the AVM image with the real environment, it may not be considered as a complete solution to address AVM distortion errors.

3.2. AVM Image Generation Using Deep Learning

In recent years, there has been a shift towards deep learning-based AVM image generation methods that use Neural Networks (NN) as viewpoint transformation functions, instead of using homography geometric information. HDMapNet [11] uses a Multilayer Perceptron (MLP) to convert each camera image from AVM into a bird's eye view, and then creates an AVM image through an attaching process using installed camera locations. Another method, BEVformer [12], uses the Transformer model to create an AVM image without an additional attaching process. However, these methods require a large amount of training data.

To estimate the vehicle pose without the need for pre-map construction or additional sensors, the algorithm of the Hybrid Bird's-Eye Edge-Based Semantic Visual SLAM [7] is selected. Instead of using the method directly, it was modified in two aspects to improve the pose estimation performance. First, parking line information is utilized as input data instead of edge information from semantic features and free space. Second, Generalized ICP [17] is used instead of ICP for pose estimation. GICP is a pointcloud registration algorithm developed from ICP that addresses inaccurate correspondences. The modified Hybrid Bird's-Eye Edge-Based Semantic Visual SLAM is used to examine the impact of distortion errors, resulting from inaccurate AVM calibration, on the performance of AVM-based SLAM.

References

1. Smith, R.C.; Cheeseman, P. On the representation and estimation of spatial uncertainty. *Int. J. Robot. Res.* 1986, 5, 56–68.
2. Karlsson, N.; Di Bernardo, E.; Ostrowski, J.; Goncalves, L.; Pirjanian, P.; Munich, M.E. The vSLAM algorithm for robust localization and mapping. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005*; IEEE: New York, NY, USA, 2005; pp. 24–29.
3. Chen, W.; Shang, G.; Ji, A.; Zhou, C.; Wang, X.; Xu, C.; Li, Z.; Hu, K. An overview on visual slam: From tradition to semantic. *Remote Sens.* 2022, 14, 3010.
4. Yang, N.; Wang, R.; Gao, X.; Cremers, D. Challenges in monocular visual odometry: Photometric calibration, motion bias, and rolling shutter effect. *IEEE Robot. Autom. Lett.* 2018, 3, 2878–2885.
5. Qin, T.; Chen, T.; Chen, Y.; Su, Q. Avp-slam: Semantic visual mapping and localization for autonomous vehicles in the parking lot. In *Proceedings of the 2020 IEEE/RSJ International*

- Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; IEEE: New York, NY, USA, 2020; pp. 5939–5945.
6. Zhang, C.; Liu, H.; Xie, Z.; Yang, K.; Guo, K.; Cai, R.; Li, Z. AVP-Loc: Surround View Localization and Relocalization Based on HD Vector Map for Automated Valet Parking. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; IEEE: New York, NY, USA, 2021; pp. 5552–5559.
 7. Xiang, Z.; Bao, A.; Su, J. Hybrid bird's-eye edge based semantic visual SLAM for automated valet parking. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; IEEE: New York, NY, USA, 2021; pp. 11546–11552.
 8. Natroshvili, K.; Scholl, K.U. Automatic extrinsic calibration methods for surround view systems. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; IEEE: New York, NY, USA, 2017; pp. 82–88.
 9. Choi, K.; Jung, H.G.; Suhr, J.K. Automatic calibration of an around view monitor system exploiting lane markings. *Sensors* 2018, 18, 2956.
 10. Lee, Y.H.; Kim, W.Y. An automatic calibration method for AVM cameras. *IEEE Access* 2020, 8, 192073–192086.
 11. Li, Q.; Wang, Y.; Wang, Y.; Zhao, H. Hdmapnet: An online hd map construction and evaluation framework. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; IEEE: New York, NY, USA, 2022; pp. 4628–4634.
 12. Li, Z.; Wang, W.; Li, H.; Xie, E.; Sima, C.; Lu, T.; Qiao, Y.; Dai, J. Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In Proceedings of the Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; Proceedings, Part IX. Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–18.
 13. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 40, 611–625.
 14. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* 2017, 33, 1255–1262.
 15. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; IEEE: New York, NY, USA, 2014; pp. 15–22.
 16. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*; SPIE: Bellingham, WA, USA, 1992; Volume 1611, pp. 586–606.

17. Segal, A.; Haehnel, D.; Thrun, S. Generalized-icp. In Proceedings of the Robotics: Science and Systems, Seattle, WA, USA, 28 June–1 July 2009; Volume 2, p. 435.

Retrieved from <https://encyclopedia.pub/entry/history/show/114023>