Visual Odometry

Christos Kokas

May 2022

Contents

1 Introduction

In robotics and computer vision, visual odometry is the process of determining the position and orientation of a robot by analyzing images from the robot's camera. It has been used in a wide variety of robotic applications, such as on the Mars Exploration Rovers [1].

The accurate localization of the robot poses a great challenge for many applications, even more in GPS-Denied environments (e.g. indoors) where GPS systems, such as RTK GPS, have difficulty producing an accurate estimate of the position of the robot. Visual Odometry has gained popularity because it can produce accurate estimates of the position of the camera even in GPS-Denied environments and at the same time be cheaper than other alternatives (e.g. RTK GPS, LiDARS). Combinations of Visual Odometry and other techniques or products, such as Visual Odometry combined with an IMU (also know as Visual Inertial Odometry) or Visual Odometry combined with Deep Learning, can also yield satisfying results.

2 Literature Review

In this section papers using Visual Odometry will be presented each using a different method of producing an accurate estimate of the camera.

2.1 ORB_SLAM2

2.1.1 Abstract

We present ORB-SLAM2 a complete SLAM system for monocular, stereo and RGB-D cameras, including map reuse, loop closing and relocalization capabilities. The system works in real-time on standard CPUs in a wide variety of environments from small hand-held indoors sequences, to drones flying in industrial environments and cars driving around a city. Our back-end based on bundle adjustment with monocular and stereo observations allows for accurate trajectory estimation with metric scale. Our system includes a lightweight localization mode that leverages visual odometry tracks for unmapped regions and matches to map points that allow for zero-drift localization. The evaluation on 29 popular public sequences shows that our method achieves state-of-the-art accuracy, being in most cases the most accurate SLAM solution. We publish the source code, not only for the benefit of the SLAM community, but with the aim of being an out-of-the-box SLAM solution for researchers in other fields [2].

2.1.2 System Overview

ORB-SLAM2 for stereo and RGB-D camera has three main parallel threads.

- Find Feature Matches to the local map with every frame to localize the camera, and apply motion-only Bundle Adjustment (BA) to minimize the reprojection error. (Tracking)
  - Bundle Adjustment is simultaneous refining of the 3D coordinates describing the scene geometry, the parameters of the relative motion, and the optical characteristics of the camera(s) employed to acquire the images, given a set of images depicting a number of 3D points from different viewpoints. It amounts to an optimization problem on the 3D structure and viewing parameters (i.e., camera pose and possibly intrinsic calibration and radial distortion), to obtain a reconstruction which is optimal under certain assumptions regarding the noise pertaining to the observed image features.
  - Reprojection Error is a geometric error corresponding to the image distance between a projected point and a measured one.
- Perform local BA to manage and optimize the local map. A local map is a simplified representation of the immediate environment around the robot. (Local Mapping)
- Detect large loops and correct the accumulated drift by performing a pose-graph optimization. A pose graph contains nodes connected by edges that represents the pose of the robot. (Loop Closure)

References

[1] Mark Maimone, Yang Cheng, and Larry Matthies. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics, Special Issue on Space Robotics*, 24:2007, 2007.

[2] Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.