平成 27 年度 博士論文（要約）

東京大学
THE UNIVERSITY OF TOKYO

RGB-D SLAM and Automatic Calibration of Camera Sensor Network for Mobile Robotic Applications in Indoor Environments

(屋内環境における移動ロボットアプリケーションのための RGB-D SLAM 及びカメラセンサネットワークの自動キャリブレーション)

東京大学大学院 工学系研究科 精密工学専攻

池 勇勳

These days, many kinds of robots have been developed for industrial applications, such as factory automation. However, lately robots are expected to operate not only in industrial areas, but also in service for daily human life (i.e., human-robot coexistence environments, such as offices, department stores, private homes, and so on).

Safe and reliable navigation technology for mobility in such areas is one of the most important requirements for such service robots, in their interaction with humans. The navigation functions for a mobile robot are basically performed based on the map information of an environment. Therefore, an accurate mapping of an unknown environment in advance is a very important task for mobile robots, including localization and path planning. However, considering human-robot coexistence environments, the map information, which is a static model, cannot deal with such dynamic environments because it cannot reflect changes in the environment (e.g., moving objects, etc.).

On the other hand, the concept of an intelligent space, which constructs distributed sensor networks in an external environment, can monitor what is occurring in it. Thus, such a space can manage environmental changes by processing distributed sensor data in real time. Therefore, additional information from the distributed sensor networks can generate the necessary dynamic information. Whereas, the map information built by SLAM is only static and may be insufficient to deal with dynamic information.

Considerable research for dense and accurate map building has been conducted and the reliable theories have already been established. Hence, the current state-of-the-art SLAM methods can generate dense 3D map information with remarkable accuracy in a large environment using light detection and ranging (LiDAR) or RGB-D sensors, which can easily obtain the surrounding 3D information. However, little attention has been given to the use of state-of-the-art map information built by SLAM. Likewise, research about constructing an intelligent space has also been intensively performed over the past decade; however, most research remains as only proposals for the concept. In other words, although a number of studies have been made on building map information and intelligent spaces, a problem in applying those to improving functions of service robots remains unanswered. Therefore, a proposal for a new concept taking full advantages of the combination of these types of information (i.e., the static map information and the dynamic information from the intelligent space) is needed.

In order to cope with these situations and achieve safe and efficient operation of service robots in environments where they co-exist with humans in the future, this thesis suggests a new concept in the following three steps:

- Accurate modeling an indoor environment to build static map information using RGB-D SLAM schemes.
- Construction of an intelligent space using a distributed camera network and developing

automatic calibration frameworks for them.

- Operation of robots by utilizing static and dynamic information from both the built environment model and the constructed intelligent space.

In order to realize the points mentioned above, prior to that, this thesis mainly focuses on the automatic calibration methods for complete external parameters of all distributed camera sensor using the static environment model, which is extremely tedious task if performed manually. Therefore, the specific aims of this thesis can be summarized as follows:

- Proposing novel feature information for an indoor SLAM framework in order to build map information of the target environment for use as space constraints in the calibration process.
- Developing automatic calibration frameworks for obtaining the complete 6 degree of freedom (DOF) parameters of a camera sensor network based on pre-given map information.
- Integrating these types of information (i.e., static, dense data from map information, and sparse, dynamic data from calibrated camera sensor networks) in order to establish a new navigation paradigm for mobile robots.

More detailed summaries for each step are as follows.

In chapter 2, a novel RGB-D SLAM method is introduced in order to build a 3D texture map composed of the popular Octree structure (i.e., 3D texture OctoMap information) and semantic information using a vanishing point and door plates as landmark information. A vanishing point and door plates are usually observed in corridor-like environments. Using these types of reliable features as landmarks maintains the stability of the SLAM. Reliable data association can be performed for the door plates with clearly distinguishable signature (room number information) by registering the information to the whole state vector. In contrast to a conventional corner feature-based SLAM scheme, the proposed approach estimates more accurate trajectories and can also significantly reduce computation time. Moreover, the recognized room number information can be useful as semantic information by combining it into the built map information. The built 3D texture map information is utilized hereafter as the static model for many applications as it can provide all of the coordinate information for the entire environment.

Chapter 3 proposes a novel method to achieve trajectory reconstruction of a mobile robot and calibration of complete 6DOF external parameters of a distributed camera sensor network to construct an intelligent space at the same time. In this study, grid map information of the environment is utilized as prior information that can be used as an additional constraint when finding the optimal solution as the camera networks are generally installed on interior wall planes. The experimental results shows that the proposed framework, which uses maximum a posterior (MAP) estimation-based optimization technique, can simultaneously estimate 6DOF poses of the entire camera network system and the robot trajectory accurately.

In chapter 4, another approach to realize a complete automatic calibration scheme for 6DOF external parameters is proposed. The proposed approach does not use the mobile agent, but instead, uses only the environmental map information to cope with situations in which the mobile agent cannot be used. There has been no previous study which attempted to solve this problem of a myriad of local minima solutions that would appear because it is impossible to achieve global estimation of complete 6DOF camera parameters without strong constraints. In this study, 3D texture OctoMap information (e.g., the results of the RGB-D SLAM described in chapter 2) is used to generate virtual 2D images from arbitrary viewpoints. The 6DOF camera pose can be calibrated by matching the virtual images generated at every viewpoint with real images from the camera sensor network. To this end, 3D geometric line parameters for the entire environment are learned from the 3D texture OctoMap information. Then, a novel image descriptor based on Quantized Line parameters in the Hough space (QLH) is proposed in order to perform a particle filter-based matching process between line features extracted from both a distributed camera image and the learned 3D geometric line parameters. The experimental results show that the proposed calibration scheme using the QLH descriptor matching based on a particle filter can calibrate complete 6DOF external camera parameters successfully.

Finally, chapter 5 describes several applications to improve navigation functions (localization and safe path generation) for a mobile robot using both a camera sensor network calibrated using the methods proposed in chapter 3 or 4, and the 3D map information built by the RGB-D SLAM. In order to take full advantage of the information combination (i.e., the static map information and the dynamic information from the camera sensor networks), a novel concept of generating a global bird's-eye view image (elevated view of the entire environment from above) is proposed in this study. By detecting moving objects using background subtraction from the global bird's-eye view images, the performance of essential navigation functions (localization and path planning) can be significantly improved. In conclusion, mobile robots are expected to be operated safely and efficiently in the future by utilizing information from such intelligent spaces with the 3D map information.