報告番号	※甲　　　第　　　　号

主　論　文　の　要　旨

論文題目　An End-to-End Multi-Sensor Fusion System for Autonomous Driving Applications
（自動運転向けのエンド・ツー・エンドの複数センサー複合システム）

氏　　名　　MONRROY CANO Abraham Israel

論 文 内 容 の 要 旨

Autonomous Driving technologies promise the reduction of accidents, facilitate the transportation of people, the automation of goods distribution, and other applications which currently suffer from a shortage of personnel to satisfy transportation needs. Self-driving technologies additionally pledge to improve the quality of life of their users and the people in charge of providing the service, such as drivers and distributors. For the above reason, we can consider this technology brings social value, convenience, and benefits to all the persons involved, from its development and production to its end-users.

Navigation robots and self-driving technologies require the precise integration of multiple sensors to help them perceive their surroundings, detect obstacles and often base their decisions on the data provided by these sensors to reach their goals safely. The simultaneous integration of data from multiple sensors is known as fusion, and it is used to overcome weaknesses in each sensor. Case in point, LiDARs (Light Detection and Ranging) fit medium and long-range sensing applications, while cameras are suitable for high-resolution sensing that brings a loss in three-dimensional as a downside. Wide-angle lenses attached to cameras provide a wide-angle field view. However, due to its perspective nature, distant objects appear smaller. On the other hand, cameras equipped with telephoto lenses provide long-range sensing. Nevertheless, their field of view gets significantly narrowed.

State-of-the-art robots use multiple cameras and LiDARs, among other sensors, to satisfy their sensing requirements according to their defined application. As described before, LiDARs and cameras have a limited field of view. For this reason, the integration of multiple LiDARS and cameras is considered when designing these types of vehicles. However, to achieve multi-sensor fusion, it is required to accurately obtain the camera intrinsic parameters and the extrinsic calibration among all the sensors to enable its integration.

These multi-sensing modules can be used in ADAS (Advanced driver-assistance systems) to facilitate the development of reliable safety systems such as Collision Warning Systems, Collision Intervention Systems, and Driving Control Assistance. These systems have significantly improved the safety of the driver and its companions over the years since its inception. Moreover, these systems are constantly improving and bringing social value in the shape of safety, reducing accidents, and, more importantly, protecting the users' lives. Cameras can be used in conjunction with computer vision algorithms to obtain imminent obstacles or feed lane detection modules to notify the driver about dangerous scenarios such as lane departure or pedestrian detection to enable emergency braking systems. LiDARs, on the other hand, thanks to their wide field of view and precise range detection, can be employed to facilitate the detection of occluded objects and the development of other systems such as adaptive cruise control.

Additionally, highly accurate multi-sensor systems are essential in self-driving systems as determined by the SAE (Society of Automotive Engineers) to generate the data required to feed the perception, localization, planning, and control modules. Perception modules use data from multiple sensors to obtain information such as the surrounding obstacles, ego lane information, traffic light state, and others. In a similar manner, localization modules require data from multiple sensors to reduce the inherent measurement and quantization error while using a single sensor. Sensor fusion techniques help mitigate the error accumulation caused while the vehicle moves and produce a more reliable localization estimation.

Significant efforts have been made towards the calibration of each sensor. However, multi-sensor calibration guidelines and fusion frameworks remain unexplored and scarce. In this dissertation, we develop an end-to-end multi-sensor calibration framework to accelerate the advancement of sensing and perception systems that require precise camera-LiDAR calibration.

The first module contained in our framework is the automatic single-shot intrinsic camera calibrator. With the help of a simulator, we generated thousands of synthetic image frames containing multiple calibration targets variating their positioning and rotation, evaluated them, and selected the optimal ones. We then constructed precise guidelines to obtain intrinsic camera parameters accurately using a single image containing multiple targets in a predefined setting; We validated these with real-world cameras and lenses commonly used in robotics and autonomous driving applications. Our findings found that using seven checkerboard targets produces repeatable and accurate camera intrinsic parameters for its use in 3D applications, such as the projection of the point cloud generated by 3D LiDARs.

The multi-LiDARs calibration module in our frame framework can accurately calculate the relative position and angle among them, or in other words, find the extrinsic calibration parameters. Our multi-LiDAR calibration module uses a method that finds the geometric similarities using the normal distribution computed on the voxelized space of each point cloud obtained from each LiDAR. After that, our method uses an optimizer to find the minimum distance between the calculated geometric similarities until it converges, resulting in the translation and rotation that relates both point clouds, and therefore obtaining the relative transformation between LiDARs. This process is repeated sequentially to calibrate all the required LiDARs extrinsically and finally adjusted to minimize the relative error caused by the voxelization step.

The camera-LiDAR extrinsic calibration. It follows a guided approach to find the euclidean transformation between the camera and the LiDAR with the help of the projection matrix obtained in the intrinsic camera parameter calibrator. The user provides feature hints related to the image and the point cloud. These features are fed to an optimizer that follows the PnP method to relate the 3D and 2D features to obtain the desired camera-LiDAR transformation.

The last module in our framework is data preprocessing and fusion. It contains methods to classify point cloud as ground, fuse data obtained from a camera, and a LiDAR using the intrinsic and extrinsic calibration parameters at pixel-cloud level, also known as low-level fusion. Additionally, it contains back-projection methods that can enable higher-level perception using deep learning methods such as semantic or panoptic segmentation, among others.

Finally, we evaluated and validated our methods on multiple sensing devices and platforms, such as robots, data collection vehicles, and Vehicle to Infrastructure (V2I) systems. We found that our framework is accurate to the sub-centimeter level, and it helps accelerate the calibration process of multiple sensors, removing the need for specialized personnel to obtain the parameters.