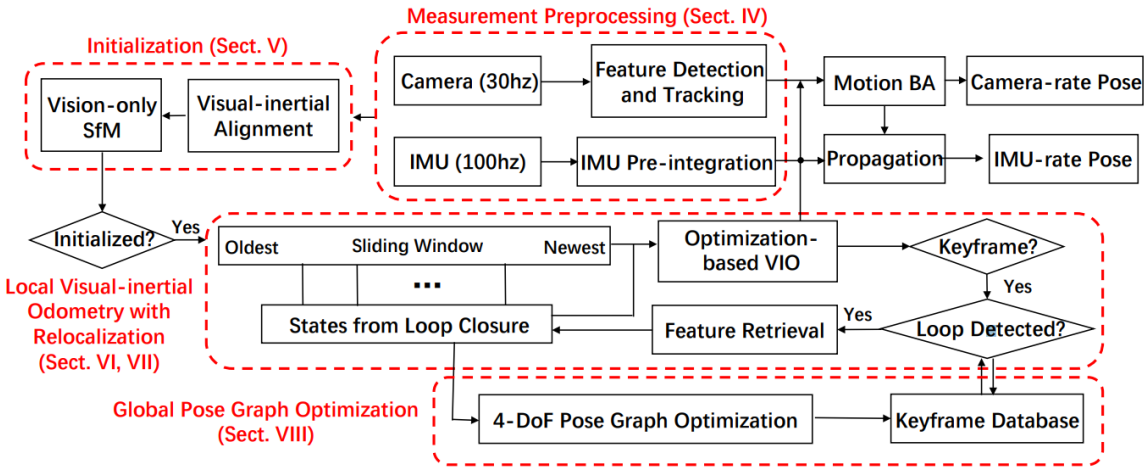
VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator

Shaozu Cao, Jie Pan, Jieqi Shi and Shaojie Shen
Hong Kong University of Science and Technology
{shaozu.cao, jpanai, jshias, eeshaojie}@connect.ust.hk



I. Original Framework

Our system is based on the VINS-MONO[1] framework illustrated in the picture.

We first extract features from images provided by a camera sensor and operate optical flow algorithms to complete a pre-treatment. Meanwhile, we do IMU pre-integration between two consecutive frames. After pre-processing everything we need, we perform a robust estimator initialization.

In the initialization procedure, we make use of visual structure from motion together with visual-inertial alignment. This loosely-coupled sensor fusion method offers us initial values we need to build a monocular tightly-coupled visual-inertial system.

To change from loosely to tightly coupled, we proceed with a sliding window-based optimization system for high accuracy and robust state estimation. We also employ marginalization to bound the computational complexity of our optimization-based VIO. What's more, a lightweight motion-only visual-inertial BA is carried out to boost the estimation.

We also consider failure detection and recovery in our framework. At the face of accumulated drifts, we propose a tightly-coupled re-localization module consisting of loop detection, feature retrieval and final re-localization steps. After re-localization, the local sliding window shifts and aligns with past poses. Further, we develop an additional pose graph optimization step to ensure the set of previous poses are registered into a globally consistent configuration.

II. Modifications

Inspired by the problems of limited by computation resources and optical flow residuals, we want to improve the efficiency and accuracy of our system. Therefore, we make some slight modifications to our framework and speed up the whole procedure.

We first add a flow-back step to the front-end of our system. After calculating optical flow between subsequent frames, we operate another flow from current to the previous frame and combine two flows to calculate a final result. It is a simple but efficient step to reduce outliers and improve robustness.

The second improvement is to increase the frequency of our original VINS procedure. We assume that operating the original steps of our framework only on keyframes will notably accelerate our system. Such an acceleration is quite important in the circumstances of poor computing power and strict frequency requirement. We fix landmarks built from features of the environment and optimize the poses of sensors between our selected frames. Such a motion-only BA ensures that the accuracy is not influenced while saving the running time.

III. References

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