Chapter 1

Introduction

1.1 The SLAM Problem

Simultaneous localization and mapping (SLAM) is one of the fundamental problems in robotics. The problem is for a mobile robot, while moving around in some unknown environment, to use its sensors to construct a map of that unknown environment. SLAM is difficult mainly because the robot cannot determine its position precisely. It might have access to some positioning sensors such as wheel encoders, GPS, or a compass, but still, some kind of environmental feedback will always be necessary to help correct the error that inevitably exists in these sensor readings. The main sensors used in SLAM for this kind of feedback are laser range finders and video cameras.

The conventional formulation of SLAM problem is to seek a recursive estimate of

$$p(\boldsymbol{s}_t, \Theta \mid \boldsymbol{z}_{0:t}, \boldsymbol{u}_{1:t}), \tag{1.1}$$

the posterior over robot pose at time t, $\boldsymbol{s}_t$ and a non-dynamic environment Θ, given a set of robot actions $\boldsymbol{u}_{1:t}$ and a set of sensor observations $\boldsymbol{z}_{0:t}$. The filter equation that recursively estimates (1.1) is given by

$$\begin{aligned} p(\boldsymbol{s}_t, \Theta \mid \boldsymbol{z}_{0:t}, \boldsymbol{u}_{1:t}) &= \eta\, p(\boldsymbol{z}_t \mid \boldsymbol{s}_t, \Theta) \\ &\quad \int p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \boldsymbol{u}_t)\, p(\boldsymbol{s}_{t-1}, \Theta \mid \boldsymbol{u}_{1:t-1}, \boldsymbol{z}_{0:t-1})\, d\boldsymbol{s}_{t-1}, \end{aligned}$$

where η is a normalization constant [1]. Here, the conditional probabilities $p(\boldsymbol{z}_t \mid \boldsymbol{s}_t, \Theta)$ and $p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \boldsymbol{u}_t)$ are generally referred to as *measurement model* and robot *motion model*, respectively. Therefore, once we define the two models appropriately, we can, in principle, infer the posterior (1.1) in a recursive manner.

SLAM is an important problem that encompasses a wide range of practical applications across academic, commercial and human-life sectors. Thanks to the automation of map generation, SLAM will save huge amount of time and human resources when applied to mapping of places such as mine fields and coral reefs. SLAM for large-scale environments will make large contributions to geoscientific projects such as marine explorations and planetary terrain modelings. Combined with the latest developments in mobile and humanoid robots, autonomous robots with SLAM capabilities are going to solve dangerous missions such as landmine removal projects and rescue robot projects without risking human lives.

1.1.1 SLAM Setup

Many different forms of setups may arise in SLAM depending on the environments that we are mapping and system's requirements to solve individual problems. The difficulty of a particular SLAM problem can be inferred from its setup. We describe the major categories of SLAM setups in the following subsections.

Scale of Environments

The scale of the environments that we want to map varies from that of a small laboratory room to that of a vast expanse of terrain. Generally, it is more difficult to cope with larger environment due to the limits of sensor's range capabilities and increasing number of landmarks that we have to store and manage. We, for example, categorize them as follows:

- indoor environments (single rooms, connected rooms, etc.)
- large-scale indoor environments (office buildings, etc.)
- outdoor environments (car parks, mine fields, disaster sites, etc.)
- large-scale outdoor environments (terrains, coral reefs, cities, etc.)

We are particularly interested in mapping large-scale environments. The number of successful large-scale SLAM demonstrations is still not very large today. While SLAM in indoor environments has been solved by many research groups, due to their relatively small scale, indoor environments provide us with good testbeds for measuring the performance of our proposed SLAM system. In this research, we first conduct a large-scale outdoor SLAM experiment in simulation to measure the scalability of our proposed system, and finally, we conduct an indoor SLAM experiment in a real laboratory environment to measure the robustness of our system under noise.

Degrees of Freedom

The degrees of freedom of target environments and mobile robots are important since they are directly related to the complexity of the SLAM problem being modeled. Early SLAM systems have been dealing with 2D environments with a mobile robot constrained to a plane. In these cases, a mobile robot has three degrees of freedom in space since its position s is fully determined as $s = [s_x, s_y, s_\psi]^T$, where s_x and s_y specify the 2D Euclidean coordinates of the robot and s_ψ is yaw of the robot.

On the other hand, in modern SLAM problems, we often seek to map 3D environments using a robot with six degrees of freedom. Robot's position s is represented as $s = [s_x, s_y, s_z, s_\phi, s_\theta, s_\psi]^T$, where s_x, s_y and s_z specify the 3D Euclidean coordinates of the robot and s_ϕ, s_θ and s_ψ represent pitch, roll and yaw of the robot, respectively.

In our case, we construct 3D sparse landmark maps of environments. For robot motion models, we consider six degrees of freedom with reasonable constraints on s_z, s_ϕ and

s_θ based on a priori knowledge that the robot is loosely constrained to the flat ground of environments.

Time Constraints

Since SLAM is one of the most fundamental parts of autonomous mobile robot navigation technology, we generally require that the system should run in real time (in other words, *online*). It is until quite recently that we see successful online demonstrations of SLAM in indoor environments [2–4]. However, none of large-scale outdoor online SLAM demonstration using lightweight equipments is yet reported to the best of our knowledge.

Our aim is to implement the prototype of a SLAM system capable of scaling to large-scale outdoor environments and running online.

1.1.2 Approaches to SLAM

There are many possible approaches to tackling the SLAM problem. Selecting and modeling sensors is one of the most important steps. In the following subsections, we first describe the sensors that are typically used in SLAM systems. We then describe the popular choices of representation of reconstructed maps, and finally introduce the major techniques to optimize the estimated robot poses and sensor observations.

Type of Sensors

In SLAM, we need some kind of sensors to incrementally register new objects into the map of unknown environments. Laser range finders are one of the most successful sensors. Lasers are, on the one hand, extremely accurate, but on the other hand, they are large, slow and expensive. Following successful SLAM demonstrations in both indoor and outdoor environments using laser range finders [1, 5], research focus has been shifting towards vision-based sensors that are small, fast and inexpensive. Vision-based sensors are generally more subject to noise than laser range finders. However, there has been various kinds of approaches to achieve robustness using vision-based sensors. We list major vision-based sensors as follows:

- a monocular camera,
- a stereo camera system,
- a trinocular stereo camera system.

Vision-based SLAM has been traditionally adopting stereo camera bench as the main sensor. Then, the recent developments of multiple-view geometry in the computer vision community encouraged researchers to use trinocular systems in the hope of more accurate sensing of 3D objects in environments (see, for examples [6,7]). More recently, monocular SLAM using a single camera as the main sensor is actively pursued [2, 8]. Monocular camera sensors have convenient since they are quite lightweight and they do not require careful calibration of extrinsic parameters as stereo cameras do. On the other hand, monocular camera sensors requires some sophisticated fusion of original sensor readings obtained from multiple images captured at different time steps in order to determine the 3D positions of landmarks. In our research, we adopt a trinocular stereo system and achieve robust sensing of 3D point features in environments.

Besides sensors for detecting objects in environments, we typically use auxiliary sensors that provide information on robot's motion (we call it *odometry*). Motion model of the robot is modeled based on the characteristics of these sensors. The major sensors of these kinds are:

- wheel encoders,
- gyro sensors,
- Global Positioning System (GPS) devices,
- vision-based odometry.

Using wheel encoders is a very popular way of obtaining odomety when cars or rovers are used as a mobile agent. However, wheel encoders are noisy sensors due to slippage. Gyro-sensors are useful to determine the orientation components of robot positions. GPS devices provide good estimates of robot positions in a global scale. However, it is not available in indoor environments and subject to occlusions even in outdoor scenes. GPS devices are mainly used to obtain ground truth position data of a mobile robot in outdoor SLAM experiments. Vision-based odometry (or visual odometry) is a technique to estimate camera motions from a sequence of images captured. In our research, we directly simulate odometry data assuming the robot is moving through a mostly smooth surface.

Map Representation

There are two major approaches to representing the map of environments: *occupancy grid mapping* and *landmark-based mapping*. In occupancy grid mapping, the entire space is divided into small grid cells each of which are assigned a binary random variable indicating whether the space is occupied or not. The set of grid cells represent the map of environments, which generally has a huge degrees of freedom even for small indoor environments. The goal of occupancy grid mapping is to estimate the posterior over maps given a robot path and all previous sensor observations. Although there are some techniques such as *oct-trees* to reduce the size of maps, occupancy grid approach is not very suitable for our case. We want to map large-scale environments that make the griding strategy inappropriate.

Instead, we take landmark-based mapping approach to represent large-scale environments more efficiently. In this approach, we register geometric objects in environments as sparse landmarks. Landmarks could be any geometric objects, but we use 3D point landmarks in our case since 3D points are more simple and straightforward than other complex objects to derive from trinocular stereo images.

1.2 Problem Statement in This Research

We are interested in SLAM for constructing metric maps of large-scale environments such as office buildings and mine fields. Laser range finders have been particularly successful sensors for these kinds of environments — see, for example, [1] — because lasers are extremely accurate. On the other hand, they are also heavy, expensive, and slow. In our work, we focus on the use of cameras as sensors due to their high speed, small size, and low cost.

In this research, we explore the use of the Shi-Tomasi point feature [9] with trinocular stereo vision as the visual sensor model for the FastSLAM algorithm (we call the algorithm *ST-SLAM*). On the one hand, Shi-Tomasi feature locations are sensitive to noise, but on the other hand, they are quite lightweight in comparison with SIFT (the scale invariant feature transform). We find that with the help of the epipolar geometry of trinocular stereo images, it is possible to reliably find corresponding Shi-Tomasi points in trinocular image sets and use them to reconstruct 3D point landmarks. To empirically measure the performance of ST-SLAM, we first conduct large-scale SLAM experiment in a simulated outdoor environment. We then apply our techniques to a trinocular image sequence collected in an indoor environment with carefully measured ground truth. We evaluate our method in terms of sensor observation likelihood and position estimation error. We conclude whether or not ST-SLAM is feasible as a lightweight alternative to RBPF methods using sensors based on rich, computationally expensive feature descriptors.

1.3 Approaches in This Research

The summary of approaches to solve our stated problem is shown as follows. The concrete descriptions of each step are written in the following chapters.

- Implement the RBPF (FastSLAM) as fundamental estimation algorithm.
- Use lightweight Shi-Tomasi point features.
- Implement a vision-based sensor based on trinocular stereo cameras.
- Conduct SLAM experiments in a simulated outdoor environment.
- Conduct SLAM experiments in a real indoor environment.
- Evaluate the performance of our proposed ST-SLAM algorithm using various measures such as log likelihood, motion error, and visual inspection of estimated robot paths.

1.4 Organization of the Thesis

This thesis is organized into six chapters:

Chapter 1 Introduction to Simultaneous Localization and Mapping (SLAM) problem, problem setups, possible approaches, and problem statement and approaches in this research.

Chapter 2 Background of vision-based SLAM.

Chapter 3 Detailed explanation of our system: system overview, algorithms, and the sensor model.

Chapter 4 Description of experimental design: target environment, experimental configurations, evaluation methods.

Chapter 5 Presentation of the results of the experiments described in Chapter 4.

Chapter 6 Discussion on the performance of our algorithm (ST-SLAM) to the SLAM problem. Limitations and future work. Conclusion.