Recovery of lost map for monocular simultaneous localization and mapping

Kaiqi Chen[1*], Jialing Liu[1*], Mengping Gui[1], Luzhen Ma[1], Jianhua Zhang[1], Shengyong Chen[2]

[1] Institute of Computer Vision, Zhejiang University of Technology, Hangzhou 310023, China

[2] College of Computer Science, Tianjin University of Technology, Tianjin 300384, China

[*] Denoting the first two authors have the same contribution to this work

1. Introduction

The traditional SLAM system requires a camera to return to the position where tracking is lost to restart location and mapping, thereby greatly limiting the application of monocular SLAM scenarios. Thus, a new and highly adaptive recovery strategy is required. In a traditional keyframe-based SLAM system, relocalization is achieved by finding the keyframes (a reference camera pose and image) that match the current frame to estimate the camera pose. In practical application, relocalization is not a user-friendly solution because users prefer to walk into previous unseen areas where no associated keyframes exist. Moreover, finding matches between the two frames is difficult when the scene changes considerably. Tracking resumption can even be difficult when the camera arrives at the place from where it has been. The motion and map message during lost time cannot be recovered, which leads to an imperfect map. Thus, this study proposes a strategy of monocular vision combined with inertial measurement unit (IMU) to recover a lost map based on map fusion.

2. Methods

The ORB_SLAM [1] system incorporates three threads that run in parallel: tracking, local mapping, and loop closing. Our algorithm is on the basis of ORB_SLAM. When tracking fails in conditions with motion blur or occlusions, our system immediately restarts a new map. The first map created before tracking lost is saved, which will be used for map fusion. Then a new map is created by re-initializing. Compared with relocalization, re-initialization is more achievable, whereas the scale of the camera trajectory computed by ORB_SLAM is arbitrary. The inconsistent scale must be solved before fusing the two maps. The absolute scale of the map is calculated by combining monocular vision with IMU. The IMU sensor can recover the metric scale for monocular vision and is not affected by the limitation of vision. The position of the camera is calculated by IMU data before the initialization is completed. The pose transformation obtained by IMU during lost time can be used as a bridge to fuse the two maps when the new map is initialized.

Because the ORB_SLAM algorithm framework is widely used and has high precision and robustness, the ORB_SLAM framework is as our basic framework of the map fusion algorithm. The algorithm framework of this work is shown in Figure 1. The algorithm framework can be divided into three threads, namely tracking, local mapping and loop closing detection. The algorithm will not relocate after tracking the thread is lost, but re-initialize and create a second map. When tracking is lost, the created map before the loss is saved. And a new map is created after the scale is estimated by IMU. This scale estimation algorithm is added to the local construction thread to convert the current map scale to the real scale, and the scale of the two maps is unified. In addition, a new thread has been added for map fusion.

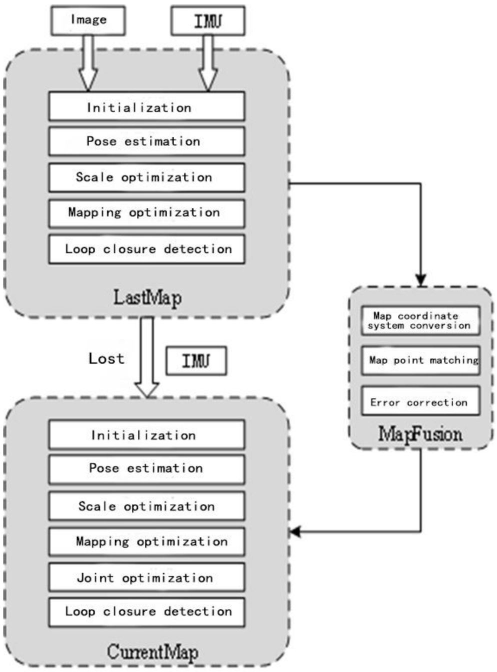

Fig.1 algorithm framework

3. Implementation Details

The process of map fusion consists of three steps: 1) Coordinate transformation. The coordinate system of the new map has changed after re-initialization. The transformation between the new map and the lost map can be computed by IMU during the lost time, which is applied to the lost map to move the lost map onto the new map. 2) Matching strategy. The data measured by IMU contains various errors unavoidably affects the result of the map fusion. Thus, the vision information is considered to eliminate errors. A jumping matching search strategy is proposed according to the covisibility relationship among keyframes to reduce matching time. We match keyframes selected from the two maps according to a certain condition instead of matching them individually. Thus, a considerable amount of matching time can be reduced. 3) Estimating the error between the two maps. After a set of matching map points is obtained, motion estimation between the matching points is solved by nonlinear optimization, which is applied to the two maps to reduce errors. Finally, the new map points in the matching map points are erased, and the two maps are merged. The relationship of the keyframes and the map points between the old and new maps is established, which are used to jointly optimize the pose of the camera and the position of the map points in subsequent tracking and mapping.

Scale estimate：

In the monocular SLAM system, since the depth information is unknown, the motion trajectory and the real scale of the map cannot be directly obtained, and the scale of the generated map is arbitrary. Therefore, there is a scale difference between the maps after each initialization, and the scale must first be unified before the map fusion. In this work, the real scale is estimated by IMU information according to [2].

Map coordinate system transformation：

In the process of creating a map by the monocular vision SLAM system, the camera coordinate system of the key frame of the first frame is often used as the world coordinate system. Therefore, when the SLAM system is reinitialized, the world coordinate system also changes. The IMU coordinate system, the world coordinate system and the camera coordinate system have a transformation relationship with each other, and the relationship between the coordinate systems is as shown in Fig. 2 .

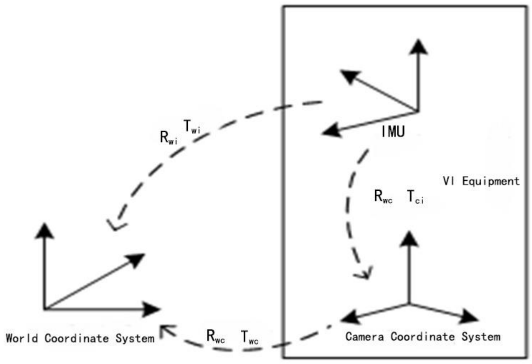

Fig.2 Transforming relationship between coordinate systems

When the visual tracking fails, the system re-initializes the created map to be different from the coordinate system of the map created before the loss. But in the world coordinate system, there is a correlation between the two maps. Map fusion can be performed by calculating the transformation, in the same world coordinate system, between the first frame keyframe of the map created before tracking failure and the keyframe of the first frame after the reinitialization. The calculation is mainly composed of two parts, as shown in Fig. 3. The first part is the transformation relationship between the first keyframe and the last frame of the map before the loss, which is obtained by visual odometry. The second part is the transformation relationship estimated by the IMU in the tracking failure part. By estimating the camera pose during the time period where the visual tracking fails, the IMU pose transformation can be used to transform two maps with different coordinate systems into the same coordinate system.

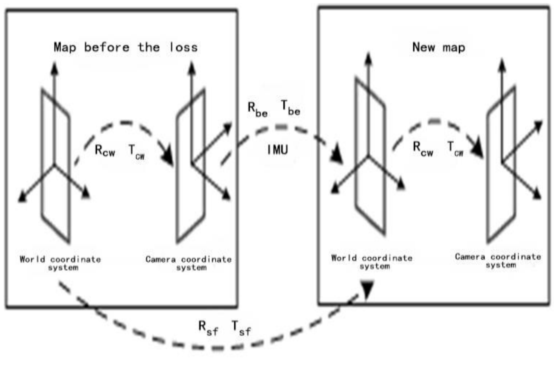

Fig.3 Map coordinate system transforming relationship

Error correction:

In order to solve the cumulative error caused by the measurement data of IMU, an error correction method is proposed to solve the motion estimation between two map points by nonlinear optimization. In order to eliminate the cumulative error, we need to find the matching keyframes in the two maps. First, we propose a jumping matching search strategy to reduce matching time. This strategy is based on the covisible relationship among keyframes. After finding the matching keyframes and the corresponding matching map points, the nonlinear optimization is used to solve the motion estimation between the two group of 3D points. The motion estimation results between the two sets of 3D points are obtained by iterative optimization, and then the old map is transformed accordingly to compensate for the error of the IMU measurement.

Map fusion:

The map before the loss is transformed according to the optimized transformation parameters, and the two maps before and after the loss are unified to the same coordinate system.

Reference:

1. Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163. [DOI: 10.1109/TRO.2015.2463671]

2. Mur-Artal R, Tardós J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803. [DOI: 10.1109/LRA.2017.2653359]