

CMPUT 631: Project Proposal

Performance Enhancement of RTAB-map Utilizing ORB-SLAM2 Depth Information

Islam Ali
iaali@ualberta.ca

Zifei Jiang
zifei.jiang@ualberta.ca

Abstract

1 Problem Definition

2 Background

SLAM (Simultaneous Localization and Mapping) is the process of generating a model of the environment (a map) and localizing the camera inside it [1]. By localization we refer to the estimation of the 6DoF of the moving camera (orientation and position). Many sensors are used in SLAM such as LiDARs, Cameras, and Radars [2]. In this work, we focus on the usage of vision-based SLAM and possibly its integration with Laser-based SLAM.

The conventional pipeline of SLAM consists of two major stages, the first one is responsible for sensor abstraction and processing, which includes features extraction and tracking as well as any long-term data association steps such as bundle adjustment or loop closure [2][1]. The second step is to use the associated data to estimate a map or a model of the environment and, at the same time, estimate the 6DoFs (complete pose) of the camera inside the generated map reliably.

3 Literature Review

3.1 Visual SLAM

A wide spectrum of research has been conducted into the SLAM problem with the objective of enhancing its performance, allow for long-term operation, and enable portable and limited resources applications to utilize it in real-time. SLAM algorithms can be categorized to either feature-based method and direct methods [3]. Feature-based methods depend on extracting features and tracking them through different captured images. The estimation of the camera motion and the map is then determined using either filter-based methods (such as EKF in MonoSLAM) or using bundle adjustment optimization (such as BA used in PTAM). It was shown that BA method can provide better performance due to its ability to use more feature key points when compared to the complex EKF-based solution [4].

On the contrary, direct feature-less methods are usually used on images directly without any abstraction [3] in order to achieve real-time operation and eliminating the accumulated

error usually occurring in the aforementioned feature-based methods. Direct methods differ, mainly, in the density of the map and can be categorized based on that to either dense, semi-dense, or sparse methods [3]. Methods such as DTAM and LSD-SLAM are examples of the trials done to achieve direct SLAM by means of dense and semi-dense maps while other methods such as SVO and DVO are more advanced steps towards fully direct SLAM algorithms utilizing sparse maps [1][3].

3.2 RTAB-Map

3.3 ORB-SLAM 1 & 2

4 Research Methodology and Procedure

5 Performance Evaluation

The three corner stones of performance in computer science are accuracy, efficiency, and storage requirements. In this work, we focus on addressing *accuracy* aspects of the system which are defined by the relation between the generated trajectory and ground truth data.

The *Absolute Trajectory Error (ATE)* was defined in [5] as a performance metric to stand on the accuracy of a SLAM system vs. a ground truth. The ATE is defined by:

$$ATE(t_i) = \|(x_{t_i}^*, y_{t_i}^*) - (x_{t_i}, y_{t_i})\| \quad (1)$$

where $(x_{t_i}^*, y_{t_i}^*)$ are the ground truth coordinates at t_i , and (x_{t_i}, y_{t_i}) are the coordinates generated by the SLAM algorithm under test at the same epoch. For a more detailed representation of this performance metric, statistical functions are applied to this metric such as Root Mean Square (RMS), mean, median, variance, and standard deviation. Additionally, the maximum and minimum *ATE* are reported to define the system limits while operating.

6 Significance of Proposed Research

References

- [1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [2] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, “Simultaneous localization and mapping: A survey of current trends in autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, 2017.
- [3] T. Taketomi, H. Uchiyama, and S. Ikeda, “Visual slam algorithms: a survey from 2010 to 2016,” *IPSS Transactions on Computer Vision and Applications*, vol. 9, no. 1, p. 16, 2017.

- [4] H. Strasdat, J. M. Montiel, and A. J. Davison, “Visual slam: why filter?,” *Image and Vision Computing*, vol. 30, no. 2, pp. 65–77, 2012.
- [5] M. Filipenko and I. Afanasyev, “Comparison of various slam systems for mobile robot in an indoor environment,” in *2018 International Conference on Intelligent Systems (IS)*, pp. 400–407, Sep. 2018.