TECHNISCHE UNIVERSITÄT WIEN

MASTER'S THESIS PROPOSAL

Visual Simultaneous Localization And Mapping Evaluation on a Mobile Robot Platform

Author:
Martin Haar, BSc

Registration Number:
01625753

Supervisors:
Ao. Univ. Prof. Dipl.-Ing. Dr.techn. Markus Vincze
Projektass. Jean-Baptiste Nicolas Weibel, MSc

April 4, 2022

1 Motivation & Problem Statement

Nowadays robots are used extensively in all sort of domains, especially in an industrial context. Progress made in recent years in sensing and human-robot interaction brought us closer to the deployment of robots in our everyday lives. One example of such a robot is the Human Support Robot (HSR) developed by Toyota [1]. These robots need to be capable of locating themselves on a map and detect obstacles in order to safely move in their environment. Furthermore they need to collect detailed 3D and semantic information of the objects in their environment to understand and execute given tasks.

Data collection and object modelling are key parts of the development and evaluation of any vision-based grasping pipeline. It is very common to first estimate the object pose to infer how to grasp it, which requires an object model, and any vision-based grasping pipeline needs to be evaluated on the specific object of interest.

These tasks are however time-consuming for a human as they require some manual annotation and need to be repeated for every application. In this work, we want to investigate the feasibility of an automated pipeline for data collection with automated annotation. A core component for such a pipeline is the ability to precisely track the camera pose across time. This general problem is known as simultaneous localization and mapping, or SLAM. Knowing the camera pose is a key component to object modelling, and can greatly reduce the annotation need, as scenes can be annotated once and the semantic information can be reprojected in every frame.

2 Aim of the Work

The aim of this work is to benchmark several SLAM algorithms and their accuracy when used for automated object reconstruction on the Toyota HSR.

The HSR has a lot of different sensors: a 2D LiDAR, a RGB-D camera, a set of stereo cameras and an IMU. It can therefore be used to test several SLAM algorithms having different sensor requirements. The focus of the evaluation will be on high accuracy object reconstruction.

A setup with a static robot arm has already been developed, where the camera pose can be determined using the kinematics. Every joint is equipped with a high quality encoder that provides its exact rotation. Through this setup a high accuracy object reconstruction can be performed without the need for localization. This setup will provide groundtruth data for the estimated camera poses of the SLAM algorithms. In a second set of experiments the SLAM is deployed on the HSR and the object models obtained by the robot are compared to reference object models to evaluate the quality of the SLAM algorithms on the mobile robot platform.

Most SLAM algorithms are able to combine visual-, IMU- and LiDAR data to increase localization accuracy. Therefore each algorithm (if possible) should be used in two different setups:

- Visual only SLAM
- Full sensor SLAM (e.g. LiDAR, IMU)

Using the camera poses obtained with the SLAM algorithms, every depth image from the RGB-D sensor will first be filtered to remove the supporting plane and only keep the objects of interest (using Random Sample Consensus [2]) and combined into a 3D model using a Truncated Signed Distance Function grid.

At the end of this thesis it should be determined if the object reconstruction accuracy with SLAM methods are good enough to be used in further projects or describe the constraints on the scene necessary for their applicability.

3 Methodological Approach

1. **Literature review:**
   An extensive literature survey is needed to get a good understanding of the topic and also show the actual research progress. Furthermore the literature research yields the necessary information to select appropriate SLAM algorithms fit for the task.

2. **Setup ROS environment and object reconstruction pipeline:**
   Since the HSR operates ROS an according development environment must be set up. With a running ROS setup the development of the object reconstruction pipeline can be conducted. Every SLAM algorithm to be tested must be integrated in said pipeline and can if necessary be simulated using Gazebo.

3. **Data collection on actual hardware:**
   After a successful software integration the feasibility of the pipeline must be tested on the actual hardware. If simulations have been conducted there is always the so called *sim to real gap* which can only be assessed by comparing the simulation results to the real world outcome.

   There are two hardware setups which are used in the course of this thesis:

   a) Reference robot-arm setup

   b) Toyota HSR

   As mentioned before the reference robot-arm is used to evaluate the quality of the estimated trajectories using the chosen SLAM algorithms. The actual object reconstruction accuracy is tested using the HSR and reference object models.

4. **Evaluation of results:**
   The final task is to find good metrics to assess the quality of each SLAM algorithm deployed on the HSR and conduct a comparison to 2D LiDAR SLAM and the static

robot arm. To assess the object models the first approach will be a point to point comparison between the groundtruth and the SLAM models. The camera poses can be compared in terms of translation- and rotation error.

4 State-of-the-Art

The first SLAM technology was introduced 1986 by Smith et al [3], since then a multitude of different SLAM algorithms have been developed. These algorithms utilize all sort of available sensors like, LiDAR, RaDAR, odometry, cameras, IMUs and so on. Since the goal of this thesis is to benchmark the performance of SLAM methods for object reconstruction, visual SLAM methods are of particular interest.

The HSR is equipped with stereo cameras as well as a RGB-D camera therefore a wide range of visual SLAM implementations can be used. A lot of the recent SLAM implementations are summarized in Barros et al [4], Zhang et al [5] and Servieres et al [6].

One of the most common SLAM algorithm is ORB-SLAM2 by Mur-Artal and Tardos [7]. There is also a very recent update called ORB-SLAM3 by Campos et al [8]. There are a lot of other recent SLAM algorithms which are probable contenders to be used in the course of this thesis ([9], [10], [11])

One necessity for the chosen algorithms is the availability of an open source implementation for ROS. The implementation of every algorithm for the HSR would go beyond the scope of this thesis. There is already a comprehensible list of open source SLAM implementation on Github [12].

5 Relevance to the Curricula of Computer Engineering

The subject area *Machine Vision and Cognitive Robotics* can be placed in an intersection of electrical engineering and informatics. In each robot setup the complex hardware platform has to be matched with according software and algorithms. Therefore this topic fits very well in the curriculum *computer engineering.*

Besides courses which are directly related to the topic of computer vision and mobile robotics there are also courses which provide the mathematical backgrounds for this thesis. Especially fundamentals in linear algebra, statistics and optimization are covered by several courses offered by the key area *Digital Signal Processing and Communication.*

- 376.054 Machine Vision and Cognitive Robotics
- 182.763 Stochastic Foundations of Cyber-Physical Systems
- 191.119 Autonomous Racing Cars
- 389.166 Signal Processing 1

- 389.170 Signal Processing 2
- 389.163 Digital Communications 1
- 389.119 Parameter Estimation Methods
- 389.040 Signal Detection

References

[1] T. Yamamoto, T. Nishino, H. Kajima, M. Ohta, and K. Ikeda, "Human support robot (hsr)," in *ACM SIGGRAPH 2018 emerging technologies*, pp. 1–2, 2018.

[2] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[3] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *The International Journal of Robotics Research*, vol. 5, no. 4, pp. 56–68, 1986.

[4] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A comprehensive survey of visual slam algorithms," *Robotics*, vol. 11, no. 1, 2022.

[5] S. Zhang, L. Zheng, and W. Tao, "Survey and evaluation of rgb-d slam," *IEEE Access*, vol. 9, pp. 21367–21387, 2021.

[6] M. Servieres, V. Renaudin, A. Dupuis, N. Antigny, and S. Potirakis, "Visual and visual-inertial slam: State of the art, classification, and experimental benchmarking," *Hindawi Journal of Sensors*, 2021.

[7] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[8] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[9] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: an open-source library for real-time metric-semantic localization and mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1689–1696, 2020.

[10] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[11] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4666–4672, 2020.

[12] IntRoLab, "List of open source slam projects," July 2021.