PLVI-SLAM: Visual–Inertial Monocular SLAM Using Point and Line Features

Xinyu Wei* Zengming Tang† Huiyan Wu‡ Jun Huang§

Shanghai Advanced Research Institute, Chinese Academy of Sciences
Shanghai, China

ABSTRACT

This paper presents PLVI-SLAM, a tightly-coupled monocular visual–inertial SLAM system using both point and line features. We fuse the multi-features from visual sensors and inertial measurements from IMU to improve the robustness and accuracy of monocular SLAM system. Compared with point features, line features are abundant in man-made environments and they can complement points well. We use parallel processing framework to speed up the multi-feature process. To efficiently fuse preintegrated IMU measurements and multi-feature observations, a joint optimization algorithm is proposed. A loop closing module is utilized in the system to implement online relocalization.

Index Terms: Visual–Inertial SLAM—Point and Line Features—Sensor Fusion;

1 INTRODUCTION

In recent years, monocular Simultaneous Localization And Mapping (SLAM) system with low-cost inertial measurement units (IMUs) attracted much attention. Many VI-SLAM [6, 8] and VIO [4, 7, 11] systems are proposed. These systems used the acceleration and angular velocity information from IMUs to improve the robustness of monocular SLAM system. The combination of visual sensors and IMUs make the SLAM system being capable of recovering the metric scale and more robust in textureless scenes.

According to the methods of data fusing, monocular visual–inertial SLAM system can be classified into two main categories: loosely-coupled and tightly-coupled algorithms. Loosely-coupled algorithms using data from IMUs and cameras to estimate the pose of camera or robot separately. And then fuse the two pose estimates to obtain final result. Tightly-coupled algorithms combining the two types of data to estimate the camera pose. Compared to loosely-coupled algorithm, tightly-coupled algorithms can usually get more accurate results. Tightly-coupled fusion is usually done by the the Extended Kalman filter (EKF) and optimization-based algorithms. But due to the EKF approaches discard the historical measurements, the graph optimization methods generally have more accuracy results and can handle large environment.

In visual SLAM, point features are commonly used to represent the image information. But in low-textured environments such as man-made scenes, it is difficult to find a reliable set of point features. In comparison, line features are usually abundant in these environments. Combination of points and line segments can improve the visual SLAM in such challenging scenes. In addition, line features can provide more geometrical structure information of the environment and is suitable for some applications such as Augmented Reality (AR).

To obtain geometrical structure information of the environment and improve the robustness of visual SLAM system, we proposed the PLVI-SLAM, a tightly-coupled monocular visual–inertial SLAM system using both point and line features. We utilized the graph optimization methods to fusing the multi-features and inertial measurements, and obtain the accuracy camera poses.

2 SYSTEM OVERVIEW

The PLVI-SLAM is built upon our previous work [10]. The system has three threads: tracking, map managing and loop closing (Fig. 1). In the tracking module, point and line features are extracted and matched with 3D features in map. The inertial measurements between two frames are combined into a single relative motion constraint. Then a graph optimization method is used to optimized the current camera pose. The map managing module maintain the system map which includes the keyframes, map points and map lines. In this thread, 3D features in map, keyframes, groscope bias, acceleration bias and gravity vector are jointly optimized. The loop closing thread searches loop and corrects it. We process the features and inertial measurements in parallel threads. In Fig. 1, the parallel processing modules are denoted with blue doted boxes.

2.1 Traking

For each new frame, ORB [9] and EDLine [1] features are detected. And IMU preintegration is conducted as [2]. In our system, the rotation of camera is represented as a manifold structure of the rotation group $SO(3)$. Then the initial camera pose is estimated with the preintegrated IMU measurements. When the initial pose is obtained. The 3D features in local map are projected in current frame and the 2D-3D feature matches is calculated using a fast matching algorithm [10]. After getting the initial pose, feature matches and preintegrated IMU measurements, a pose graph optimization is conducted to refine the initial pose . The optimization procedure includes the optimization on reference keyframe and optimization in local map. At last, the keyframe is selected based on the tracking quality.

2.2 Map Managing

When a new keyframe is selected. Map managing module add the keyframe in map. In this thread, old map features and keyframes are culled and new map features are created. All variables including the poses of keyframes, map features, gyroscope bias, acceleration bias, metric scale and gravity vector are jointly optimized. And this graph-based optimization is conducted in the local window as [6]. In addition, for the horizontal plane can be determinated by the gravity vectors, which means the absolute roll and pitch angles can be observed, we use the 4-DOF pose graph optimization method as [7].

In map managing. The maintaining of map features is difficult. We use multi criteria including life cycle, observation number and tracking quality to cull features, and multiple optimizations to optimized features and keyframes.

*e-mail: weixy@sari.ac.cn
†e-mail:tangzengming2019@sari.ac.cn
‡e-mail:wuhuiyan2019@sari.ac.cn
§e-mail:huangj@sari.ac.cn

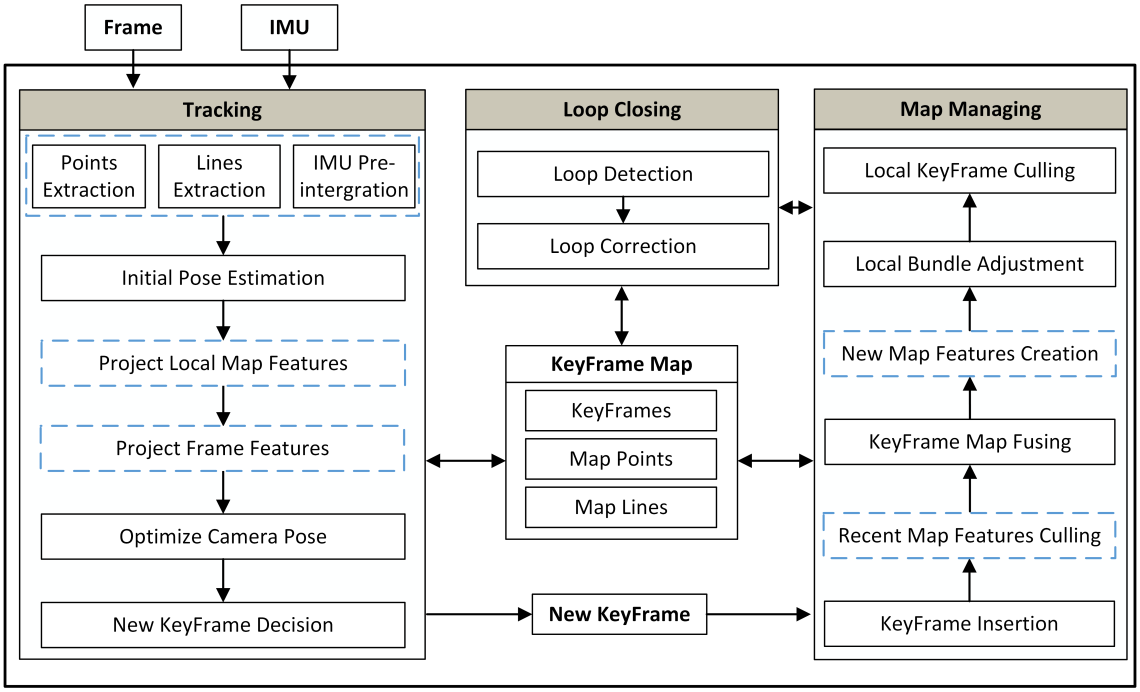

Figure 1: Overview of PLVISLAM system

2.3 Loop Closing

To reduce drifts, we used DBoW2 [3] to identify the places that have been visited. In the loop closing module, a loop is detected as [5]. After loop detection, a pose-graph optimization on 4 degrees of freedom is performed to refine the map.

References

[1] C. Akinlar and C. Topal. Edlines: A real-time line segment detector with a false detection control. *Pattern Recognition Letters*, 32(13):1633–1642, 2011.

[2] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. On-manifold preintegration for real-time visual–inertial odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2016.

[3] D. Gálvez-López and J. D. Tardos. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 28(5):1188–1197, 2012.

[4] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015.

[5] 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*, 33(5):1255–1262, 2017.

[6] R. Mur-Artal and J. D. Tardós. Visual-inertial monocular slam with map reuse. *IEEE Robotics and Automation Letters*, 2(2):796–803, 2017.

[7] T. Qin, P. Li, and S. Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.

[8] M. Quan, S. Piao, M. Tan, and S.-S. Huang. Accurate monocular visual-inertial slam using a map-assisted ekf approach. *IEEE Access*, 7:34289–34300, 2019.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski. Orb: An efficient alternative to sift or surf. In *ICCV*, vol. 11, p. 2. Citeseer, 2011.

[10] X. Wei, J. Huang, and X. Ma. Real-time monocular visual slam by combining points and lines. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 103–108. IEEE, 2019.

[11] F. Zheng, G. Tsai, Z. Zhang, S. Liu, C.-C. Chu, and H. Hu. Trifo-vio: Robust and efficient stereo visual inertial odometry using points and lines. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3686–3693. IEEE, 2018.