INFLUENCE OF IMU QUALITY ON OPTIMIZATION-BASED VISUAL INERTIAL ODOMETRY

[1]*Kuan-Wei Tseng* (曾冠維), [2]*Meng-Wei Hsu* (許孟偉), [2]*Peng-Yuan Kao* (高鵬淵), [1,2]*Yi-Ping Hung* (洪一平)

[1]Dept. of Computer Science and Information Engineering, National Taiwan University, Taiwan
[2]Graduate Institute of Networking and Multimedia, National Taiwan University, Taiwan
E-mail: hung@csie.ntu.edu.tw

ABSTRACT

Low-cost inertial measurement units (IMU) has been applied to visual Simultaneous Localization and Mapping (vSLAM) to increase the positioning accuracy and robustness under challenging situations. In recent years, with the development of computing power, optimization-based visual inertial SLAM that minimizes the reprojection error and IMU residual has become popular. Intuitively speaking, using a carefully-calibrated high quality IMU with lower noise density and bias instability will achieve decent localization accuracy, and compromising performance will be expected if using consumer grade IMUs. However, the influence of IMU quality (i.e. noise parameters) on the optimization-based visual inertial odometry has not been quantified yet. In this research, we provide a comparative analysis on the visual inertial odometry using different level of IMU by simulation in Gazebo. The result shows that although the performance degrades when using poor IMU, there is no significant enhancement in the localization accuracy when the noise is smaller than a certain level. Also, further analysis indicates that except for extremely noisy sensors, scale and gravity estimation is accurate enough for most application.

Index Terms— Visual Inertial Odometry, IMU, Error Analysis

1. INTRODUCTION

Visual Simultaneous Localization and Mapping (vSLAM) has become a popular solution for robot localizations. However, the robustness and accuracy of vision-only method are known to degrade in several situations, including dynamic environment, texture-less area, low illumination, and pure rotation. To overcome such adversity, Inertial Measurement Units (IMUs) are integrated to establish a Visual-Inertial Odometry (VIO) system due to its complementariness with camera.

IMUs are low cost sensors consist of accelerometer and gyroscope, which measure the acceleration and angular velocity of the body, respectively. Operating at higher frequency than cameras, it is an indisputable fact that the quality of sensor has significant influence on the performance of VIO system. The quality here refers to the noise parameters, that is, bias stability and noise density. Generally speaking, pricey IMUs for navigational or tactical purpose has small bias and noise, while the quality of cheap IMUs on the consumer electronics are doubtful.

VINS-Mono [1], a state-of-the-art VIO system, has successfully demonstrate the reliability of their algorithm with the IMU on smartphone. However, the influence of IMU noise parameters on the VIO system has not been quantified. To analyze such influence, simulation is an appropriate option since we can get IMU measurement with different level of noise parameters at the same time. We built up simulated environment by Gazebo Simulator [2, 3], in which we manipulate a drone with camera and IMU. Several different level of bias and noise data are added to the IMU measurements. We tested these data on VINS-Mono and provide a comparative analysis in this research.

The followings are the structure of the rest of this paper. In Section II, we discuss on several related works. In Section III, we comprehensively elaborate on the experimental environment, including our test data, measurement model, and parameter settings, etc. We present our result in section IV with some discussion, and arrive at the conclusion in section V.

2. RELATED WORKS

As we mentioned in the previous section, Visual Inertial Odometry (VIO) is an alternative for vision-only system when encountering challenging situations. According to their back-end, they can be classified into filter-based and optimization-based method.

Especially for filter-based method, depending on the composition of their state vector, can be further categorized into loosely-coupled and tightly-coupled method [4]. In loosely-coupled method, visual and inertial measurements are processed separately and then fused together to get the output trajectory. Such framework requires less computational power, which is suitable for running on embedded system with limited resource. In contrast, visual and inertial measurements

are incorporated in tightly-coupled method. Since feature tracking process can be boosted by IMU integration, and the drift of sensors can be mutually corrected, tightly-coupled approaches provides more accurate results with the cost of more computational complexity.

With the development of computing power, optimization-based method has become an active research area in SLAM. It estimates the pose by a loss function to minimize the reprojection error and IMU measurement residuals. Since IMU operates at high rate, real-time optimization is an arduous problem. Pre-integration techniques [5] has successfully reduce the computational cost. Integration on the body frame instead of global frame can avoid re-propagating IMU measurement when the poses are adjusted. In addition, re-propagation due to bias change during optimization can also be avoided by analytic correction if the change of biases are small. Therefore, this framework is applied to many VIO systems such as VI-ORB [6] and VI-DSO [7], which are visual inertial extension from a well-known vision-only ORB-SLAM2 [8] and DSO [9], respectively.

VINS-Mono is an open-source nonlinear optimization-based visual inertial state estimator running on Robot Operating System (ROS). On one hand, similar to other VIO frameworks, pre-integration and marginalization strategies are adopted to bound the computational complexity. On the other hand, it features its robust initialization and loop closure technique. Despite its high resource usage, its outstanding performance among several famous VIO algorithms has been demonstrated by benchmark comparison [10] on EuRoC [11] dataset. Thus, we apply VINS-Mono to analyze the influence of IMU Noise on state estimation in this work.

Many researches have been conducted for IMU modeling, calibration, and simulation. For instance, Kalibr toolbox [12, 13] provides the calibration of IMU noise parameters, i.e., biases and noises. Lee et. al [14] had test three different odometry algorithms with FlightGoggles Simulator. They focus on the accuracy and resource usage of each algorithm. Their results shows that although ORB-SLAM2 is more accurate than VINS-Mono, the CPU usage is much more higher because the it need to track many features. However, in their work, the variance of IMU noise parameters are fixed, so we cannot evaluate the effect of IMU quality. Ma. et. al [15] had analyzed the influence of IMU's quality on MSCKF [16], a tightly-coupled filter-based visual inertial odometry, by simulation on KITTI odometry benchmark [17]. They synthesis ideal IMU measurments by inverse propogating ground truth pose collected by GPS, and add four different levels of IMU noise and bias. Their result shows that pure inertial navigation (Dead Reckoning) performs better than VIO if using high quality IMU. For bad IMU, performance of VIO can be improved by tracking more features. Nonetheless, there still lacks quantitative study on optimization based method.

3. SIMULATION

3.1. Environments

Gazebo is an open source robot simulator with robust physics engine. It features high quality graphic and compatibility with ROS, which is suitable for simulating multiple robots in complex environments. It provides abundant ready-made models that can be directly imported to create realistic simulation environments. We established two sets of scenes, indoor and outdoor, with those models. Figure 1 shows the simulated scenes we created. Vision landmarks are placed closer to the camera in the indoor scenes, resulting in more accurate visual measurements. In contrast, for outdoor scenes, the larger scale and farther vision landmarks reduces the visual measurement accuracy.

(a) Outdoor scene

(b) Indoor scene

Fig. 1. Simulation scenes

We set up a drone in the simulator with a forward-looking monocular camera and IMU. The camera captures images at 30 FPS, whose resolution is 640×480. We apply the ideal pinhole camera model without any distortion. The horizontal field of view is 60 degrees. The IMU measures the acceleration and angular velocity at 200 Hz. The ground-truth trajectory is collected at 1000 Hz which is the same rate as the simulator update. The drone is driven by a prescribed motion

Table 1. Different Levels of IMU Noise Parameters

Level	Perfect	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Unit
Acc. Noise	0	1.00E-04	5.00E-04	1.00E-03	5.00E-03	1.00E-02	3.60E-02	8.00E-02	$\frac{m}{s^2}\frac{1}{\sqrt{Hz}}$
Acc. Bias	0	5.00E-06	2.50E-05	5.00E-05	2.50E-04	5.00E-04	1.80E-03	4.00E-03	$\frac{m}{s^3}\frac{1}{\sqrt{Hz}}$
Gyr. Noise	0	5.00E-08	2.50E-07	5.00E-07	2.50E-06	5.00E-06	1.80E-05	4.00E-05	$\frac{rad}{s}\frac{1}{\sqrt{Hz}}$
Gyr. Bias	0	2.50E-09	1.25E-08	2.50E-08	1.25E-07	2.50E-07	9.00E-07	2.00E-06	$\frac{rad}{s^2}\frac{1}{\sqrt{Hz}}$

program, in which force and torque are exerted. As most visual inertial benchmarks do, we add translation and rotation on each axis (6DoF) at the beginning to enhance the accuracy and efficiency of initialization. In terms of VINS-Mono, this boosts the loosely-coupled initialization process which aligsn the pose obtained fro visual only Structure from Motion(SfM) and the posed constructed by IMU integration. Also to circumvent constant velocity, which jeopardize the stability of VIO, uniform distributed perturbation is added to 6-DoF motion program during runtime.

The specification of indoor and outdoor sequence are summarized in Table 2. Note that since the pose error of IMU propagates quadratically over time, the flight time of each sequence must be as consistent as possible. For both sequence, it is about one minute. Although the duration is the same, the length of the trajectory and the way of flight are different. The drone travels 102.31 meters in the outdoor sequence, with maximum linear velocity and angular velocity at 6 m/s and 0.8 rad/s, respectively. We make its trajectory form a loop to activate the loop closure in VINS-Mono. This has successfully eliminate the error caused by drift, which is particularly obvious when utilizing poor IMUs. As for the indoor sequence, the drone travels 54.96 meters in a small museum, with maximum linear velocity 4.39 m/s and maximum angular velocity 0.7 rad/s. Its trajectory forms a loop that covers the left half of the room. This also activates the place recognition module for loop clousre.

Table 2. Dataset Specifications

	Outdoor	Indoor
Trajectory length	102.31 m	54.96 m
Duration	61s	64s
Mean linear velocity	1.68 m/s	0.86 m/s
Max linear velocity	6 m/s	4.39 m/s
Mean angular velocity	0.5 rad/s	0.2 rad/s
Max angular velocity	0.8 rad/s	0.7 rad/s

3.2. IMU Modeling

We follow the classic IMU measurement model in Kalibr. Measured acceleration $\widehat{\mathbf{a}}$ and angular velocity $\hat{\boldsymbol{\omega}}$ in body coordinate is the actual acceleration $\mathbf{a}$ and actual angular velocity $\boldsymbol{\omega}$ added with bias $\mathbf{b}$ and noise $\mathbf{n}$. In terms of accelerometer, gravity $\mathbf{g}$ in world coordinate should be considered, and thus, rotational matrix from world coordinate to body coordinate R_w^b is applied.

$$\widehat{\mathbf{a}}(t) = \mathbf{a}(t) + \mathbf{b}_a(t) + \mathbf{n}_a(t) + R_w^b\,\mathbf{g}(t) \tag{1}$$

$$\hat{\boldsymbol{\omega}}(t) = \boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \tag{2}$$

where $\mathbf{b}_a$ and $\mathbf{b}_g$ are accelerometer and gyroscope bias modeled by Wiener process whose derivative are Gaussian white noise, that is,

$$\dot{\mathbf{b}}_a = \mathcal{N}\left(0, \sigma_{ba}^2\right) \quad \dot{\mathbf{b}}_g = \mathcal{N}\left(0, \sigma_{bg}^2\right) \tag{3}$$

$\mathbf{n}_a$ and $\mathbf{n}_g$ are the noise for accelerometer and gyroscope, which is modeled by Gaussian white noise, that is,

$$\mathbf{n}_a = \mathcal{N}\left(0, \sigma_a^2\right) \quad \mathbf{n}_g = \mathcal{N}\left(0, \sigma_g^2\right) \tag{4}$$

$\boldsymbol{\sigma}_{ba}, \boldsymbol{\sigma}_{bg}, \boldsymbol{\sigma}_a, \boldsymbol{\sigma}_g$ are the standard deviations that determine the bias stability and noise density. In our experiment, different levels of these standard deviations are applied to simulate various IMU performance level. Note that since these models are designed for continuous-time, we have to discretize before we add them to discrete-time IMU measurements.

$$\mathbf{b}_{k+1} = \mathbf{b}_k + \sigma_b\,\mathcal{N}(0,1)\sqrt{\Delta t} \tag{5}$$

$$\mathbf{n}_{k+1} = \sigma_n\,\mathcal{N}(0,1)\frac{1}{\sqrt{\Delta t}} \tag{6}$$

where Δt is the sampling time, σ_b and σ_n are the standard deviations for bias and noise.

3.3. Experimental Setup

We simulated 8 different levels of Inertial Measurement Units, including one one ideal (unbiased, noise-free) and 7 degraded IMUs. The corresponding noise and bias standard deviations are shown in Table 1. Note that Level 0 refers to ideal (perfect) IMU. Noisy measurement are generated by adding noise and bias after ideal IMU readings are captured in the manner we introduced in the previos subsection. These noise parameters are empirically inspired by real IMU devices. Typically, commercially available IMUs are grouped into four performance categories: navigation grade, tactical grade, industrial grade, and consumer grade. In our IMU performance level

Table 3. Median RMSE of ATE using different level IMUs

		Level 0	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Indoor	***7-DoF***	0.051	0.052	0.053	0.053	0.061	0.098	0.222	0.390
	6-DoF	0.084	0.085	0.086	0.085	0.084	0.111	0.282	0.499
	4-DoF	0.084	0.085	0.086	0.085	0.084	0.112	0.282	0.499
Outdoor	***7-DoF***	0.146	0.145	0.148	0.152	0.202	0.250	0.665	1.765
	6-DoF	0.303	0.302	0.148	0.325	0.385	0.467	0.912	1.765
	4-DoF	0.304	0.303	0.307	0.326	0.386	0.468	0.954	1.930

setting, we regard Level 0 and Level 1 as high-end IMUs better than tactical grades. These IMUs are installed on aircrafts or ships for accurate navigational purposes. Level 2 and Level 3 are considered as industrial-grade IMUs. They have good cost performance ratio and are often used for visual inertial sensors on drone. Level 4, Level 5, Level 6 are treated as consumer grade IMUs. Despite the fact that those IMUs lack logn term accuracy, they are functionally sufficient to meet the needs of consumer electronics. As for Level 7, as far as we know, most available IMUs can outperfom, and therefore, it is set as a lower bound for poor IMUs.

3.4. Software and Hardware Specifications

We specify the configuration settings for VINS-mono here. We enable the optimization on extrinsic parameters with an initial value. The maximum number of feature is increased to 200 and the minimum distance between features is decreased to 10 pixels. We also enable the loop closure and time offset estimation. Other parameters not mentioned are the same as those provided by the author. As for the hardware, our computer has a hex-core Inter Core i7-9700 with multi-threading, operates at 3.0GHz, and 32GB of RAM, which is sufficient resource for running VINS-Mono.

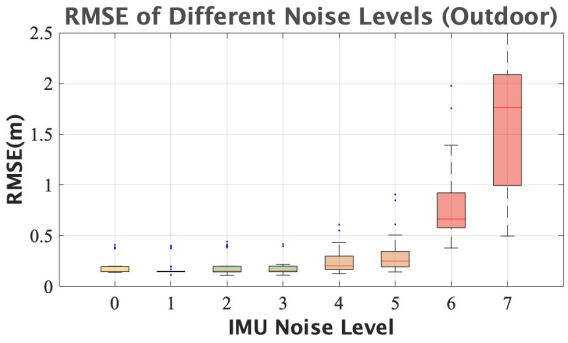
4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Evaluation Metric

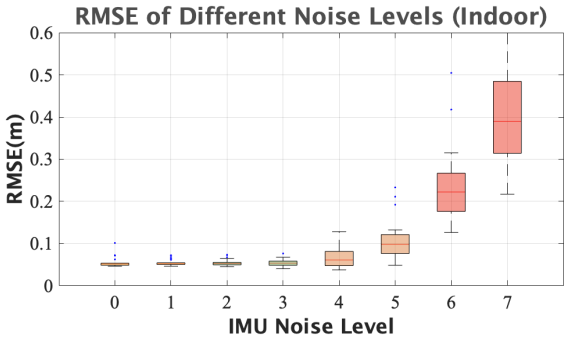
We use absolute trajectory error (ATE) [18] to evaluate the accuracy for each estimation since it is a single number metric, which is straight-forward to compare. It seeks a optimal similarity transform $\mathbf{S}' = \{s', R', t'\}$ between estimated positions $\{\widehat{p}\}_{i=0}^{N-1}$ and ground-truth positions $\{p\}_{i=0}^{N-1}$ with the following cost function:

$$\mathbf{S}' = \underset{\mathbf{S}=\{s,\mathbf{R},\mathbf{t}\}}{\arg\min} \sum_{i=0}^{N-1} \|\mathbf{p}_i - s\mathrm{R}\hat{\mathbf{p}}_i - \mathbf{t}\|^2 \tag{7}$$

The above equation is a generalized form which optimizes an up-to scale transformation with seven degrees of freedom, including translation, rotation, and scale. This approach is utilized to evaluate most monocular SLAM methods. In practice, different alignment criteria can be applied. On one hand, if the system provides an up-to-scale measurements (e.g. SLAM using stereo or RGB-D cameras), we can align the trajectory with only rigid body transform (six degrees of freedom) to analyze its scale accuracy. On the other hand, if the system can recover the gravity with scale and direction (e.g. Visual-Inertial SLAM), which makes its roll and pitch angle become observable, we can align the trajectory with Yaw-only rigid body transform (four degrees of freedom) the trajectory to analyze its performance on gravity estimation.



(a) Outdoor Sequence



(b) Indoor Sequence

Fig. 2. RMSE of ATE using 7-DoF alignment

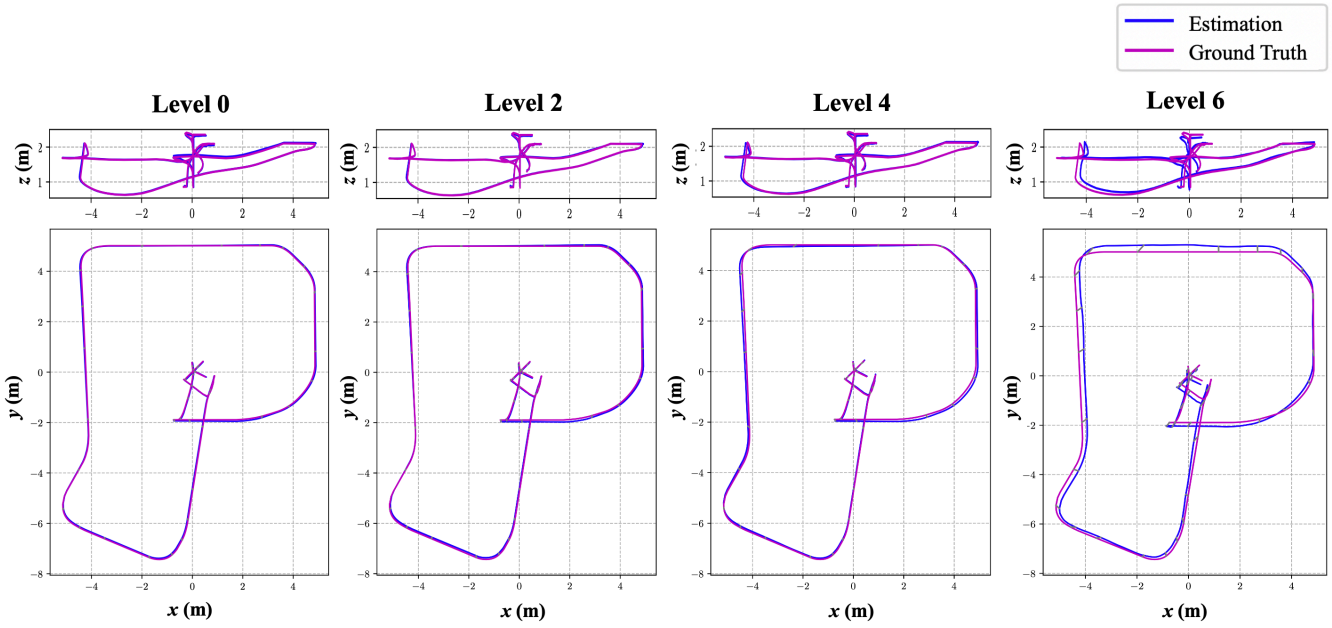


Fig. 3. Trajectory of indoor sequence after 7-DoF alignment

4.2. Major Results

In this part, we discuss the major finding through the indoor and outdoor experiments. We ran 30 times with different random seeds for each IMU performance level. Figure 2 shows the box graph of RMSE of ATE using similary transform alignment (7-DoF alignment) for both indoor and outdoor sequences. Figure 3 and Figure 4 show the estimated and ground-truth trajectory of indoor and outdoor sequence after 7-DoF alignment. Table 3 shows the meidan RMSE of ATE using differenct level IMUs under 4, 6, or 7 degrees of freedom alignment criteria.

First of all, due to the smaller scale and shorter distrance to the vision landmarks, it is expectable result that it produces less localization error in indoor sequence. For example, when using an Level 2 IMU, the meidan RMSE of indoor sequence is about is 0.054m compared to 0.148m for outdoor sequence. Secondly, we observed that no matter indoor or outdoor, they share the same degrading trend. We could hardly identify the influence of IMU noise from Level 0 to Level 3, we could hardly identify the influence of IMU noise. Performance in both sequence starts to degrade from Level 4. Moreover, owing to randomized large noise, the result varies, producing more unstable results with higher standard deviations. Failure in loop closure makes catastrophic error frequently happen at Level 7. The PnP test with RANSAC [19] in VINS-Mono after place recognition rejects the loop candidates that lacks spatial connectivity. However, if the geometric constraint is constructed incorrectly, it also eliminates the possibility of recovery through loop clousre. Thanksfully, in most commercially available IMU, such extreme condition are less likely to happen. Therefore, we can conclude that most devices are applicable to VINS-Mono to reach an acceptable performance. If user wants to get more precise and accurate measurements, industrial level IMU, which is about Level 2 or 3 in our experiment, should be considered. Though it the improvement seems minor, more stable result can be expected. However, the result also indicates that utilizing navigational level IMU or even perfect IMU cannot enhance the positioning accuracy in terms of visual inertial odometry.

4.3. Analysis on Scale and Grivity Direction

In this part, we look further into the estimation result. We discuss the scale error and gravity direction of VINS-Mono under different IMU levels.

4.3.1. Scale

In VINS-Mono, scale estimation is done at loosely-coupled initialization where IMU and camera measurments are processed separately and aligned afterwards. Same as most visual inertial odometry, it recovers metric scale from IMU information. In detail, it is obtained from gravity magnitude measured by accelerometers. Thus, noise and bias on the gravity direction will affect the scale accuaracy.

When applying 7-DoF alignment criteria, we find an optimal scale factor s to optimize the absoulte trajectory error. If the scale factor s close to 1, then we can say the scale estimation is accurate. Table 4 shows the median RMSE scale

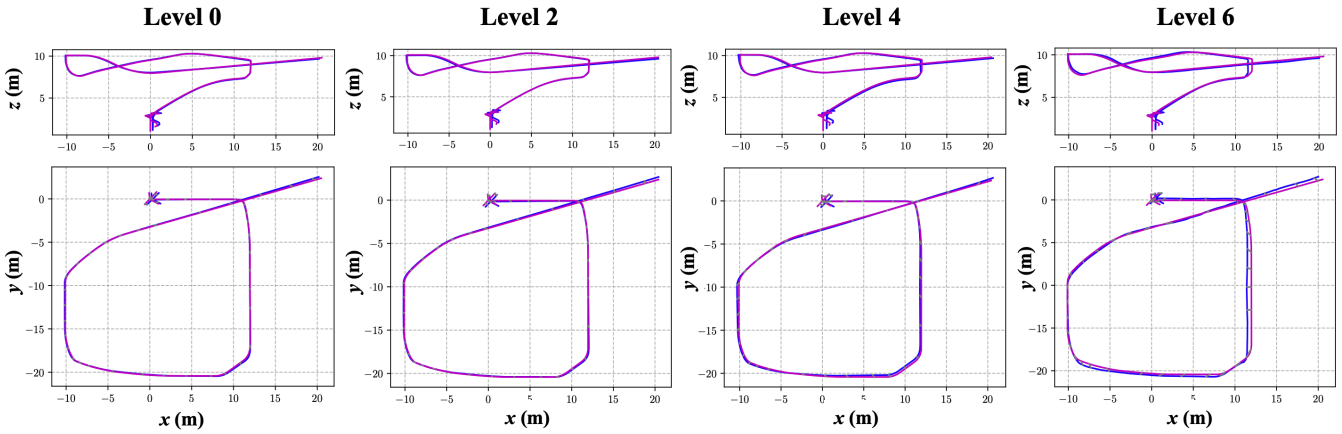


Fig. 4. Trajectory of outdoor sequence after 7-DoF alignment

errors of some selected IMU levels, and Figure 5 shows the box graph of all levels.

Table 4. Median RMSE of scale (%)

	Level 0	Level 2	Level 4	Level 6
Indoor	1.200	1.224	1.464	3.438
Outdoor	2.743	2.761	2.643	6.742

We can observe that the scale estimation is consistently accurate if the IMU quality is better than Level 5. Since the magnitude of acceleration on gravity direction does not change a lot. Most scale error lies under 5%, which can be considered reliable. A significant increase in the error and instability starts from Level 6. For small scale application, it provides a scale error at about 3%, which may still be acceptable. However, for the large scale case, where the accumulation of scale error is obvious, it might be fatal for precise robot control.

4.3.2. Gravity Direction

During the initilization of VINS-Mono, it performs gravity refinement after linear initialization step. It finds an optimal gravity vector iteratively by 2-DoF perturbation (since magnitude of gravity is known) until the loss is converged. Once the gravity direction is determined, the normal of x-y plane is also set in the odometry system.

When applying 6-DoF alignment criteria, we find an optimal rigid body transformation $T = \{R, t\}$, whereas in 4-DoF alignment criteria, roll and pitch are removed. This strategy can help us to evaluate the how accurate gravity direction is estimated. If there is noticeable difference between translational RMSE of 6-DoF and 4-DoF alignment, we can infer that the roll and pitch angles that VIO has estimated might be unreliable.

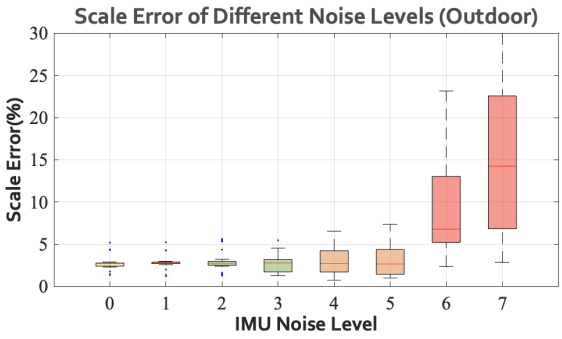
Table 5 shows the median RMSE of outdoor sequence for the selected IMU levels. Experiment shows that influence of IMU noise and bias on the estimation of gravity direction are small if using high quality IMUs. For outdoor case, we can observe trivial difference at IMU worse than Level 4. However, from our perspective, 1mm or 4cm error is neglectable compare to 100-meter long trajectory. As for the indoor case, there is no RMSE difference across all levels. It shows that dual optimization strategy that perturbs gravity vector after linear optimization adopted by VINS-Mono is effective when dealing with large IMU noise and bias.

Table 5. Median RMSE of outdoor sequence

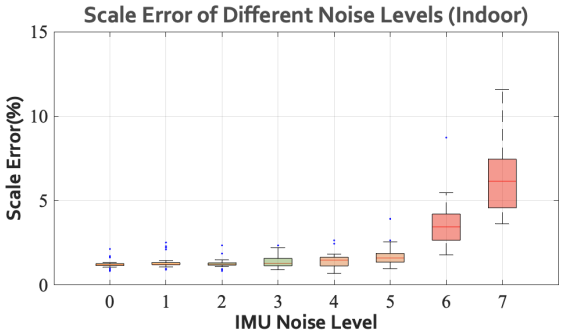
	Level 0	Level 4	Level 5	Level 6
6-DoF (m)	0.304	0.385	0.467	0.912
4-DoF (m)	0.304	0.386	0.468	0.954
Difference (m)	0	0.001	0.001	0.042

5. CONCLUSION AND FUTURE WORKS

In this study, we quantitatively analyzed the influences of IMU's quality on VINS-Mono. Simulated indoor and outdoor scenes are set up for testing different IMU noise parameters. We found that the performance of VINS-Mono, which is tightly-coupled and optimization-based method, is not susceptible to sensor quality when utilizing high quality sensors. With some engineering assumption, we show that consumer grade IMUs, which are not calibrated strictly, might have some impact on localization accuracy and precision. Since VI-SLAM obtain scale information from accelerometers, if the noise on gravity direction are sufficiently large, the scale estimation will become inaccurate. As for the estimation on gravity direction, the result is consistently correct except for some extreme cases due to the robust initialization process in VINS-Mono. We believe that these conclusions are helpful



(a) Outdoor Sequence



(b) Indoor Sequence

Fig. 5. RMSE of Scale Estimation

for users to select suitable grade sensor for their application.

There are still many directions for future research. First of all, we can test the algorithm with more versatile scenes to make the result more persuasive. We can either create more simulated scenes or generate ideal IMU measurements from various benchmark datasets. Secondly, we can test more visual inertial frameworks. Although many VIO adopt pre-integration under the nonlinear optimization framework, their initialization and optimization strategies still have many differences. In addition, loosely-coupled method can also be analyzed. Last but not least, real world experiments based on real IMU devices can be conducted. We realize that modeling in simulation may not perfectly cover things that will happen in the real world. Verifying the result with real world experiment is helpful to reinforce our argument.

ACKNOWLEDGEMENT

This study was partially supported by Ministry of Science and Technology, Taiwan, under grant no. MOST 107-2221-E-369-001-MY2, and no. MOST 109-2218-E-369 -001 -.

REFERENCES

[1] Tong Qin, Peiliang Li, and Shaojie Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[2] Nathan Koenig and Andrew Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, Sep 2004, pp. 2149–2154.

[3] C.E. Aguero, N. Koenig, I. Chen, H. Boyer, S. Peters, J. Hsu, B. Gerkey, S. Paepcke, J.L. Rivero, J. Manzo, E. Krotkov, and G. Pratt, "Inside the virtual robotics challenge: Simulating real-time robotic disaster response," *Automation Science and Engineering, IEEE Transactions on*, vol. 12, no. 2, pp. 494–506, April 2015.

[4] Davide Scaramuzza and Zichao Zhang, "Visual-inertial odometry of aerial robots," *Springer Encyclopedia of Robotics*, 2019.

[5] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.

[6] Raúl Mur-Artal and Juan D Tardós, "Visual-inertial monocular slam with map reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, 2017.

[7] L. Von Stumberg, V. Usenko, and D. Cremers, "Direct sparse visual-inertial odometry using dynamic marginalization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 2510–2517.

[8] 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*, vol. 33, no. 5, pp. 1255–1262, 2017.

[9] Jakob Engel, Vladlen Koltun, and Daniel Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[10] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 2502–2509.

[11] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, 2016.

[12] Paul Furgale, Joern Rehder, and Roland Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

[13] Joern Rehder, Janosch Nikolic, Thomas Schneider, Timo Hinzmann, and Roland Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4304–4311.

[14] E. M. Lee, I. Wee, T. Kim, and D. H. Shim, "Comparison of visual inertial odometry using flightgoggles simulator for uav," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, 2019, pp. 1166–1169.

[15] Xianglu Ma, Xiaoshan Yao, and Rensong Ding, "Influence of IMU' s quality on VIO: based on MSCKF method," in *Sixth Symposium on Novel Optoelectronic Detection Technology and Applications*, Junhao Chu and Huilin Jiang, Eds. International Society for Optics and Photonics, 2020, vol. 11455, pp. 1018 – 1024, SPIE.

[16] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007, pp. 3565–3572.

[17] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[18] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 7244–7251.

[19] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155, 2009.