Performance Analysis of Visual/Inertial Integrated Positioning in Typical Urban Scenarios of Hong Kong

Xiwei Bai[1] Weisong Wen[2] and Li-Ta Hsu*[1]

[1]Interdisciplinary Division of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, Hong Kong

[2]Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong

*Corresponding author: Li-Ta Hsu

E-mail: lt.hsu@polyu.edu.hk, Tel.: +852-3400-8061

Abstract

There is an increasing demand for accurate and robust positioning in many application domains, such as the unmanned aerial vehicle (UAV) and autonomous driving vehicles (ADV). The integration of visual odometry and inertial navigation system (INS) is extensively studied to fulfill the positioning requirement. The visual odometry can provide aided positioning by matching consecutive frames of images. However, it can be sensitive to illumination conditions and features availability in urban environment. In this paper, we propose to evaluate the performance of tightly coupled visual/inertial integrated positioning in a typical urban scenario of Hong Kong based on existing state-of-the-art visual/inertial integration algorithm. The performance of visual/inertial integrated positioning is tested and validated in a typical urban scenario of Hong Kong which includes numerous dynamic participants, vehicles, pedestrians and trunks. The result shows that the visual/inertial integration can be affected in scenes with excessive dynamic objects

Keywords: Positioning; Navigation; Sensor fusion; Visual/inertial integrated positioning system

1. Introduction

Accurate and robust positioning is significant for the unmanned aerial vehicle (UAV) [1] and autonomous driving vehicles (ADV) [2] in an urban area. The integration of visual odometry [3] and inertial navigation system (INS) is extensively studied to fulfill the positioning requirement. The visual/inertial integrated positioning method is a promising solution for autonomous systems. The visual odometry can use camera images extracted feature points and match them with previous frames to provide aided positioning [4]. However, it can be sensitive to illumination conditions and features

availability. The low-cost inertial navigation system could provide high-frequency attitude and acceleration measurements. The recently proposed tightly coupled visual/inertial integration method [5] can obtain prominent positioning performance in constraint scenarios with enough environment features and ideal illumination conditions. The factor graph [6] is employed to integrate the visual odometry, visual loop closure [5] and INS raw measurements. At present, the main methods of Visual Odometry are divided into two parts based on feature points and direct methods without features. Based on the former, a method proposed to reduce the noise effects in the sequential trajectory reconstruction process, which could improve the accuracy in the feature points matching [7]. The Direct Sparse Odometry (DSO) is a visual odometry based on a direct structure and motion formulation, so it can sample pixels from all image regions with intensity gradient [8]. Unfortunately, the sole visual odometry are not robust to light conditions, dynamic changes, image conditions and the motion estimate still drifts without loop closure. These factors make localization difficult in outdoor environments [9-11]. However, the performance of visual/inertial integration can be challenged or degraded in a scenario with excessive dynamic objects. Since the violation of geometric constraints in a dynamic environment, and the optical flow characteristic will be affected [12].

In this paper, we propose to evaluate the performance of tightly coupled visual/inertial integrated positioning in a typical urban scenario of Hong Kong based on the work in [5]. The tested scenario can have numerous dynamic participants, vehicles, pedestrians and trunks, etc. Firstly, the state-of-the-art INS pre-integration technique [13] is employed to get the transformation between consecutive frames INS raw measurements to derive the INS factor. Then the feature-based visual odometry is conducted based on features matching to derive the visual odometry factor. Finally, we make use of the Ceres [14] to solve the factor graph optimization to get the optimal estimation of the positioning state set. The performance of visual/inertial integrated positioning is tested and validated in a typical urban scenario of Hong Kong. The result shows that the visual/inertial integration can be affected in scenes with excessive dynamic objects.

The main contributions of this paper are listed as follows:
1) We evaluate the performance of visual/inertial integrated positioning system in an urban scenario of Hong Kong with numerous dynamic objects.
2) We analyze the performance of the visual/inertial integration positioning versus the quality of visual feature tracking.

The rest of paper is organized as following: we discuss the methodology in Section 2 based on the VINS-Fusion framework. The performance analysis is shown in Section 3. Finally, the conclusion of this research is summarized in Section 4.

2. Methodology

The evaluated visual/inertial integrated positioning framework is based on the work in [5]. The flowchart is shown in Figure 1. The inputs are the sensor measurements from IMU and images from monocular camera. This system starts from raw measurements pre-processing of camera and IMU. And then the initialization provides all necessary values for nonlinear optimization. During the initialization, the loosely coupled sensor fusion is used to obtain the initial value. Firstly, the pure visual estimation of the pose of all the frames in the sliding window is performed by SFM, and then aligned with the IMU pre-integration, therefore obtaining the attitude, velocity, gravity vector and 3D feature location. The output is the position and orientation estimation. The detail of the evaluated visual/inertial integrated positioning algorithm can be found in [5]. We evaluate this technique on vehicle localization in deep urban area.
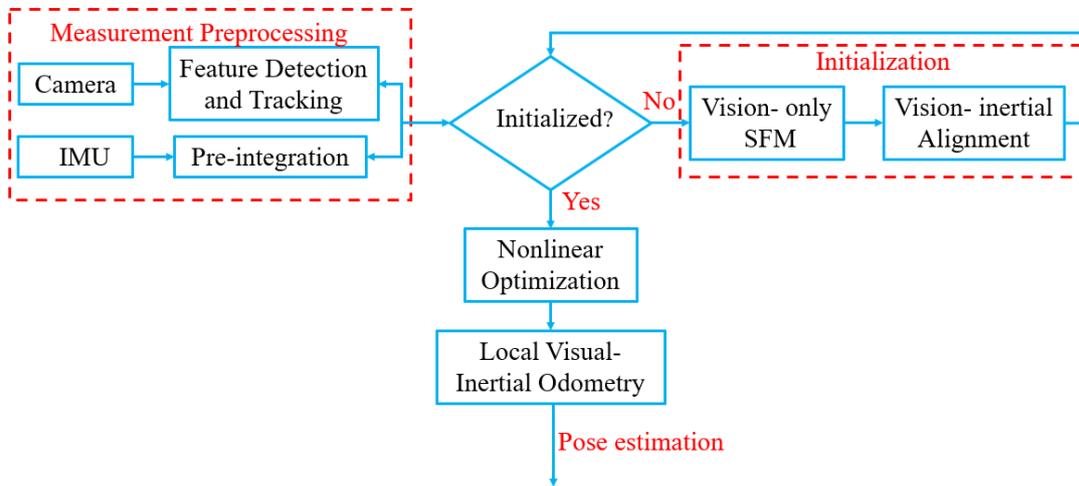
Fig. 1. Flowchart of the evaluated visual/inertial positioning system.

3. Experiment Results

a) Experiment Setup

The sensor setup is shown in the left-hand side of Figure 2 and the data is collected on 12th, April 2019. The IMU (Xsens Mti 30) is used to collect the high-frequency attitude and acceleration measurements. The monocular camera is used to capture consecutive images. Both IMU and camera are installed on top of a vehicle. The reference trajectory is provided by NovAtel SPAN-CPT (RTK GNSS/INS integrated positioning system). A dynamic experiment is conducted in an urban scenario in Hong Kong. The yellow curve in right-hand side of Figure 2 shows the tested trajectory.

The tested scenarios are shown in Figure 3. We can find that the illumination is varable during the test which can introduce significant challenge. The dynamic vehicles are passing through which can severely distort the performance of feature tracking process [5]. In short, the illumination and dynamic objects are the two major challenges. We believe that the evaluated scenarios can really be a challenging case for visual/inertial integrated positioning which is crucial for autonomous driving.

Fig. 2. The experiment setup
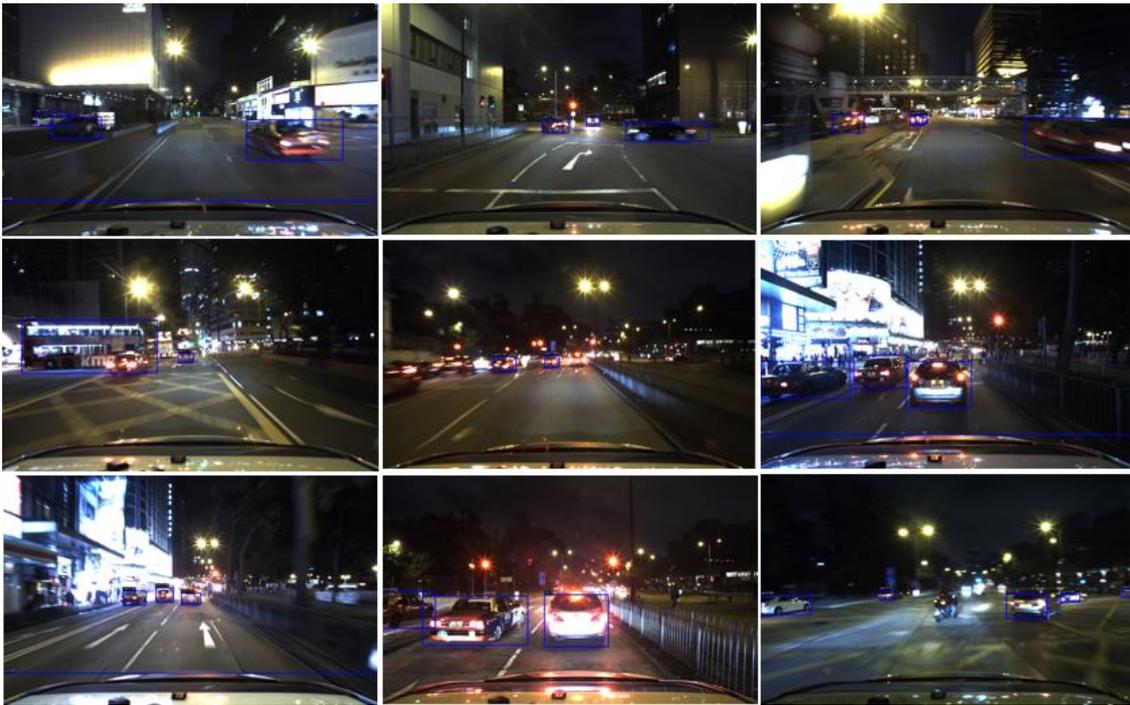
Fig. 3. Snapshots of the evaluated scenarios with numerous dynamic objects

b) Performance analysis

In this paper, we focus on analyzing the performance of the existing state-of-the-art visual/inertial integrated positioning performance in urban canyons.

In order to evaluate the performance of visual/inertial integrated positioning system, three aspects are analysed:

1) 2D positioning error VS velocity error: this part analyses the relationship between the 2D positioning error and the velocity error of visual/inertial integrated positioning.

2) 2D positioning error VS mean number of features tracking [5]: this part analyses the impact of average number of times each visual feature is tracked in each key frame for 2D positioning [5]. For more details of feature tracking can be found in [5].

3) 2D positioning error VS feature track difference: this part analyses mean number of features tracking difference between two consecutive frames which indicates the current features tracking number of times minus the one at last time

The trajectories of the visual/inertial and reference positioning are shown in Figure 4. Table I shows the 2D positioning performance of the evaluated visual/inertial integrated positioning system. The red curve represents the reference trajectory and the green curve represents the positioning from evaluated visual/inertial integrated positioning. Firstly, 34.21 meters of mean positioning error is obtained based on the evaluated method with a standard deviation of 15.49 meters. Moreover, the maximum error reaches 67.32 which is not acceptable for autonomous driving vehicle localization. The second column shows the 2D velocity error during the experiment. The mean error and standard deviation are 0.92 and 0.79 respectively. The detailed results of 2D positioning error and velocity error can be found in Figure 5. The third column shows the mean number of feature tracking with mean error of 35.56 and a standard deviation of 49.31 respectively. The last column shows mean number of feature tracking difference during the experiment with mean error and standard deviation are 5.72 and 9.29 respectively.

TABLE I

POSITIONING PERFORMANCE OF THE EVALUATED VISUAL/INERTIAL INTEGRATED POSITIONING

Items	2D error	2D velocity error	mean number of feature tracking	Mean number of feature tracking difference
Mean error	34.21m	0.92	35.56	5.72
Std	15.49	0.79	49.31	9.29
Maximum error	67.32	4.13		
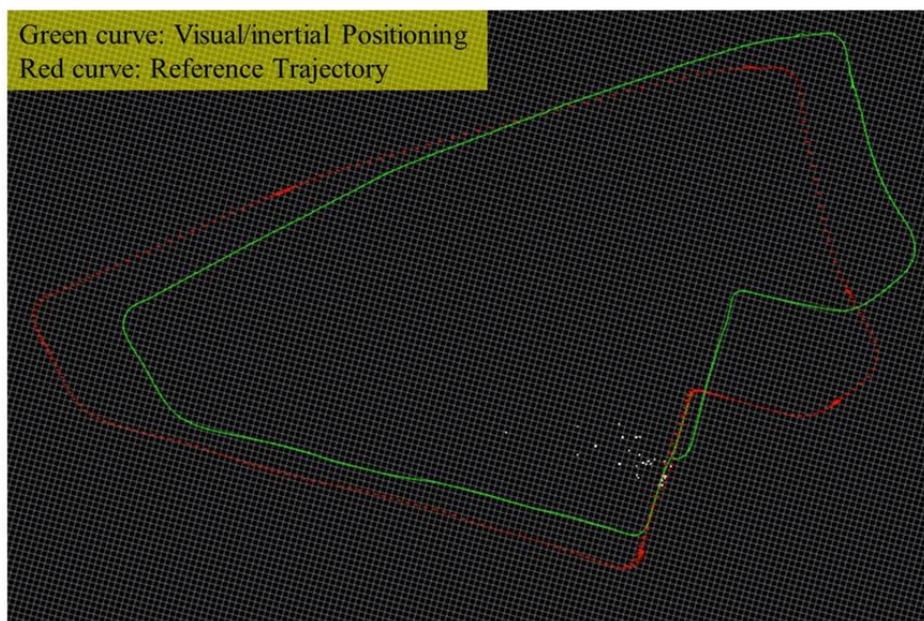
Fig. 4. The trajectories of the visual/inertial positioning (green curve) and reference

As the Figure 5 shows, the top panel shows the reference and VIO velocity and the bottom panel shows the 2D positioning error. We find that the 2D error increases significantly during epoch 20 to 60 and 100 to 200 when 2D velocity error is more than 0.92 (mean error). Moreover, when the VIO velocity during epoch 330 to 350 is more than 10, the 2D error also increases slightly. In short, the performance of the visual/inertial integrated positioning is correlated with the reference and VIO velocity.
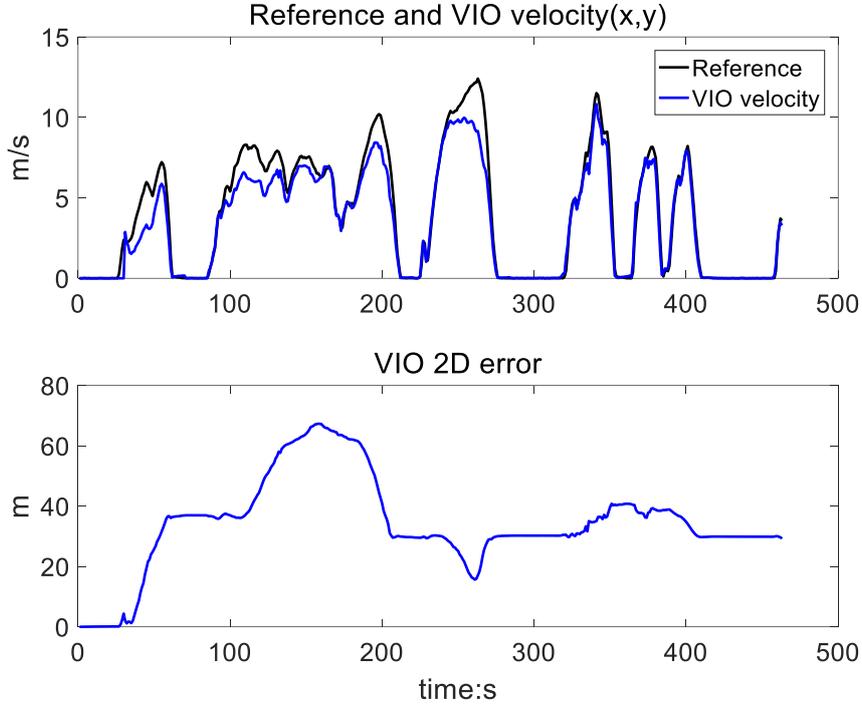
Fig. 5. Positioning error VS reference and VIO velocity

As the Figure 6 shows, the top panel shows the mean number of feature tracking and the bottom panel shows the 2D positioning error. We find that the bottom panel 2D error increases significantly during epoch 100 to 200 when mean number of feature tracking is limited (less than 20). As the Figure 3 shows, many dynamic objects and the illumination environment cause the features decrease. Moreover, when the velocity is zero (the vehicle stops), mean number of feature track rises significantly (epoch 70~80 and 270~330). In short, the performance of the visual/inertial integrated positioning is correlated with the number of feature tracking.
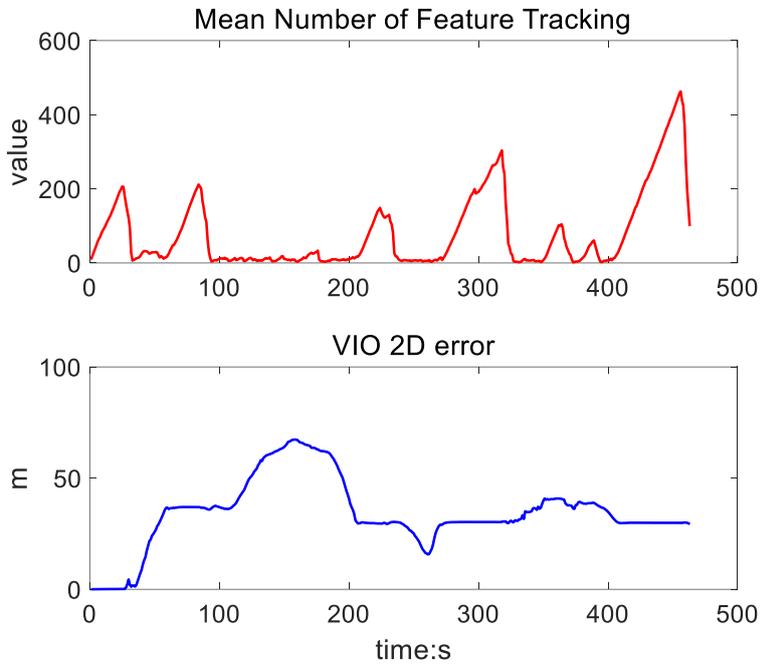
Fig. 6. Positioning error VS mean number of feature tracking

As the Figure 7 shows, the top panel shows the mean number of feature tracking difference and the bottom panel shows the 2D positioning error. We find that the bottom panel 2D error increases dramatically during epoch 20 to 60 when mean number of feature tracking difference changed significantly (more than 50). Interestingly, when the 2D error is very large during epoch 100 to 200, the feature tracking difference fluctuates slightly. As the Figure 3 shows, many moving objects affect the number of features over the period (less than 20). In short, the performance of the visual/inertial integrated positioning is correlated with the number of feature tracking difference.
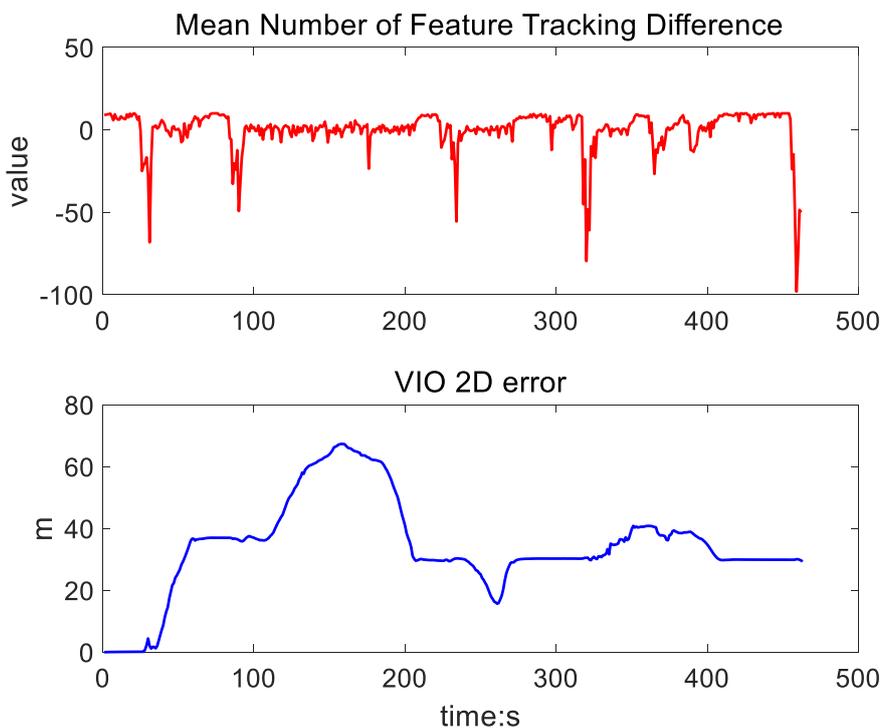
Fig. 7. Positioning error VS mean number of feature tracking difference

In conclusion, the performance of visual/inertial is affected by these factors: the velocity between reference and VIO, mean number of feature tracking, mean number of feature tracking difference. More importantly, numerous dynamic objects in the road will reduce the features tracking number and lead to heavy 2D positioning error. In the future, we plan to use YOLO, which is a real-time object detection system, to detect the moving objects and then remove them.

References

[1] F. Nex and F. Remondino, "UAV for 3D mapping applications: a review," Applied geomatics, vol. 6, no. 1, pp. 1-15, 2014.

[2] C. Urmson et al., "Autonomous driving in urban environments: Boss and the urban challenge," Journal of Field Robotics, vol. 25, no. 8, pp. 425-466, 2008.

[3] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, 2015.

[4] F. Steinbrücker, J. Sturm, and D. Cremers, "Real-time visual odometry from dense RGB-D images," in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 719-722: IEEE.

[5] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004-1020, 2018.

[6] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Factor graph based incremental smoothing in inertial navigation systems," in Information Fusion (FUSION), 2012 15th International Conference on, 2012, pp. 2154-2161: IEEE.

[7] S. Fiani and M. Fravolini, "A robust monocular visual odometry algorithm for autonomous robot application," IFAC Proceedings Volumes, vol. 43, no. 16, pp. 551-556, 2010.

[8] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 3, pp. 611-625, 2017.

[9] R. Gonzalez, F. Rodriguez, J. L. Guzman, C. Pradalier, and R. Siegwart, "Control of off-road mobile robots using visual odometry and slip compensation," Advanced Robotics, vol. 27, no. 11, pp. 893-906, 2013.

[10] K. Nagatani, A. Ikeda, G. Ishigami, K. Yoshida, and I. Nagai, "Development of a visual odometry system for a wheeled robot on loose soil using a telecentric camera," Advanced Robotics, vol. 24, no. 8-9, pp. 1149-1167, 2010.

[11] N. Nourani‑Vatani and P. V. K. Borges, "Correlation‑based visual odometry for ground vehicles," Journal of Field Robotics, vol. 28, no. 5, pp. 742-768, 2011.

[12] M. R. U. Saputra, A. Markham, and N. Trigoni, "Visual SLAM and structure from motion in dynamic environments: A survey," ACM Computing Surveys (CSUR), vol. 51, no. 2, p. 37, 2018.

[13] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2:

Incremental smoothing and mapping using the Bayes tree," The International Journal of Robotics Research, vol. 31, no. 2, pp. 216-235, 2012.

[14] S. Agarwal and K. Mierle, "Ceres solver: Tutorial & reference," Google Inc, vol. 2, p. 72, 2012.