

Sensor applications

Visual-Gyroscope-Wheel Odometry With Ground Plane Constraint for Indoor Robots in Dynamic Environment

Ming Ouyang^{1,2}, Zhiqiang Cao^{1,2*}, Peiyu Guan^{1,2}, Zhonghui Li³, Chao Zhou^{1,2}, and Junzhi Yu^{4**}

¹State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

²School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

³Beijing Engo Technology Company Ltd., Beijing 101318, China

⁴Department of Mechanics and Engineering Science, BIC-ESAT, College of Engineering, Peking University, Beijing 100871, China

*Senior Member, IEEE

**Fellow, IEEE

Manuscript received January 21, 2021; accepted January 31, 2021. Date of publication February 4, 2021; date of current version March 8, 2021.

Abstract—Localization is an essential prerequisite for autonomous robots to execute tasks. In this letter, a multisensor fused odometry is proposed for the robot applied to dynamic indoor environments. The gyroscope and wheel encoder information is integrated into visual odometry, where the robot pose is parameterized with 3-D position and 3-D Euler angle to relieve the estimation drift during planar motion and conveniently fuse wheel encoder measurements. According to the results of semantic segmentation, the dynamic elements are removed. On this basis, ground plane constraint is exerted on the optimization procedure. As a result, the estimation accuracy is improved. The experimental results on the OpenLORIS-scene dataset demonstrate the effectiveness of the proposed odometry.

Index Terms—Sensor applications, dynamic environment, indoor robot, semantics, visual-gyroscope-wheel odometry.

I. INTRODUCTION

Visual odometry (VO) utilizes visual information to localize the robot [1]. Depending on the utilized visual information, VO can be divided into direct and feature-based methods. The former utilizes all the pixel information in an image to minimize the photometric error [2]. For the latter, it needs to first extract visual features and then track these features, where the cost function considers reprojection error. ORB-SLAM2 is a representative feature-based method [3].

In practice, visual information can be degenerated under poor illumination conditions, which possibly makes VO become invalid. The researchers turn to the fusion with other information including inertial measurement unit (IMU), wheel encoder and semantics. In [4], wheel encoder information is fused in a graph optimization-based visual inertial odometry to enhance the pose estimation. Qin *et al.* [5] proposed a monocular visual-inertial system (VINS-Mono), which extracts visual features and IMU preintegration information for pose estimation with graph optimization. For dynamic environments, Zhao *et al.* [6] proposed a semantic visual-inertial simultaneous localization and mapping (SLAM) by adding segmentation on VINS-Mono. In [7], a dynamic semantic SLAM (DS-SLAM) is proposed where semantic segmentation network with moving consistency check method are combined to reduce the impact of dynamic objects. Li *et al.* [8] combined the inertial measurements with dynamic semantics to build a visual inertial semantic SLAM for unmanned aerial vehicle running in underground tunnel. In [9], a multisensor fusion scheme is presented where wheel encoder measurements are fused with semantics; however, only vision information is involved in the pose optimization. Overall, multisensor and semantics fusion provides a

preferable way to enhance the accuracy and robustness of robot odometry in dynamic environments. In this letter, with the combination of vision, gyroscope, wheel encoder, and semantics, a multisensor fused odometry is proposed for indoor robots. The wheel encoder exerts additional preintegration constraint between consecutive keyframes, which improves the processing ability when there exists scene with few visual features. Furthermore, the robot pose is parameterized by 3-D position and 3-D Euler angle instead of widely used special Euclidean group $SE(3)$ [4], [5], [7]. On one hand, the preintegration constraint of wheel encoder is conveniently imposed. On the other hand, the pose drift can be reduced. Also, the ground plane constraint is introduced and the estimation accuracy is further improved.

II. VISUAL-GYROSCOPE-WHEEL ODOMETRY

In this letter, the information provided by camera, gyroscope, and wheel encoder is integrated. The flow chart of the proposed odometry termed as VGWO is given in Fig. 1, which is composed of sensors measurements, front end, and back end. In the front end, all sensor information is processed to obtain features and constraints between two consecutive keyframes, which are sent to the back end for pose estimation. The camera is used to extract visual features of the environment for feature tracking. The wheel encoder is adopted to enhance the short-term pose estimation quality, especially in the case where visual localization is weak. Its measurements related to robot velocity are integrated between two consecutive keyframes to give a relative pose constraint. Besides, considering that the rotation speed calculated by the wheel encoder is perhaps inaccurate, the gyroscope is employed to provide rotation measurements, which generates a relative rotation constraint. We label the robot motion speed as $V(v_k^b, w_k^b)$, where v_k^b and w_k^b refer to the linear speed and rotation speed, respectively. The sensors information is processed in the front end to track the

Corresponding author: Zhiqiang Cao (e-mail: zhiqiang.cao@ia.ac.cn).

Associate Editor: F. Costa.

Digital Object Identifier 10.1109/LSENS.2021.3057088

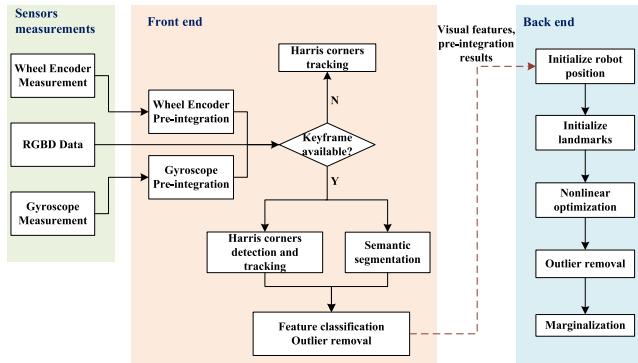


Fig. 1. Flow chart of the visual-gyroscope-wheel odometry VGWO.

visual features and generate the relative pose constraint between two consecutive keyframes for the back end, which utilizes the graph optimization to estimate the robot pose. To reduce the computation burden of the back end, only keyframes are considered. Note that both wheel encoder and gyroscope are inefficient for long-term pose estimation. In this letter, a preintegration procedure [10] is applied. On this basis, current frame is judged as a keyframe when the relative pose between it and last keyframe is large.

We extract Harris corner [11] as the image feature and then utilize Kanade–Lucas–Tomasi optical flow tracker [12] to track the features. To make the odometry be applied to dynamic environments, image semantic segmentation is also implemented to label the environment. Feature extraction and semantic segmentation only run in the keyframes with two parallel threads.

In the back end, the image features from the front end are classified into background, dynamic, and ground features according to semantic labels. Dynamic features do not involve in the optimization of the back end. The remaining features are used to create 3-D landmarks by associating these features with corresponding depth measurement from the depth image or triangulation. With the combination of initial pose estimation that is obtained according to the robot kinematic model and preintegrations, a nonlinear optimization problem is given. This problem is iteratively solved using Levenberg–Marquardt algorithm [13], and then, we acquire the estimated robot pose. Also, the positions of landmarks are updated. In addition, sliding window is used to bound the number of keyframes where marginalization is adopted to drop the oldest keyframe while keeping its prior information [5].

A. Full State Vector

The full state vector of the robot environment with n keyframes and m landmarks in a sliding window is given by

$$\begin{aligned} \chi &= [b_0^w, b_1^w, \dots, b_{n-1}^w, l_0^w, l_1^w, \dots, l_{m-1}^w, T_{bc}, R_{bg}, G^w] \\ b_k^w &= [x_k^w, y_k^w, \theta_k^w, \delta z_k^w, \delta \beta_k^w, \delta \varphi_k^w] \\ l_i^w &= [l_{i_x}^w, l_{i_y}^w, l_{i_z}^w], \quad T_{bc} = [R_{bc}, t_{bc}], \quad G^w = [\alpha, \tau, d] \end{aligned} \quad (1)$$

where $b_k^w (k = 0, 1, \dots, n-1)$ refers to the robot pose related to k th keyframe in the world coordinate frame $O_w X_w Y_w Z_w$. $O_w X_w Y_w Z_w$ is set to the robot body frame at the first keyframe, which means that this letter concerns a relative pose estimation of the robot. b_k^w contains 3-D position $(x_k^w, y_k^w, \delta z_k^w)$ and 3-D Euler angle $(\theta_k^w, \delta \beta_k^w, \delta \varphi_k^w)$. Herein, θ , β , φ represent the yaw, roll, pitch angles, respectively, and the prefix δ means that the value of corresponding variable is near 0. l_i^w denotes the observed i th landmark position in $O_w X_w Y_w Z_w$. T_{bc} refers to the extrinsic parameter matrix from $O_b X_b Y_b Z_b$ to $O_c X_c Y_c Z_c$,

where $O_b X_b Y_b Z_b$ refers to the robot body coordinate frame attached to the wheel odometry with X_b -axis being in coincidence with the moving direction of the robot, and $O_c X_c Y_c Z_c$ is labeled as the camera coordinate system with its origin O_c and Z_c -axis being the optical center and the optical axis of the camera, respectively. R_{bg} denotes the rotation transform from $O_b X_b Y_b Z_b$ to $O_g X_g Y_g Z_g$, which is the gyroscope coordinate frame with its origin O_g being its center. G^w is the ground plane coefficients and the plane equation is $\cos(\alpha) \cos(\tau)x + \cos(\alpha) \sin(\tau)y + \sin(\alpha)z + d = 0$. For b_k^w , its transformation matrix T_{bk}^w is described as follows:

$$T_{bk}^w = T(x_k^w, y_k^w, \theta_k^w) * T(\delta z_k^w) * T(\delta \beta_k^w) * T(\delta \varphi_k^w) \quad (2)$$

where

$$\begin{aligned} T(x_k^w, y_k^w, \theta_k^w) &= \begin{bmatrix} \cos(\theta_k^w) & -\sin(\theta_k^w) & 0 & x_k^w \\ \sin(\theta_k^w) & \cos(\theta_k^w) & 0 & y_k^w \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \\ T(\delta z_k^w) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \delta z_k^w \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ T(\delta \beta_k^w) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\delta \beta_k^w) & -\sin(\delta \beta_k^w) & 0 \\ 0 & \sin(\delta \beta_k^w) & \cos(\delta \beta_k^w) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \\ T(\delta \varphi_k^w) &= \begin{bmatrix} \cos(\delta \varphi_k^w) & 0 & \sin(\delta \varphi_k^w) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\delta \varphi_k^w) & 0 & \cos(\delta \varphi_k^w) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

B. Preintegration

As mentioned previously, both gyroscope and wheel encoder require preintegration procedures. Between two consecutive keyframes k_1 and k_2 , the gyroscope can provide multiple measurements. By adopting the method in [10], these gyroscope measurements are integrated to get the relative rotation transformation $\Delta R_{k_1 k_2}$ from k_1 to k_2 , as well as preintegration measurement covariance $\Sigma_{G_{k_1 k_2}}$.

For wheel encoder, we focus on the former three states x_k^w , y_k^w , and θ_k^w . Let γ_k^w denote $[x_k^w, y_k^w, \theta_k^w]^T$. The discrete dynamic motion formulation within Δt_o is given by

$$\begin{bmatrix} x_{t_2}^w \\ y_{t_2}^w \\ \theta_{t_2}^w \end{bmatrix} = \begin{bmatrix} x_{t_1}^w \\ y_{t_1}^w \\ \theta_{t_1}^w \end{bmatrix} + \begin{bmatrix} \cos(\theta_{t_1}^w) v_{t_1}^b \\ \sin(\theta_{t_1}^w) v_{t_1}^b \\ w_{t_1}^b \end{bmatrix} * \Delta t_o \quad (3)$$

where $\Delta t_o = t_2 - t_1$. For two consecutive keyframes k_1 and k_2 , there are N encoder measurements and we can propagate the state using Euler integral as

$$\gamma_{k_2}^w = \gamma_{k_1}^w + \sum_{i=0}^{N-1} \begin{bmatrix} \cos(\theta_{t_i}^w) v_{t_i}^b \\ \sin(\theta_{t_i}^w) v_{t_i}^b \\ w_{t_i}^b \end{bmatrix} * \Delta t_o. \quad (4)$$

Preintegration of wheel encoder refers to the relative pose $\gamma_{k_2}^{k_1}$ between keyframes k_1 and k_2

$$\gamma_{k_2}^{k_1} = \begin{bmatrix} R(-\theta_{k_1}^w) & 0 \\ 0 & 1 \end{bmatrix} (\gamma_{k_2}^w - \gamma_{k_1}^w) = \sum_{i=0}^{N-1} \begin{bmatrix} \cos(\theta_i^w - \theta_{k_1}^w) v_i^b \\ \sin(\theta_i^w - \theta_{k_1}^w) v_i^b \\ w_i^b \end{bmatrix} \Delta t_o \quad (5)$$

where $R(\theta_k^w) = \begin{bmatrix} \cos(\theta_k^w) & -\sin(\theta_k^w) \\ \sin(\theta_k^w) & \cos(\theta_k^w) \end{bmatrix}$.

The preintegration measurement covariance $\Sigma_{O_{k_1 k_2}}$ can be deduced by using covariance propagation method in [10].

C. Semantic Segmentation

BiseNet [14] is adopted for semantic segmentation. It is responsible for detecting dynamic objects and ground plane so that each Harris corner feature is attached a semantic label. Also, it is executed in parallel with feature detection and tracking.

D. Optimization in Back End

A multisensor bundle adjustment [15] formulation is used to describe the nonlinear optimization problem, where the sum of the Mahalanobis norm [16] of all measurement residuals is minimized to obtain maximum posteriori estimation

$$\min_{\chi} \left\{ \sum_{k=0}^{n-2} \|r_{O_{k,k+1}}\|_{\Sigma_{O_{k,k+1}}^{-1}}^2 + \sum_{k=0}^{n-2} \|r_{G_{k,k+1}}\|_{\Sigma_{G_{k,k+1}}^{-1}}^2 \right. \\ \left. + \sum_{k=0}^{n-1} \sum_{i=0}^{m-1} \|r_{C_{ki}}\|_{\Sigma_C^{-1}}^2 + \sum_{j=0}^{s-1} \|r_{P_j}\|^2 \right\} \quad (6)$$

where $r_{O_{k,k+1}}$, $r_{G_{k,k+1}}$, $r_{C_{ki}}$, and r_{P_j} are residuals for the wheel encoder, gyroscope, vision, and ground feature constraint, respectively. $\Sigma_{O_{k,k+1}}^{-1}$, $\Sigma_{G_{k,k+1}}^{-1}$, and Σ_C^{-1} are the information matrices, which are the inverse of the covariance matrices. s is the number of ground landmarks. After solving this optimization problem, we can obtain the optimized poses of keyframes as well as the sparse map formed by landmarks. Next, the aforementioned four residuals are given.

For wheel encoder, the residual $r_{O_{k_1 k_2}}$ between two consecutive keyframes k_1 and k_2 is

$$r_{O_{k_1 k_2}} = \gamma_{k_2}^{k_1} - \begin{bmatrix} R(-\theta_{k_1}^w) & 0 \\ 0 & 1 \end{bmatrix} (\gamma_{k_2}^w - \gamma_{k_1}^w) \quad (7)$$

With $\Delta R_{k_1 k_2}$, the gyroscope residual $r_{G_{k_1 k_2}}$ can be obtained by [10].

For i th landmark whose position is l_i^w , $i = 0, 1, \dots, m-1$, given the robot position b_k^w and extrinsic parameter matrix T_{bc} , we can get its position in the camera coordinate frame $O_c X_c Y_c Z_c$

$$l_{ki}^c = \begin{bmatrix} l_{ki_x}^c & l_{ki_y}^c & l_{ki_z}^c \end{bmatrix}^T = (T_{bk}^w * T_{bc})^{-1} * l_i^w. \quad (8)$$

Then, its normalized coordinate is $l_{ki}^{nc} = [\frac{l_{ki_x}^c}{l_{ki_z}^c} \quad \frac{l_{ki_y}^c}{l_{ki_z}^c}]^T$. Meanwhile, the corresponding measured position \bar{l}_{ki}^{nc} is obtained, and then the residual $r_{C_{ki}}$ is given by

$$r_{C_{ki}} = l_{ki}^{nc} - \bar{l}_{ki}^{nc}. \quad (9)$$

Finally, for each ground landmark l_j^w , $j = 0, 1, \dots, s-1$, the corresponding ground feature constraint residual is described as

$$r_{P_j} = (\cos(\alpha) \cos(\tau) l_{j_x}^w + \cos(\alpha) \sin(\tau) l_{j_y}^w + \sin(\alpha) l_{j_z}^w + d). \quad (10)$$

Table 1. Accuracy Comparison of Different Methods Without Semantics in Terms of APE and RPE in Meters.

| Sequence | ORB-SLAM2 | | | OpenVSLAM | | VINS-Mono | | VGWO-I | |
|------------|-----------|------|-------|-----------|-------|-----------|-------|--------|-------|
| | EL/TL | APE | RPE | APE | RPE | APE | RPE | APE | RPE |
| market 1-1 | 47.05/145 | 1.32 | 0.077 | 2.53 | 0.095 | 1.67 | 0.192 | 1.04 | 0.013 |
| market 1-2 | 46.43/215 | 0.72 | 0.069 | 3.70 | 0.121 | 1.99 | 0.187 | 1.13 | 0.014 |
| market 1-3 | 99.10/224 | 2.11 | 0.074 | 2.80 | 0.099 | 1.56 | 0.485 | 1.07 | 0.033 |

Table 2. Accuracy Comparison of Different Methods With Semantics in Terms of APE and RPE in Meters.

| Sequence | DS-SLAM | | | VGWO | | VGWO-II | |
|------------|------------|------|-------|------|-------|---------|-------|
| | EL/TL | APE | RPE | APE | RPE | APE | RPE |
| market 1-1 | 43.97/145 | 1.17 | 0.243 | 0.83 | 0.013 | 0.97 | 0.014 |
| market 1-2 | 120.31/215 | 3.16 | 0.284 | 1.00 | 0.010 | 1.08 | 0.011 |
| market 1-3 | 120.15/224 | 1.78 | 0.299 | 0.86 | 0.027 | 0.96 | 0.033 |

Table 3. Average Computational Time of VGWO in One Experiment.

| Sequence | Average time of the front end | | Average time of the back end | Number of keyframes /total frames |
|-----------|-------------------------------|--------------|------------------------------|-----------------------------------|
| | keyframe | non-keyframe | | |
| Market1-1 | 36 ms | 10 ms | 152 ms | 1201/6144 |
| Market1-2 | 35 ms | 10 ms | 148 ms | 1503/7660 |
| Market1-3 | 35 ms | 10 ms | 152 ms | 1616/8814 |

III. EXPERIMENTS

The experiments are conducted to demonstrate the proposed VGWO method on OpenLORIS-scene market dataset [17] with three sequences (market 1-1, market 1-2, and market 1-3), which is collected in a real supermarket with dynamic pedestrians by a mobile robot. The dataset provides red, green, blue, and depth (RGBD) images with 30 fps, gyroscope data with 400 Hz and wheel encoder data with 40 Hz. Partial images of market 1-1 were randomly selected as the training set for semantic segmentation.

In this letter, EVO [18] is adopted as the evaluation tool of different methods. For better demonstration, accuracy comparison is illustrated without or with semantics, and the results are given in Tables 1 and 2. Table 1 provides the results of our method without semantics (VGWO-I) with visual-inertial method VINS-Mono [5], in terms of absolute pose error (APE) and relative pose error (RPE). Also, the results of pure visual methods including ORB-SLAM2 (RGBD) [3] and OpenVSLAM (RGBD) [19] are presented to only demonstrate the general performance of multisensor scheme. ORB-SLAM2 fails to run on the whole trajectory of each sequence, and only the data before failure involves statistics. EL and TL refer to effective running length and total trajectory length, respectively. One can see from Table 1 that multisensor methods can generally improve the accuracy and our method performs better. Table 2 describes the accuracy comparison of the methods with semantics including DS-SLAM (RGBD) [7], VGWO, and VGWO-II, where VGWO-II does not consider ground plane constraint. Note that DS-SLAM fails to run on the whole trajectory of each sequence, which may be due to the influence of dynamic factors with a large proportion. In contrast, our method always works. Besides, compared to VGWO-I and VGWO-II, VGWO with dynamic semantics and ground plane constraint achieves the best performance.

Taking the sequence market 1-3 as an example, the trajectories of different methods are depicted in Fig. 2. It is seen that VGWO achieves results closer to the Ground truth. Fig. 3 presents the comparison of position accuracy of VGWO and VINS-Mono on the sequence market 1-1. VGWO with parameterized pose of the 3-D position and the 3-D Euler angle as well as ground plane constraint achieves higher position accuracy. Table 3 presents the average computational time of VGWO in one experiment. With the combination of image frame rate of 30 fps, one can obtain that the average time interval between consecutive keyframes is longer than that of the back end, which means that the back end of VGWO has time to execute the pose estimation.

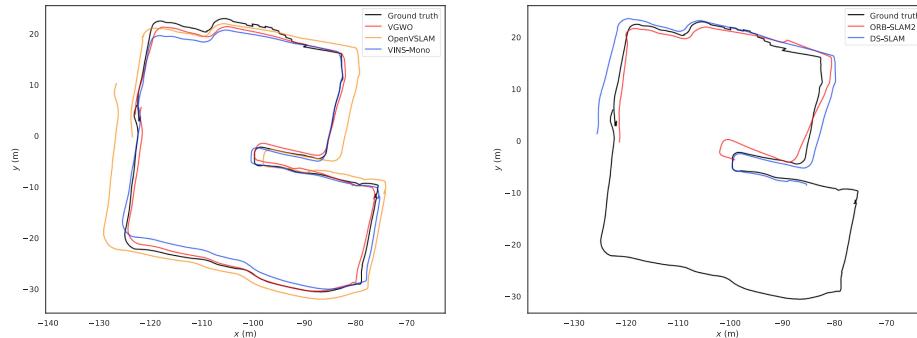


Fig. 2. Trajectories of different methods on the sequence market 1–3 of OpenLORIS-scene dataset.

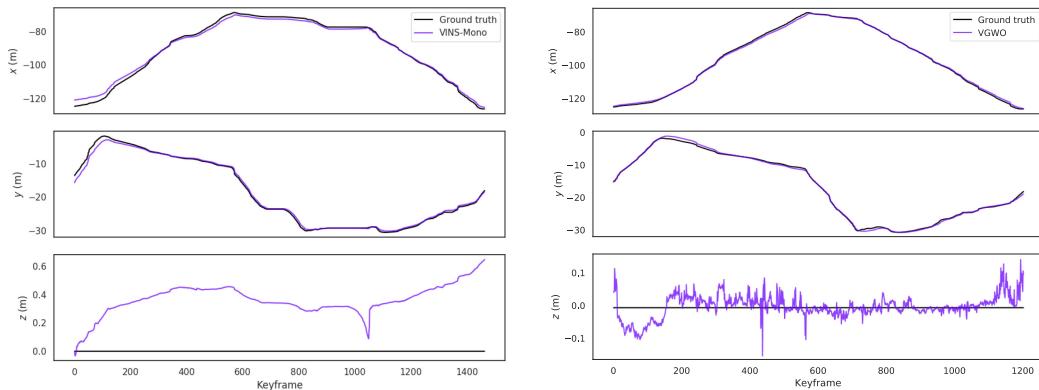


Fig. 3. Comparison of position accuracy of VGWO and VINS-Mono on the sequence market 1–1 of OpenLORIS-scene dataset.

IV. CONCLUSION

In this letter, a multisensor odometry for indoor mobile robots is proposed, where wheel encoder, gyroscope, and vision information are integrated. Robot pose parameterization with the 3-D Euler angle and the 3-D position instead of $SE(3)$ can reduce the estimation drift during planar motion. With the ground plane constraint, graph optimization can provide better estimation. In our near future, the proposed method shall be further verified on office and canteen environments with the combination of navigation.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61633020, Grant 62073322, and Grant 61836015.

REFERENCES

- [1] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, “An overview to visual odometry and visual SLAM: Applications to mobile robotics,” *Intell. Ind. Syst.*, vol. 1, no. 4, pp. 289–311, 2015.
- [2] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, Mar. 2018.
- [3] R. Mur-Artal and J. D. Tardós, “ORB-SLAM2: An open-source SLAM system for monocular stereo, and RGB-D cameras,” *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.
- [4] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, “VINS wheels,” in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 5155–5162.
- [5] T. Qin, P. L. Li, and S. J. Shen, “VINS-Mono: A robust and versatile monocular visual-inertial state estimator,” *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.
- [6] X. Zhao, C. Wang, and M. H. Ang, “Real-Time visual-inertial localization using semantic segmentation towards dynamic environments,” *IEEE Access*, vol. 8, pp. 155047–155059, 2020.
- [7] C. Yu *et al.*, “DS-SLAM: A semantic visual SLAM towards dynamic environments,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 1168–1174.
- [8] D. J. Li *et al.*, “A visual-inertial localization method for unmanned aerial vehicle in underground tunnel dynamic environments,” *IEEE Access*, vol. 8, pp. 76809–76822, 2020.
- [9] H. Xu, C. Yang, and Z. Li, “OD-SLAM: Real-time localization and mapping in dynamic environment through multi-sensor fusion,” in *Proc. Int. Conf. Adv. Robot. Mechatronics*, 2020, pp. 172–177.
- [10] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, “On-Manifold preintegration for real-time visual-inertial odometry,” *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, Feb. 2017.
- [11] J. B. Shi and C. Tomasi, “Good features to track,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 593–600.
- [12] C. Tomasi and T. Kanade, “Detection and tracking of point features,” Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, 1991.
- [13] D. Marquardt, “An algorithm for least squares estimation on nonlinear parameters,” *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.
- [14] C. Q. Yu *et al.*, “BiSeNet: Bilateral segmentation network for Real-time semantic segmentation,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 334–349.
- [15] B. Triggs *et al.*, “Bundle adjustment—A modern synthesis,” in *Proc. Int. Workshop Vis. Algorithms*, 1999, pp. 298–372.
- [16] P. C. Mahalanobis, “On the generalised distance in statistics,” *Proc. Nat. Inst. Sci. India*, vol. 2, no. 1, pp. 49–55, 1936.
- [17] X. S. Shi *et al.*, “Are we ready for service robots? The openloris-Scene datasets for lifelong SLAM,” in *Proc. IEEE Int. Conf. Robot. Autom.*, 2020, pp. 3139–3145.
- [18] EVO. [Online]. Available: <https://github.com/MichaelGrupp/evo.git>
- [19] S. Sumikura, M. Shibuya, and K. Sakurada, “OpenVSLAM: A versatile visual SLAM framework,” in *Proc. ACM Int. Conf. Multimedia*, 2019, pp. 2292–2295.