別紙４－１（課程博士（英文））

Date of Submission (month day, year) : July 7th, 2021

Department of Computer Science and Information Engineering	Student ID Number	D189302
Applicant's name	Yubao Liu	

Supervisors	Jun Miura Shigeru Kuriyama

Abstract (Doctor)

Title of Thesis	Real-time Visual Simultaneous Localization and Mapping under Dynamic Environments

Approx. 800 words

Visual simultaneous localization and mapping (SLAM) is a fundamental technology in robotics, augmented reality, computer vision, and self-driving. Visual SLAM mainly uses an onboard camera to perform camera ego-motion estimation and reconstruct the unstructured environment. There are two basic requirements for visual SLAM: robust tracking and real-time performance. Many visual SLAM algorithms use the rigid/static world assumption, limiting a wide deployment in the real world, such as populated scenes, as this strong assumption often results in poor tracking in dynamic environments.

Many studies have tried using semantic information to help visual SLAM reduce the influence of dynamic objects. The challenge is to trade off the tracking with semantic information and real-time performance. For example, Mask R-CNN can achieve good segmentation results and help improve tracking accuracy. However, it usually consumes much time, around 200ms. The efficiency of tracking is significantly limited by waiting for the segmentation results. We call such a model a *blocked model*. To achieve robust tracking while keeping the real-time performance, we proposed a *non-blocked model*, in which the tracking process is no longer blocked by waiting for the semantic results.

This thesis proposes four visual SLAM methods, which we explored better solutions to the problems mentioned above. The first two methods are developed based on the blocked model, while the others on the non-blocked model.

First, we present RTS-vSLAM, which achieves robust tracking by detecting and removing outliers on the dynamic objects using PSPNet and SegNet, and builds a static semantic map by excluding dynamic objects. However, the efficiency of the performance is limited due to the blocked model.

Second, to deal with dynamic non-pre-defined objects, in KMOP-vSLAM, we try to use k-means to segment all clusters or objects. We use OpenPose to judge which clusters belong to persons. However, k-means tend to over-segment, and the segmentation accuracy is worse than semantic segmentation methods. Time consumption is large due to the blocked model, too.

Third, we present RDS-SLAM, a real-time visual SLAM for dynamic environments, developed based on semantic segmentation and the non-blocked model. RDS-SLAM runs in real-time by evaluating the tracking and the segmentation thread in parallel. According to the semantic segmentation results, the motion information of features, represented as moving probability, is updated in the global map. Then we classify the features in the map into dynamic, static, and unknown using the moving probability. We use as many static features as possible in the data association and the bundle adjustment process to obtain robust tracking. We tested two semantic segmentation methods, SegNet and Mask R-CNN, using the TUM dynamic sequences. RDS-SLAM can run at 30 Hz with SegNet, while only at 15 Hz with Mask R-CNN to trade off the tracking robustness and the real-time performance.

Fourth, we present RDMO-SLAM, an extension of RDS-SLAM, which can run the Mask R-CNN version at 30 HZ with the help of dense optical flow. Since optical flow estimation is faster than Mask R-CNN segmentation, to get more segmented frames, we predict the semantic labels using dense optical flow and the segmented frames of Mask R-CNN. To cope with unknown dynamic objects, we also estimate the velocity of landmarks and then use it as another constraint to lower the influence of dynamic objects. These improvements make RDMO-SLAM with Mask R-CNN run at 30 Hz while keeping the tracking performance.