Title	Combining Lidar and Camera via Attention Mechanism for Simultaneous Localization and Mapping in Outdoor Environments
Author(s)	LAM, Kha Han
Citation	
Issue Date	2022-12
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18165
Rights	
Description	Supervisor: 吉高 淳夫, 先端科学技術研究科, 修士(情報科学）

Combining Lidar and Camera via Attention Mechanism for Simultaneous Localization and Mapping in Outdoor Environments

2010437 LAM Kha Han

Autonomous vehicles are self-control of moving on their own without human intervention. Autonomous vehicles are important in reducing pressure and distraction for drivers to ensure safety when participating in traffic. The area of autonomous vehicles is attractive to researchers. Many problems need to be solved to develop autonomous vehicles, such as navigation, path planning, obstacle avoidance, etc.

Simultaneous Localization and Mapping (SLAM) aims to estimate the position and reconstruct the map of the autonomous vehicle. SLAM is vital in autonomous vehicles and other areas such as drones, recuse systems, robot navigation, etc. SLAM consists of two main modules: front-end and back-end. The front-end module with the other name is odometry, which aims to localize the vehicle's pose and draw a map of the unknown environment by information collected by sensors mounted in this vehicle. Some external factors (e.g., persons, cars) affect odometry estimation. Therefore, the back-end module optimizes the odometry module result. This thesis focus on the odometry module of the SLAM component.

The purpose of odometry is to estimate the autonomous vehicle's path while it is on the fly. Because of work in outdoor environments, autonomous vehicles do not know spatial information or surrounding objects. So, the odometry depends on data collected by sensors. Several sensors can be applied to this problem. Cameras are commonly used since they are inexpensive and convenient. However, RGB cameras without depth information are insufficient for estimating the position, while RGB-D cameras are sensitive to light in outdoor environments. An alternative is to use LiDAR sensors to capture depth information in outdoor environments. This thesis combines visual and depth information from the camera and LiDAR for the odometry of SLAM in outdoor environments.

The main challenge when designing an odometry system is combining information from sensors. Traditional methods use a handcraft feature extracting and matching pipeline to estimate the vehicle's pose. These methods archive good performance but depend on handcraft feature extraction designs. Deep learning-based approaches have devised breakthroughs in various topics in recent years. The learning-based approach applies in the multi-stage of the odometry framework. The advantage of learning-based methods is independent of the geometric model or handcrafted features. Therefore, this thesis follows the learning-based odometry approach.

1

The purpose of this work is to combine information from the camera and LiDAR in odometry. The thesis's main contribution is to propose an end-to-end Visual-LiDAR odometry via an attention-driven mechanism. The proposed method extracts the essential regions from RGB images by a self-attention layer instead of feature points in previous works. A region on an image consists of many adjacent points, which is more robust than individual feature points. In previous works, the depth information from the 3D point cloud is used to complement the feature from 2D images. Guided attention is applied to emphasize the interaction between depth and visual information.

The proposed method is evaluated using the KITTI dataset, a data set in traffic with pre-collected image sequences, 3D points cloud from Lidar, and vehicle GPS - considered as ground truth. The estimation result is compared with other learning-based odometry methods.