

An EKF-based Fusion of Visual-Inertial Odometry and GPS for Global Robot Pose Estimation

Nguyen Hoang Khoi Tran, Vinh-Hao Nguyen
Department of Control Engineering and Automation
Ho Chi Minh City University of Technology (HCMUT), VNU-HCM
Ho Chi Minh City, Vietnam
thknguyen@hcmut.edu.vn, vinhhao@hcmut.edu.vn

Abstract—Globally accurate and drift-free pose estimation is essential for long-term navigation of autonomous robots. In this paper, we present an efficient method to fuse visual-inertial odometry and GPS measurements using the Extended Kalman Filter (EKF). The filter state is propagated by relative visual-inertial estimates and updated by absolute GPS readings to eliminate the accumulated drift on position and heading. We built the sensor hardware and implemented the algorithm on an embedded computer for real-time computation. The experimental datasets were collected in an outdoor environment with 6-degree-of-freedom ground truth. Evaluation results showed that the proposed algorithm achieved consistent and accurate pose estimation with different motion trajectories.

Keywords—multi-sensor fusion, robot localization, extended Kalman filter, visual-inertial odometry, GPS

I. INTRODUCTION

Localization is a fundamental demand for autonomous robot navigation. In recent years, many studies on Visual-Inertial Odometry (VIO), which combines cameras and inertial measurement unit (IMU) for pose estimation, have achieved impressive accuracy and high update rate in both indoor and outdoor environments [1], [2]. However, as a local localization approach, VIO algorithms accumulate drift in position and heading over time and are not sufficient for long-term navigation. To solve this problem, global localization measurements such as the Global Positioning System (GPS) can be employed. Thanks to satellite observations, GPS is able to deliver globally consistent position estimates regardless of the distance traveled. Nevertheless, GPS measurements are noisy, available at low frequencies and often degraded by obstacles, making them unreliable for precise localization. With complementary nature, the integration of VIO and GPS can overcome the shortcomings of each modality to facilitate high-rate, globally accurate and robust pose estimation.

In the literature, there are two major approaches on fusing visual, inertial and GPS data for robot navigation: smoothing and filtering. Smoothing algorithms estimate a chain of recent states using graph-based optimization to provide statistically optimal results. In [3], VIO estimations were loosely fused with global measurements in a sliding-window pose graph optimization. From a similar graph structure, [4] added an extra node representing the local frame to constraint the absolute orientation. In contrast to [3] and [4], [5] formulated a tightly-coupled optimization that considers the correlations among all sensors to improve estimation accuracy. Although giving superior performance, smoothing methods come with high computational costs that limit their applicability on

embedded systems. By marginalizing old measurements and only updating the latest state, filtering approaches enable efficient recursive computation that is favorable for online estimation. In [6], delayed measurements from different sensors were fused with IMU-driven states using the Iterative Extended Kalman filter (IEKF), achieving sub-milliseconds processing time. [7] constructed a robust Unscented Kalman Filter (UKF) that combines visual odometry, GPS, barometer and IMU measurements for micro aerial vehicle navigation. In [8], online spatiotemporal GPS-VIO calibration was performed based on the tightly-coupled Multi-State Constraint Kalman Filter (MSCKF) framework [9].

Despite exhibiting excellent results, only a few researches provided orientation evaluation due to the lack of orientation ground truth in their experimental datasets. Moreover, by updating all three axes of orientation in the fusion, most algorithms let the especially noisy altitude measurement of GPS affect roll and pitch estimates.

In this paper, we propose an EKF-based fusion of VIO and GPS measurements for globally accurate and drift-free pose estimation. The filter takes VIO pose estimations for state prediction and GPS position measurements for state update. Only position and yaw angle are included in the state vector to keep the original precision of VIO roll and pitch estimates. The experimental datasets are recorded in real-world conditions, with 3D position and 3D orientation ground-truth for full pose evaluation.

In short, the key contributions of our work are:

- a simple but effective algorithm that fuses VIO and GPS measurements to achieve high-rate and globally accurate robot pose estimation.
- a light-weight EKF framework for local-global fusion to eliminate drift on position and heading.
- a hardware and software development and real-world experiments to evaluate the sensor performance in outdoor environments.

The remainder of this paper is organized as follows. Section II introduces the notation, coordinate frames and system structure. Section III provides a brief overview of the employed VIO system. Section IV describes the proposed EKF-based VIO-GPS fusion algorithm. Section V demonstrates the system setup, experiments and assessment results. Finally, the work conclusion is given in Section VI.

II. SYSTEM OVERVIEW

A. Notation

Let $\mathcal{F}_A, \mathcal{F}_B, \mathcal{F}_C$ denote three arbitrary coordinate frames in 3D space. The pose of $\mathcal{F}_B$ with respect to $\mathcal{F}_A$ is represented by the position vector $\mathbf{p}_{AB} \in \mathbb{R}^3$ and the rotation matrix $\mathbf{R}_{AB} \in SO(3)$. The reference frame for the pose of $\mathcal{F}_B$ can be transformed from $\mathcal{F}_A$ to $\mathcal{F}_C$ using $\mathbf{p}_{CB} = \mathbf{R}_{CA}\mathbf{p}_{AB} + \mathbf{p}_{CA}$ and $\mathbf{R}_{CB} = \mathbf{R}_{CA}\mathbf{R}_{AB}$. The body-center velocity of $\mathcal{F}_B$ is denoted by ${}_B\mathbf{v}$. The skew symmetric matrix of a vector is marked as $[\cdot]^\times$.

B. Coordinate frames

The coordinate frames involved in this work are depicted in Fig. 1. The camera, IMU and GPS sensors are mounted on a same mobile vehicle. The VIO estimation is performed between the local world frame $\mathcal{F}_W$ and the IMU frame $\mathcal{F}_I$, which is chosen as the vehicle body frame. The GPS measurement is made between the global reference frame presented in East-North-Up (ENU) sequence, $\mathcal{F}_G$, and the GPS antenna frame $\mathcal{F}_A$. Both the Z axes of $\mathcal{F}_W$ and $\mathcal{F}_G$ are aligned to the Earth's gravitational vector.

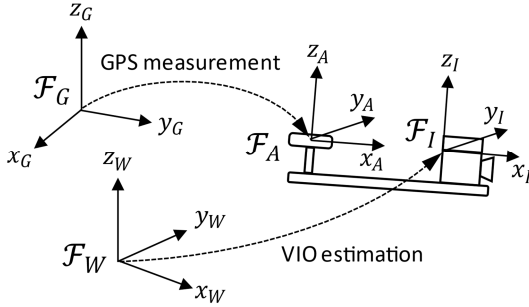


Fig. 1. Coordinate frames used in the localization system.

C. System structure

Our localization system follows a cascading structure as shown in Fig. 2. First, stereo images Π_j^{c1}, Π_j^{c2} are coupled with inertial measurements including linear acceleration $\mathbf{a}_k^i$ and angular velocity $\boldsymbol{\omega}_k^i$ to create the VIO state estimate $\hat{\mathbf{x}}_k^v$. Then, the VIO state is corrected with GPS measurements $\mathbf{z}_l^g$ through the EKF to produce the VIO-GPS state estimate $\hat{\mathbf{x}}_k^e$. The uncertainty terms of VIO and GPS measurements, denoted by $\boldsymbol{\sigma}_k^v$ and $\boldsymbol{\sigma}_l^g$, are continuously provided by the two subsystems to help the filter adapt to the current noise condition. The time indices of camera, IMU and GPS frames are marked as j, k and l , respectively.

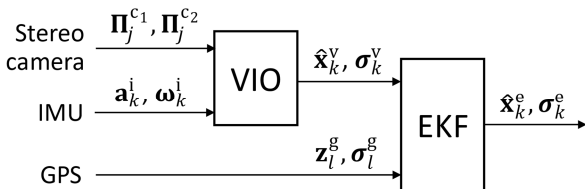


Fig. 2. Structure of the sensor fusion system.

III. VISUAL-INERTIAL ODOMETRY SYSTEM

The VIO subsystem is developed as our previous work in [10]. To begin with, we take special care of synchronization

and calibration between the camera and the IMU. The calibration datasets are collected under ideal light and motion conditions and processed by the well-developed Kalibr toolbox [11] to produce intrinsic and extrinsic calibration parameters of the stereo camera, alongside the spatial and temporal transformation between the camera and the IMU. Moreover, we continuously read and compensate for the time-varying camera's shutter period to ensure tight synchronization between images and inertial measurements.

After the preparation steps, we perform the VIO estimation by utilizing the state-of-the-art Open Keyframe-based Visual-Inertial SLAM (OKVIS) method [12]. This algorithm solves for the optimal robot states by minimizing a joint cost function of reprojection errors of visual landmarks and IMU errors in a bounded sliding window. Its front-end detects, describes and matches image keypoints to associate camera frames and initialize 3D landmarks. A number of important camera frames called keyframes are maintained in a sliding window to reduce the observation uncertainty of landmarks. Older and unimportant frames are marginalized out for computational efficiency.

IV. EKF-BASED VIO-GPS FUSION ALGORITHM

In this section, we will describe the proposed EKF-based algorithm for combining VIO and GPS data. The filter state vector is defined as:

$$\mathbf{x}^e = [\mathbf{p}_{WI}^T \quad \mathbf{v}_{WI}^T \quad \psi_{WI}]^T \in \mathbb{R}^7 \quad (1)$$

where $\mathbf{p}_{WI}$ and $\mathbf{v}_{WI}$ represent the vehicle position and velocity in the local world frame. Since the VIO system already provides observability of global roll and pitch angles [13], only the yaw angle ψ_{WI} is corrected in this fusion.

A. Process model

The continuous-time state kinematics is expressed as:

$$\begin{aligned} \dot{\mathbf{p}}_{WI} &= \mathbf{R}_{WI} \mathbf{v}^v + \mathbf{n}_p \\ \dot{\mathbf{R}}_{WI} &= \mathbf{R}_{n_\psi} \dot{\mathbf{R}}_{WI}^v \end{aligned} \quad (2)$$

where $\mathbf{n} = [\mathbf{n}_p^T, n_\psi]^T$ includes Gaussian white noise processes of VIO estimation with covariance $\mathbf{Q} = \mathbf{E}\{\mathbf{n}\mathbf{n}^T\} = (\boldsymbol{\sigma}^v)^2 \mathbf{I}_4$, and $\mathbf{R}_{n_\psi}$ is the rotation matrix of yaw noise. We provide the model with relative terms such as VIO linear velocity ${}_I\mathbf{v}^v$ and rotation velocity $\dot{\mathbf{R}}_{WI}^v$ to prevent adding back the accumulated drift to the filter state.

For EKF implementation, we apply the first-order approximation on (2) to obtain the discrete-time process model as:

$$\begin{aligned} \mathbf{x}_k^e &= \mathbf{f}(\mathbf{x}_{k-1}^e, \mathbf{u}_k^v, \mathbf{n}_k) \Leftrightarrow \\ \begin{bmatrix} \mathbf{p}_{WI_k} \\ \mathbf{v}_{WI_k} \\ \psi_{WI_k} \end{bmatrix} &= \begin{bmatrix} \mathbf{p}_{WI_{k-1}} + \mathbf{R}_{WI_{k-1}} \mathbf{p}_{I_{k-1}I_k}^v + \mathbf{n}_{p_k} \Delta t^v \\ \mathbf{R}_{WI_{k-1}} \mathbf{v}_{I_{k-1}I_k}^v + \mathbf{n}_{p_k} \\ \psi(\mathbf{R}_{WI_{k-1}} \mathbf{R}_{I_{k-1}I_k}^v) + n_{\psi_k} \Delta t^v \end{bmatrix} \end{aligned} \quad (3)$$

where Δt^v is the sampling period of VIO system. $\psi(\mathbf{R}) = \text{atan2}(R_{21}, R_{11})$ computes the yaw angle of the provided rotation matrix. The input vector $\mathbf{u}^v$ consists of the relative deviations of VIO position, velocity and orientation between two consecutive frames as:

$$\mathbf{u}_k^v = \begin{bmatrix} \mathbf{p}_{I_{k-1}I_k}^v \\ \mathbf{v}_{I_{k-1}I_k}^v \\ \mathbf{R}_{I_{k-1}I_k}^v \end{bmatrix} = \begin{bmatrix} (\mathbf{R}_{WI_{k-1}}^v)^T (\mathbf{p}_{WI_k}^v - \mathbf{p}_{WI_{k-1}}^v) \\ (\mathbf{R}_{WI_{k-1}}^v)^T \mathbf{v}_{WI_k}^v \\ (\mathbf{R}_{WI_{k-1}}^v)^T \mathbf{R}_{WI_k}^v \end{bmatrix} \quad (4)$$

Linearizing (3) around the last state estimate $\hat{\mathbf{x}}_{k-1}^e$, the Jacobian matrices of state transition and process noise are:

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_3 & 0 & [\mathbf{e}_z]^T \hat{\mathbf{R}}_{WI_{k-1}} \mathbf{p}_{I_{k-1}I_k}^v \\ 0 & 0 & [\mathbf{e}_z]^T \hat{\mathbf{R}}_{WI_{k-1}} \mathbf{v}_{I_{k-1}I_k}^v \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$\mathbf{L}_k = \begin{bmatrix} \mathbf{I}_3 \Delta t^v & 0 & 0 \\ \mathbf{I}_3 & 0 & 0 \\ 0 & \Delta t^v & 0 \end{bmatrix} \quad (6)$$

where $\mathbf{e}_z = [0 \ 0 \ 1]^T$ is the unit vector of Z axis.

The EKF state estimate and error covariance are propagated as in [14]:

$$\hat{\mathbf{x}}_k^{e-} = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^e, \mathbf{u}_k^v, 0) \quad (7)$$

$$\mathbf{P}_k^- = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{L}_k \mathbf{Q}_k \mathbf{L}_k^T \quad (8)$$

B. Measurement model

The filter measurement model for GPS antenna position $\mathbf{p}_{GA_l}^g$ and velocity $\mathbf{v}_{GA_l}^g$ is:

$$\mathbf{z}_l^g = \mathbf{h}(\mathbf{x}_l^e, \mathbf{m}_l) \Leftrightarrow \begin{bmatrix} \mathbf{p}_{GA_l}^g \\ \mathbf{v}_{GA_l}^g \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{GW}(\mathbf{R}_{WI_l} \mathbf{p}_{IA} + \mathbf{p}_{WI_l}) + \mathbf{p}_{GW} + \mathbf{m}_{pl} \\ \mathbf{R}_{GW}(\mathbf{R}_{WI_l} [\mathbf{I}_l \boldsymbol{\omega}_l^v]^{\times} \mathbf{p}_{IA} + \mathbf{v}_{WI_l}) + \mathbf{m}_{vl} \end{bmatrix} \quad (9)$$

where the measurement noise vector $\mathbf{m} = [\mathbf{m}_p^T, \mathbf{m}_v^T]^T$ contains Gaussian white noise processes with covariance $\mathbf{R} = \mathbf{E}\{\mathbf{m}\mathbf{m}^T\} = (\sigma^g)^2 \mathbf{I}_6$. The transformation between the local and global reference frames, expressed by $\mathbf{p}_{GW}$ and $\mathbf{R}_{GW}$, is obtained by aligning the beginning part of VIO and GPS trajectories using the *posyaw* method in [15]. The position offset $\mathbf{p}_{IA}$ is measured by hand on the real sensor setup. The angular velocity $\mathbf{I}_l \boldsymbol{\omega}^v$ is continuously provided by the VIO system.

Linearizing (9) around the propagated state estimate $\hat{\mathbf{x}}_l^{e-}$, we get the measurement and measurement noise Jacobian matrices as:

$$\mathbf{H}_l = \begin{bmatrix} \mathbf{R}_{GW} & 0 & \mathbf{R}_{GW}[\mathbf{e}_z]^T \hat{\mathbf{R}}_{WI_l}^- \mathbf{p}_{IA} \\ 0 & \mathbf{R}_{GW} & \mathbf{R}_{GW}[\mathbf{e}_z]^T \hat{\mathbf{R}}_{WI_l}^- [\mathbf{I}_l \boldsymbol{\omega}_l^v]^{\times} \mathbf{p}_{IA} \end{bmatrix} \quad (10)$$

$$\mathbf{M}_l = \mathbf{I}_6 \quad (11)$$

The Kalman gain along with the correction of state estimate and error covariance are derived as in [14]:

$$\mathbf{K}_l = \mathbf{P}_l^- \mathbf{H}_l^T (\mathbf{H}_l \mathbf{P}_l^- \mathbf{H}_l^T + \mathbf{M}_l \mathbf{R}_l \mathbf{M}_l^T)^{-1} \quad (12)$$

$$\hat{\mathbf{x}}_l^e = \hat{\mathbf{x}}_l^{e-} + \mathbf{K}_l (\mathbf{z}_l^g - \mathbf{h}(\hat{\mathbf{x}}_l^{e-}, 0)) \quad (13)$$

$$\mathbf{P}_l = (\mathbf{I} - \mathbf{K}_l \mathbf{H}_l) \mathbf{P}_l^- \quad (14)$$

V. EXPERIMENTS

A. Experimental setup

To evaluate the proposed algorithm, we developed a sensor unit as shown in Fig. 4. The hardware consisted of a Point Grey Bumblebee2 stereo camera with a 12 cm baseline, an ADIS16488 MEMS IMU, a U-blox M8N GPS receiver, and

an embedded computer. The stereo camera captured pairs of grayscale images of 640×480 pixels at a rate of 20 Hz. The IMU read inertial measurements at 500 Hz with an angular random walk of 0.3 deg/ $\sqrt{\text{hr}}$ and a velocity random walk of 0.029 m/s/ $\sqrt{\text{hr}}$. The GPS receiver took global measurements at 5 Hz with an accuracy of 2.5 m for horizontal position and 0.05 m/s for velocity. The camera and IMU were hardware synchronized and closely mounted on an aluminum plate to keep the calibration parameters unchanged. The embedded computer was equipped with an Intel i7-2760QM CPU to process the system software in real time. The VIO and EKF algorithms were purely programmed in C++ for high-speed execution. The device drivers were written on the Robot Operating System (ROS) for efficient data management.

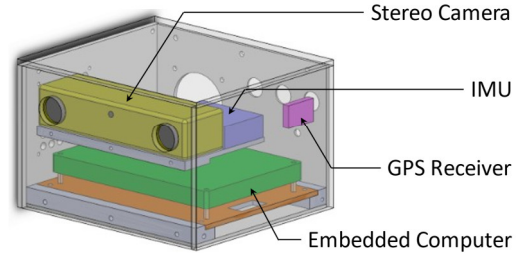


Fig. 3. Visualization of the VIO-GPS sensor.

To record evaluation datasets, we attached the sensor onto a test cart with a height of 1.5 m similar to a normal car. Its roof was tilted 10 degrees to reduce the featureless sky area in the camera view. The GPS antenna was placed 1 m away from the IMU to ensure good observability of the yaw angle. Ground-truth data were provided by a South S82 GPS-RTK system with an accuracy of 1 cm for 3D position, and a Novatel SPAN-CPT GPS/INS system made from fiberoptic gyroscopes with an accuracy of 0.02 degrees for roll and pitch angles and 0.06 degrees for yaw angle.

B. Evaluation procedure

To properly calculate the error of a localization method, we need to convert its output to the same coordinates as the ground truth. Let $\mathcal{F}_R$ and $\mathcal{F}_S$ denote the body frames of RTK rover and SPAN sensor, the ground-truth position and orientation are given as $\mathbf{p}_{GR}$ and $\mathbf{R}_{GS}$. The pose estimate of VIO-GPS fusion, $\hat{\mathbf{p}}_{WI}$ and $\hat{\mathbf{R}}_{WI}$, can be transformed to the ground-truth coordinates using the local-global alignment and sensor offsets as:

$$\hat{\mathbf{p}}_{GR} = \mathbf{R}_{GW}(\hat{\mathbf{R}}_{WI} \mathbf{p}_{IR} + \hat{\mathbf{p}}_{WI}) + \mathbf{p}_{GW} \quad (15)$$

$$\hat{\mathbf{R}}_{GS} = \mathbf{R}_{GW} \hat{\mathbf{R}}_{GI} \mathbf{R}_{IS} \quad (16)$$

In our real setup, the position offset between IMU and RTK rover is $\mathbf{p}_{IR} = [0.018, -0.001, 0.247]^T$. The rotation offset between IMU and SPAN is $\mathbf{R}_{IS} = \mathbf{I}$ as the two sensors were installed with the same orientation.

While the ground-truth form of VIO estimation is obtained in the same way as (15) and (16), the GPS measurement is transformed by:

$$\mathbf{p}_{GR}^g = \mathbf{R}_{GA} \mathbf{p}_{AR} + \mathbf{p}_{GA}^g \quad (17)$$

where $\mathbf{R}_{GA} = \mathbf{R}_{GS}$ because we set $\mathbf{R}_{SA} = \mathbf{I}$, and the offset $\mathbf{p}_{AR} = [-0.975, 0.152, -0.005]^T$.

From those, we can derive the position and orientation errors of each localization method compared to ground truth as:

$$\Delta \mathbf{p} = \hat{\mathbf{p}}_{GR} - \mathbf{p}_{GR} \quad (18)$$

$$\Delta \mathbf{R} = \hat{\mathbf{R}}_{GS} \mathbf{R}_{GS}^T \quad (19)$$

where $\Delta \mathbf{R}$ is then decomposed to roll, pitch and yaw errors. To evaluate the overall error on a dataset, we compute the root mean square error (RMSE) in each axis as:

$$\text{RMSE}(\hat{x}) = \sqrt{\frac{\sum_{k=1}^N (\Delta x_k)^2}{N}} \quad (20)$$

C. Experimental results

The sensor performance was examined in an outdoor setting inside the HCMUT campus. Two datasets with opposed trajectory shapes were recorded. In the Circle dataset, the vehicle moved smoothly on 5 circular loops with a diameter of 20 m at 0.5 m/s. After each loop, the vehicle was pulled back in a straight line to the starting point to examine the repeatability of position and yaw estimates. In the Square dataset, the vehicle went on 6 square loops with 4 straight edges of 20 m and 90-degree corners at a speed of 0.7 m/s. In both datasets, small motions were conducted around the starting point after each loop to demonstrate the local estimation precision. Fig. 4. illustrates the test condition and trajectory of the Circle dataset. The full experiments can be seen at [16].

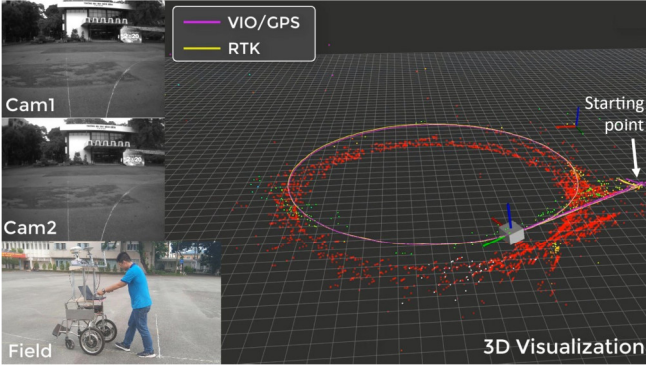


Fig. 4 Test condition and trajectory of the Circle dataset: The environment at HCMUT campus captured by the stereo camera (Cam1 and Cam2). The test vehicle driven by a person on a 20 m diameter circle (Field). The estimated trajectory and ground truth along with sensor pose and visual landmarks displayed on 3D graphic (3D Visualization).

The localization results of the Circle and Square datasets are shown in Fig. 5 and Fig. 6. In both cases, the VIO estimations are initially precise but gradually accumulates drifts in position and yaw. The GPS measurements have bounded errors but with large fluctuations compared to the ground truth. Overcoming those drawbacks, the VIO-GPS integration produces a globally accurate and drift-free pose estimation. Note that while GPS does not provide orientation measurements, our algorithm still successfully eliminates the drift in yaw angle, which is crucial for robot heading control. No loop closure method was used. The VIO roll and pitch estimates are already drift-free and thus do not need to be updated by GPS.

The quantitative localization errors are shown on Table I. Compared to separated localization methods, our VIO-GPS fusion achieve the best accuracy of about 0.4 m in 3D position and 1 degree in 3D orientation in both datasets.

VI. CONCLUSION

In this work, we presented a fusion method of visual-inertial system and GPS for globally drift-free and accurate pose estimation. An EKF-based algorithm was developed to effectively combine the relative estimation of VIO with the absolute measurement of GPS. A real sensor setup was developed with sufficient computing power to process the system algorithms in real time. Real-world experiments showed that the sensor achieve consistent and accurate pose estimation with different motion trajectories. The developed sensor can be used in a variety of autonomous mobile robots. Further research is necessary to improve the accuracy of position estimate and cope with sensor degradation scenarios.

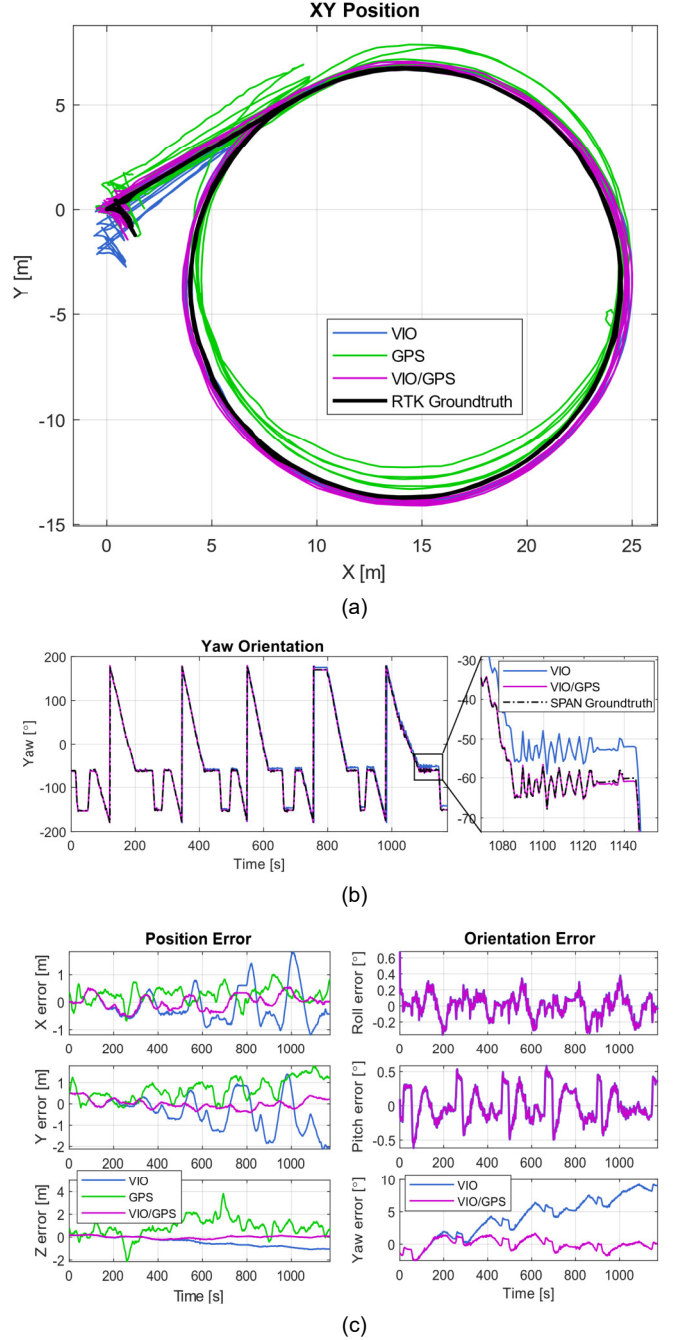


Fig. 5. Evaluation results of the Circle dataset. (a) Top-view trajectory. (b) Yaw angle estimates. (c) 3D position and 3D orientation errors.

TABLE I. LOCALIZATION ERROR ON OUTDOOR DATASETS

Dataset	Localization Method	Position RMSE [m]				Orientation RMSE [°]			
		X	Y	Z	3D	Roll	Pitch	Yaw	3D
Circle	VIO	0.62	0.88	0.60	1.23	0.13	0.23	5.15	5.16
	GPS	0.37	0.80	1.17	1.46	-	-	-	-
	VIO-GPS	0.26	0.22	0.11	0.36	0.13	0.23	0.88	0.92
Square	VIO	0.83	1.01	0.39	1.37	0.20	0.23	4.79	4.80
	GPS	0.45	0.68	0.92	1.23	-	-	-	-
	VIO-GPS	0.32	0.24	0.13	0.42	0.20	0.23	1.06	1.11

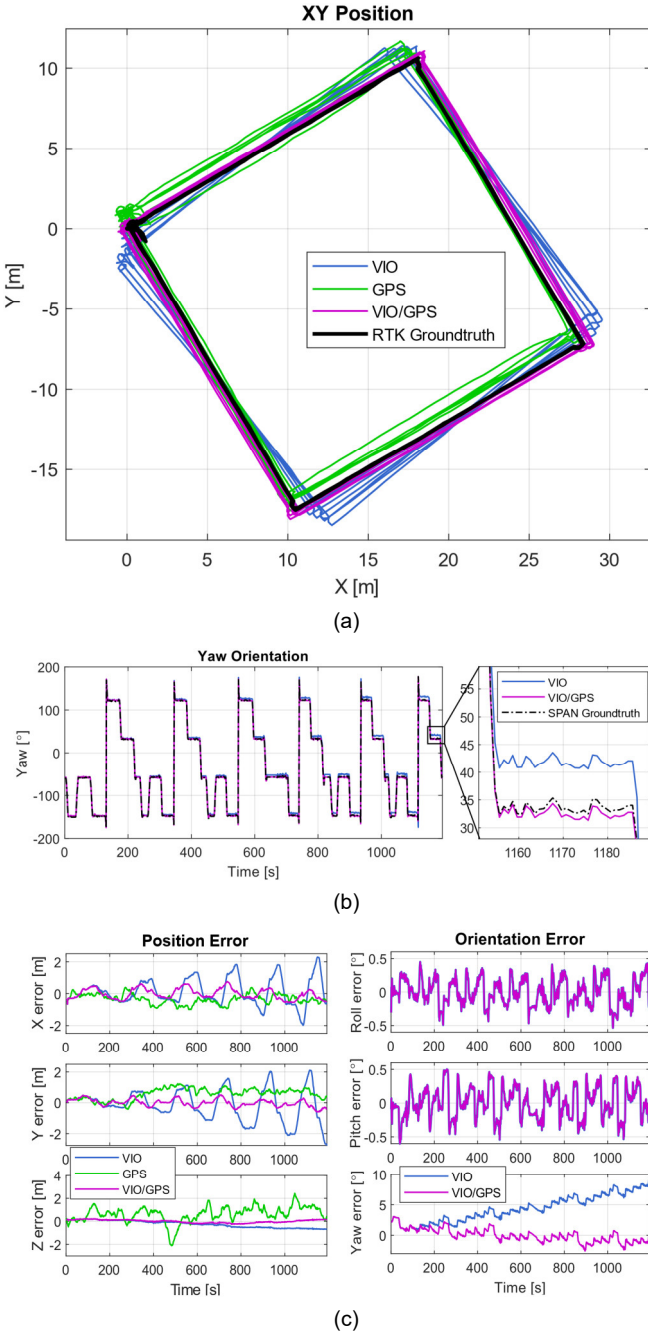


Fig. 6. Evaluation results of the Square dataset. (a) Top-view trajectory. (b) Yaw angle estimates. (c) 3D position and 3D orientation errors.

ACKNOWLEDGMENT

We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

REFERENCES

- [1] 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*, 2018.
- [2] S. A. S. Mohamed, M. -H. Haghbayan, T. Westerlund, J. Heikkonen, H. Tenhunen and J. Plosila, "A Survey on Odometry for Autonomous Navigation Systems," *IEEE Access*, vol. 7, pp. 97466-97486, 2019.
- [3] T. Qin, S. Cao, J. Pan and S. Shen, "A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors," in *arXiv preprint arXiv:1901.03642*, 2019.
- [4] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart and M. Chli, "GOMSF: Graph-Optimization Based Multi-Sensor Fusion for robust UAV Pose estimation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [5] G. Cioffi and D. Scaramuzza, "Tightly-coupled Fusion of Global Positional Measurements in Optimization-based Visual-Inertial Odometry," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [6] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to MAV navigation," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.
- [7] A. Chambers, S. Scherer, L. Yoder, S. Jain, S. Nuske and S. Singh, "Robust multi-sensor fusion for micro aerial vehicle navigation in GPS-degraded/denied environments," in *2014 American Control Conference*, 2014.
- [8] W. Lee, K. Ekenhoff, P. Geneva and G. Huang, "Intermittent GPS-aided VIO: Online Initialization and Calibration," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [9] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *2007 IEEE International Conference on Robotics and Automation*, 2007.
- [10] Nguyen H. K. Tran and Vinh-Hao Nguyen, "A Tightly-Coupled Visual-Inertial System with Synchronized Time for Indoor Localization," *International Journal of Mechanical Engineering and Robotics Research*, vol. 10, no. 4, pp. 202-208, 2021.
- [11] P. Furgale, J. Rehder and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, 2013.
- [12] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 4, pp. 314-334, 2014.
- [13] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli and R. Siegwart, "Real-time Onboard Visual-Inertial State Estimation and Self Calibration of MAVs in Unknown Environments," in *2012 IEEE International Conference on Robotics and Automation*, Saint Paul, MN, USA, 2012.
- [14] Dan Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, 2006.
- [15] 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)*, Madrid, Spain, 2018.
- [16] [Online]. Available: <https://youtu.be/aT3N4ZHg42g>.