

Development of Cloud-Based Parking Map Generation Technology for Automated Valet Parking

Morihiko Sakano¹⁾ Takahiro Sakai¹⁾ Yoshinobu Ogasawara²⁾ Hidehiro Toyoda²⁾

1) Hitachi, Ltd.

7-1-1 Omika, Hitachi, Ibaraki, 319-1292, Japan (E-mail: morihiko.sakano.ty@hitachi.com)

2) Astemo, Ltd.

2520, Takaba, Hitachinaka, Ibaraki, 312-8503, Japan

KEY WORDS: Parking Map Generation, Structure from Motion, Localization, Dual Descriptor Map

Automated Valet Parking (AVP) has emerged as a promising application of automated driving technologies, particularly in low-speed and spatially constrained parking environments. Since GNSS signals are often unavailable in indoor or underground parking facilities, reliable camera-based self-localization and high-precision parking maps are essential. Such maps must include detailed drivable lane structures, parking space locations, and landmarks suitable for localization. Cloud-based map generation using sensor data collected from multiple vehicles is an effective approach to construct large-scale parking maps without relying on infrastructure sensors. However, large-scale map generation based on Structure from Motion (SfM) involves high computational costs, and deep learning-based localization is generally too computationally intensive for real-time execution on in-vehicle edge devices.

This study proposes a cloud-based parking map generation and high-speed localization framework for AVP that addresses these challenges by introducing a dual-descriptor map representation and a scalable map generation strategy. In the proposed system, multi-camera image data collected from multiple vehicles are uploaded to the cloud, where a high-precision parking map is generated. The map is then distributed to vehicles, where localization and automated driving are performed on edge devices.

To achieve both high map accuracy and real-time localization performance, a dual-descriptor map is introduced as shown in Fig.1. During map generation, deep learning-based local features are employed to ensure robust and accurate SfM reconstruction. In parallel, handcrafted features are extracted and triangulated using camera poses estimated by SfM, allowing both deep learning-based and handcrafted features to coexist within a single map. During vehicle operation, localization is performed using only handcrafted features, enabling fast pose estimation suitable for real-time processing. Any loss in localization accuracy is compensated by sensor fusion with wheel odometry using a Kalman filter, achieving a well-balanced trade-off between accuracy and computational efficiency.

For large parking facilities, the number of input images can exceed tens of thousands, causing SfM processing time to grow prohibitively due to bundle adjustment complexity. To mitigate this issue, a divide-and-conquer strategy is adopted. The input image sequence is partitioned into smaller subsets, and independent maps are generated for each subset. These sub-maps are then efficiently merged using geometric consistency constraints without additional bundle adjustment. This approach enables practical processing times even for very large-scale parking environments.

In addition to point cloud map generation, vehicle trajectories estimated by SfM are utilized to automatically generate node-link structures representing drivable paths in the parking lot. Furthermore, road markings such as parking spaces are projected into the map coordinate system by estimating the road surface plane from camera trajectories and point cloud distributions. This allows consistent integration of structural map information required for path planning and automated driving.

Experimental evaluations using real-world data collected in a large commercial parking facility demonstrated the effectiveness of the proposed approach. The average localization error was 0.14 m, with approximately 95% of the results within 0.5 m. The average localization processing time was 46 ms, satisfying real-time requirements for AVP systems. Moreover, a large-scale parking map generated from over 30,000 images was successfully completed as shown in Fig.2 in approximately 10 hours using the proposed scalable map generation method, confirming its practicality for real-world AVP deployment.

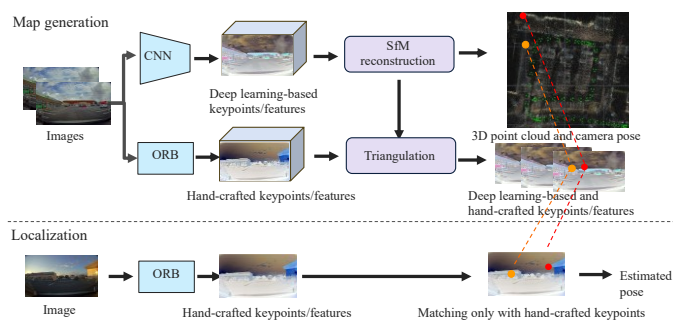


Fig.1 Block diagram of proposed method (Dual Descriptor map gen.)

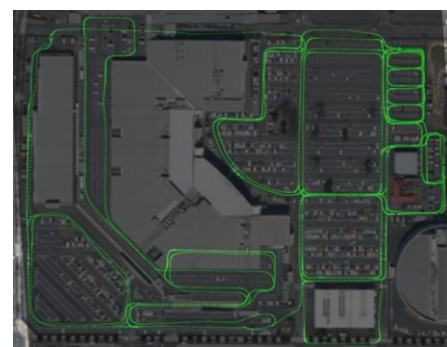


Fig.2 Integrated map