A Filter-Based Visual-Inertial Odometry with RAFT

Can Huang[*], Ran Yan[*], Xiao Liu[*]

Abstract

*This report summarizes Megvii-3D team's approach to IROS 2020 UZH-FPV VIO Competition. The system carries out **V**isual-**I**nertial **O**dometry by taking stereo images together with IMU measurements. Our method includes FAST keypoints extraction and an optical flow neural network to perform feature tracking. The pose optimization is achieved by a filter-based MSCKF estimator implemented in OpenVINS, with some functional adaptions and parameter tuning. We consider the pose of IMU frame in our state vector. The camera-IMU time offset as well as camera intrinsics and extrinsics are also calibrated online.*

1 Summary of the Approach

1.1 Front-End Design

For visual tracking, FAST keypoints are extracted and tracked via optical flow. In order to guarantee those keypoints relatively equally distributed across the image, the image is preprocessed to be divided into small cells with fixed number of row and column. Different from the Open-VINS[1] system which implements KLT[2] to provide indirect sparse feature tracking, we use the state-of-the-art dense optical flow network **R**ecurrent **A**ll Pairs **F**ield **T**ransforms for Optical Flow [3]. It simultaneously outputs correspondences between temporal consecutive images as well as stereo images at the same timestamp.

Outperforming the traditional KLT method, RAFT extracts features using a modern convolutional neural network and uses a recurrent net with Gated Recurrent Units[4] to update the optical flow. Volumes are pooled and correlated at multiple scales. We utilize the pre-trained weights available online.

1.2 Back-End Design

The backend is mostly identical to MSCKF[5]. It is an Extended Kalman filter-based approach on manifold, which reports the IMU poses in the frequency of visual measurements. The idea of keyframes are not implemented. All frames are treated the same and the pose estimation is in the manner of sliding window. We fix the size of frames involved in the pose estimation and always marginalize out the older frame regardless of covisibility or temporal distance.

1.3 Corner Case Strategies

- Detection of bad IMU measurements and divergence of velocity
  When the drone attempts to land off in the sequences like *indoor forward 12*, it may undergo multiple collisions and bounces off the floor. These instances introduce significant disturbance to the IMU measurements, which lead to a temporarily divergent estimation of the velocity. Whereas, under that circumstance, the visual optical flow showing relatively small magnitude implies small displacements of the drone. Thus, we develop the system to automatically adjust its confidence towards the IMU and the visual measurements, based on the magnitude of optical flow. On top of that, if the magnitude of the dense optical flow is below a threshold, it is confident to announce the system is in stationary, and thus the velocity is forced to be zero. We achieve this feature in the same way as of OpenVINS[1] zero velocity update.

- Adaptive system parameters
  The values of parameters related to sensor noises will be automatically updated, according to the change of optical flow as well as IMU measurements.

- Adjustments according to map points distribution
  While performing the indoor deg 45 sequences, the drone flies in close proximity to the ground during some period. Other than lack of feature points due to low texture, the majority of the map points are rather close to the camera, causing poor rotation estimation. Under such condition, the drone is said to be in the *close* mode. A set of criteria are established to identify whether the drone is in the close mode. During the *close* mode, the reliability of IMU measurements in terms of rotation will be increased.

[*] 3D Team, Megvii Research {huangcan, yanran, liuxiao@megvii.com}

2 Parameters Setting

Identical parameters were used throughout all sequences. The details for parameter setting are demonstrated in Table 1.

Additionally, some parameters are dynamically adjusted, such as the FAST threshold and the threshold of optical flow to force zero velocity, as explained in the previous sections. The initial values of those parameters are provided here with annotation 'init'.

Parameter	Value
Slide Window Size	12
Max num. of Features	300
FAST Threshold(init)	10
Flow Magnitude Threshold	0.25

Table 1. Parameter Settings

3 Processing Time & Hardware

All the dataset were processed on a local device with Intel Core i7 CPU@1.80GHz×8 and 16GB memory with Graphical Card GeForce GTX 2080ti. The time used to evaluate each sequence is summarized in Table 2.

	Processing Time [Sec]
Indoor Forward 11	154.1
Indoor Forward 12	99.4
Indoor Deg45 3	155.8
Indoor Deg45 16	76.2
Outdoor Forward 9	141.9
Outdoor Forward 10	196.5

Table 2. Processing Time for Each Evaluation Sequence

4 Conclusion & Further Developments

In this report, we present our solution based on MSCKF[5] applied to IROS2020 UZH-FPV VIO Competition. Although it shows a decent result on the evaluation, we still believe there are many other promising strategies to further advance the behaviors of the system as a whole, such as loop closing with local bundle adjustment or global bundle adjustment, though it may significantly increase the workload of the system. Also, despite that the performance of RAFT optical flow net as a whole is consistently better than the KLT method, it requires longer time to process and relies on the GPU acceleration. Combining the usage of RAFT and KLT and developing a strategy to choose between them shows a potential to retain the current performance and meanwhile improve the efficiency. Those are not implemented in our system at the moment due to time restriction.

References

[1] Patrick Geneva, Kevin Eckenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. Openvins: A research platform for visual-inertial estimation. In *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020. 1

[2] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. of the International Joint Conference on Artificial Intelligence*, Vancouver, BC, 1981. 1

[3] Zachary Teed and Jia Deng. Raft: Recurrent all pairs field transforms for optical flow. In *Proceedings of the European Conference on Computer Vision*, Rome, Italy, 2020. 1

[4] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014. 1

[5] A. I. Mourikis and S. I. Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy, 2007. 1, 2