

360-VIO: A Robust Visual–Inertial Odometry Using a 360° Camera

Qi Wu *Student Member, IEEE*, Xiangyu Xu *, Xieyuanli Chen *, Ling Pei *, Senior Member, IEEE*, Chao Long *, Junyuan Deng *, Guoqing Liu *, Sheng Yang *, Shilei Wen *, and Wenzian Yu *, Senior Member, IEEE********

Abstract—Visual–inertial odometry (VIO) is an important component for robots working in industrial environments to obtain accurate and robust pose estimation. In this article, we introduce a novel VIO algorithm designed specifically for 360° cameras, exploiting their wide field of view. We build our algorithm based on an EKF-based filtering framework and propose a novel measurement model for 360° camera leveraging the reprojection error on the tangent plane of the spherical surface. By this, our 360-VIO effectively mitigates errors caused by the image distortion and fully exploit the omnidirectional observation, resulting in superior accuracy and robustness toward illumination changes and fast camera movements. To the best of our knowledge, no public 360-based visual–inertial dataset is currently available. To address this gap, we have constructed a novel 360° camera dataset under different challenging environments to evaluate the performance of our proposed algorithm. The results of our experiments demonstrate the impressive robustness and accuracy of our method compared with other state-of-the-art VIO methods across diverse and challenging environments. To facilitate future research on 360-VIO, we will release the method, self-collected dataset and the calibration tool at: <https://github.com/Gatsby23/360-VIO>.

Index Terms—360° camera, robotics applications, visual odometry, visual–inertial odometry (VIO).

I. INTRODUCTION

THE estimation of the six-degree-of-freedom pose is a fundamental requirement for many applications,

Manuscript received 21 March 2023; revised 19 June 2023, 5 September 2023, and 24 October 2023; accepted 29 October 2023. Date of publication 25 December 2023; date of current version 5 June 2024. This work was supported by the National Nature Science Foundation of China (NSFC) under Grant 62273229. (*Corresponding author: Ling Pei.*)

Qi Wu, Ling Pei, Junyuan Deng, Guoqing Liu, and Wenzian Yu are with the Shanghai Jiaotong University, Shanghai 200240, China (e-mail: robotics_qi@sjtu.edu.cn; ling.pei@sjtu.edu.cn; d.junye@sjsu.edu.cn; guoqing_liu@sjtu.edu.cn; wxyu@sjtu.edu.cn).

Xiangyu Xu, Chao Long, and Shilei Wen are with the ByteDance, Beijing 100190, China (e-mail: xuxiangyu.fs@bytedance.com; longchao@bytedance.com; zhengmin.666@bytedance.com).

Xieyuanli Chen is with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: xieyuanli.chen@nudt.edu.cn).

Sheng Yang is with the Shanghai Jiaotong University, Shanghai 200240, China, and also with Autonomous Driving Dept, Cainiao Inc., Alibaba Group (e-mail: shengyang@alibaba-inc.com).

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/TIE.2023.3337541>.

Digital Object Identifier 10.1109/TIE.2023.3337541

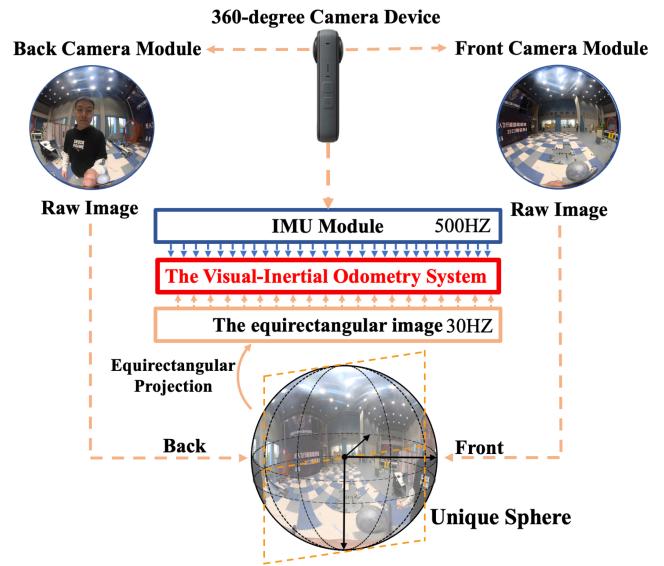


Fig. 1. Our proposed 360-VIO system exploits equirectangular images from the 360° camera device to do the odometry estimation.

including autonomous driving, unmanned aerial vehicles, and augmented/virtual reality. In GNSS-denied environments, visual–inertial odometry (VIO) has emerged as a favorable solution, considering the limitations of energy consumption and hardware costs of mobile systems [1], [2]. There are two types of VIO algorithms: optimization-based and filtering-based methods. Optimization-based methods perform global pose optimization to maintain long-term trajectory consistency, but incur a heavy computational cost due to postprocessing. Filtering-based methods, on the other hand, use visual constraints to correct IMU biases and maintain trajectory accuracy, making them more lightweight and suitable for online trajectory estimation in real-world applications. However, the current works in this research area are focused on the pinhole camera model, which has a limited sensing area with a narrow field of view (FOV).

Using 360° cameras is a potential solution to the problem mentioned previously. The definition of the 360° camera is a back-to-back attached stereo cameras with wide-angle lenses, allowing for a wide FOV in both the horizontal and vertical directions. The FOV of this type of camera is 360° horizontally and 180° vertically. The captured images can be combined into a single equirectangular image, which can be used for

visualization and some downstream tasks, such as odometry estimation. However, this wide FOV also leads to distortion problems, particularly for features located at the edge of the FOV. In addition, the back-to-back attached structure of the 360° camera does not allow for the recovery of metric information from the baseline. Although there are some omnidirectional VIO systems composed of multiple cameras with wide baselines, research on the 360° VIO (360-VIO) system is still lacking.

To bridge this gap, we propose a novel 360-VIO method utilizing the equirectangular image to extract visual features for triangulation and state estimation, while also incorporating a measurement update module to mitigate the effects of feature distortions, as shown in Fig. 1. Benefiting from the omnidirectional observations and the proposed measurement update module, our 360-VIO algorithm maintains robustness and accuracy even in the presence of illumination changes or fast camera movements. To the best of our knowledge, this is the first 360-VIO system that employs equirectangular images in conjunction with an IMU module. However, the lack of 360° dataset makes it challenging to evaluate the proposed VIO system thoroughly. To address this, we propose a new 360-VIO dataset with different settings for quality and quantity comparison. In addition to the dataset, we expanded the existing calibration tool [3] to support 360° camera model. Evaluations of the proposed method on our dataset reveal its accuracy and robustness over state-of-the-art VIO methods.

In sum, our key contributions are threefold as follows.

- 1) To the best of our knowledge, we are the first VIO that focuses on the 360° camera by utilizing the equidistant projection imagery. We fully leverage the wide-angle characteristics of commercial 360° cameras, ensuring comprehensive observation and tracking of the surrounding environment, thereby achieving a more robust VIO performance under different challenging environments.
- 2) Our study introduces a novel measurement model that offers a significant contribution to the 360° camera-based perception. We have modeled the uncertainty of the observed features based on the equidistant projection principle. By adopting this approach, we were able to mitigate the impact of feature distortions on the system according to the location of the features on the projected image and improve the robustness against illumination variations and rapid camera movements.
- 3) In order to promote the development of the community, we provide a novel 360-VIO dataset and will make it publicly available together with the implementation of our method. Furthermore, we also extend the calibration tool, which is the first calibration tool to support the extrinsic estimation between the 360° camera with the IMU. We believe that this work will facilitate the future research on the 360-VIO task.

II. RELATED WORK

In this section, we review works on the VIO algorithms, especially those with wider lenses. The seminal work on the VIO estimation was carried out by [4], [5], [6], [7], including the

optimization-based methods [8], [9] and the filter-based methods [10], [11], [12], [13]. The batch optimization methods use the preintegration of the high-rate IMU measurement to form the optimization-based method, which are usually time consuming. Filter-based methods are typically more efficient [9], [10], [11], [12], [14], [15] and robust due to improvements in the filter consistency [16], [17], [18]. In Mourikis's definition [6], the filtering-based method is mainly formed by the propagation module and the updation module. The propagation module uses the IMU measurements to predict the current body state, and the updation module uses observations from the visual corner features to eliminate the bias of the system error. The evaluation result of these works proved that the VIO system can achieve high accuracy under a natural lighting environment with a smooth and moderate camera movement. However, neither of the abovementioned systems can work well in challenging environments, particularly under overly bright or dim lighting conditions. While OpenVINS alleviates issues in low-light conditions through histogram equalization, however, in many practical applications, the combination of challenging lighting conditions and fast camera movements can present a significant challenge, resulting in system crashes or failure.

The limited sensing scope is one of the main reasons for the system vulnerability. The early works on this topic try to use multiple cameras to realize the omnidirectional observations [19], [20], [21], [22], [23], [24], [25]. The increased FOV can make the VIO system observe more textures, and the salient visual features can be tracked for a longer duration. However, these works rely on a customized hardware and the calibration of these cameras is complex. Meanwhile, advancements in lens manufacturing have made fisheye lenses available as commercial-grade CMOS sensors. The specialized camera models, such as Kannala Brandt [26] or the MEI [27], are designed to support the projection for these ultrawide cameras. Since traditional VIO systems are designed for pinhole cameras, they [4], [28], [29] have upgraded their previous works [7], [30], [31] to support the wide-angle fisheye lenses. While the upgraded ones have shown impressive localization abilities, they still lack robustness in challenging environments.

Compared with previous omnidirectional simultaneous localization and mapping (SLAM) systems equipped with at least four cameras, the 360° camera has only two back-to-back attached cameras. It can directly get the equirectangular images from the hardware device, which simplifies the calibration process. Different from the previous customized omnidirectional hardware, the commercial device of the 360° camera (e.g., Insta-360, RICOH THETA) is more lightweight and has a shorter baseline. The absence of baseline metrics is incompatible with the existent methods. OpenVSLAM [32] is the first 360° SLAM system, which supports the equirectangular image processing. However, the author neglects the highly distorted pixels and still uses simple reprojection error as the measurement constraints, which makes the system weak in practical applications. The following work 360-VIO [33] modified the measurement model and used the photometric errors as the visual constraints. Although there are several SLAM algorithms have explored the direct or sparse features [34], [35] on the 360° camera, but the

360-VIO System

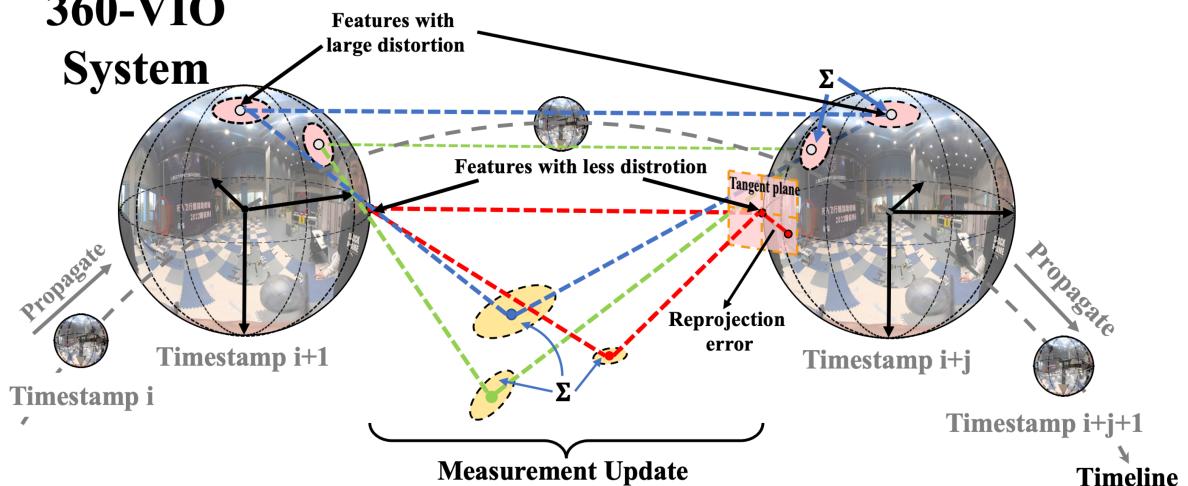


Fig. 2. Overview of the 360-VIO system: The whole system is based on an EKF framework. The propagation module uses IMU measurements to estimate the current body state, and the proposed update module utilizes the reprojection error on the tangent plane of the spherical surface to achieve robust and accurate 360-VIO.

research on the 360° VIO system is still absent and the evaluation dataset is also desired for the open community. To the best of our knowledge, our work is the first 360-VIO system that works well in different challenging environments with high accuracy.

III. METHOD

The overview of our system is shown in Fig. 2. Our method uses an EKF framework to fuse the 360° visual perception with IMU pose estimates. The entire 360-VIO system is built on top of HybVIO [36] and uses the static initialization module to initialize the system. There are four main modules in this section. Section III-A describes the projection model used in the 360° camera. Section III-B briefly introduces the developed extrinsic calibration tool for the 360° camera with the IMU device. Section III-C and Section III-D describes the propagation and the update modules of the 360-VIO system.

Before diving into details, we will briefly introduce the notations used in the following sections. Following [13], we define the state vector at timestamp k as follows:

$$\mathbf{x}_k = \left[{}^G\mathbf{T}_{\text{IMU}_k}^{(0)} \ {}^G\mathbf{v}_I^\top \ \mathbf{b}_a^\top \ \mathbf{b}_\omega^\top \ \mathbf{T}_a^\top \ {}^G\mathbf{T}_{\text{IMU}_k}^{(1)}, \dots, {}^G\mathbf{T}_{\text{IMU}_k}^{(n)} \right]^\top \quad (1)$$

where ${}^G\mathbf{T}_{\text{IMU}_k}^{(i)} = [{}^G\mathbf{R}_I^{(i)} \ {}^G\mathbf{P}_I^{(i)}]^\top$ represents the IMU state. The right superscript represents the sample $i \in (0, \dots, n)$ in a sliding window with size of n . The ${}^G\mathbf{T}_{\text{IMU}_k}^{(0)}$ represents the current state. The left superscript represents the coordinate frame and the right subscript represents the body frame of current sensor. ${}^G\mathbf{R}_I \in \text{SO}(3)$ and ${}^G\mathbf{P} \in \mathbb{R}^3$ represent the orientation and position of the IMU body frame in the global coordinate. \mathbf{T}_a^\top is a diagonal matrix that represents the multiplicative correction between the gyroscope and the accelerometer. ${}^G\mathbf{v}_I^\top$ represents the current speed of the IMU sensor. The biases of the accelerometer and gyroscope, denoted by \mathbf{b}_a^\top and \mathbf{b}_ω^\top , respectively, are typically modeled as random walk variables. The measurements from the gyroscope or the accelerometer can be modeled as



Fig. 3. Overview of the equirectangular projection. The left is the spherical representation, which we defined it as the 360-image and the right is the UV representation, which we defined it as the equirectangular image. (a) Features on the 360-image. (b) Features on the equirectangular image.

follows:

$$\begin{aligned} \boldsymbol{\omega}_m(t) &= {}^I\boldsymbol{\omega}(t) + \mathbf{b}_\omega(t) + \mathbf{n}_\omega(t) \\ \mathbf{a}_m(t) &= \mathbf{R}_{IG} \cdot \mathbf{T}_a({}^G\mathbf{a}(t) - {}^G\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t). \end{aligned} \quad (2)$$

The \mathbf{R}_{IG} means the transformation from global frame to the IMU frame. The $\mathbf{n}_\omega, \mathbf{n}_a$ are modeled as the white Gaussian noise of the gyroscope and the accelerometer.

A. Camera Model

Since the 360° camera has been rarely used for VIO, we first define the model of such a novel sensor. There are two types of representations of the 360° image: the spherical representation and the uv representation [32], [33]. We define the spherical representation as the 360-image and the UV representation as the equirectangular image [32]. Similar to the pinhole camera projection, the equirectangular projection describes the projection from the unit spherical surface onto the image plane. The description of this projection is depicted in Fig. 3. ${}^C\mathbf{P}_{f,j}$ is the visual landmark j in global coordinates, which is an extension of the unit vector ${}^C\mathbf{P}_{u,j} = [p_x, p_y, p_z]^\top$. The ${}^C\mathbf{P}_{u,j}$ is on the camera space and parameterized by the longitude θ and the latitude ϕ , where the $-\pi < \theta < \pi$ and $-\pi/2 < \phi < \pi/2$. The

parameterization of the unit vector is

$$p_x = \cos(\theta)\sin(\phi), p_y = -\sin(\phi), p_z = \cos(\theta)\cos(\phi). \quad (3)$$

So the equirectangular projection of the ${}^C\mathbf{P}_{u,j}$ to the image pixels is defined as follows:

$$\pi({}^C\mathbf{P}_{u,j}) = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} W/2\pi & 0 & W/2 \\ 0 & -H/\pi & H/2 \end{bmatrix}}_{\mathbf{K}} \begin{bmatrix} \theta \\ \phi \\ 1 \end{bmatrix} \quad (4)$$

where θ and ϕ can be derived from (3) as follows:

$$\begin{bmatrix} \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \arctan(\frac{p_x}{p_z}) \\ -\arcsin(p_y) \end{bmatrix}. \quad (5)$$

In line with other camera models, we set the matrix in the (4) as an ideal camera intrinsic \mathbf{K} . The parameters \mathbf{K} is defined by the size of the equirectangular image $[H, W]^\top$, where the H and the W represent the height and the width of the equirectangular image.

B. Calibration

In this section, we use the predefined projection model to calibrate the extrinsic parameters between the IMU sensor with the optical center of the 360-image, which is different from those captured by either the front or back camera on the 360° device. Since the developed tool is an extension of the calibration tool Kalibr [3], we will give an introduction of the calibration process first. The kernel of Kalibr is to use continuous-time batch optimization to estimate the extrinsic between the exteroceptive sensor and the IMU frame. First, we derive IMU poses as the time-varying states, which are the weighted sum of a finite number of known analytical basis functions. Considering that both the relative translation and the relative rotation have three degrees of freedom, we parameterize the pose as a 6×1 spline. Then, we use the extracted target from the camera to build the constraint for the trajectory. In this work, we extend the nonlinear perspective camera model \mathbf{g} in the [3] and the formulation is derived as follows:

$$\mathbf{y}_{j,k} = \mathbf{g}({}^G\mathbf{T}_{C,I} {}^G\mathbf{T}_{\text{IMU}_c}^{-1} {}^G\mathbf{P}_{f,j}) + \mathbf{n}_{y,k}. \quad (6)$$

With the accurate parameter \mathbf{K} defined previously, we directly analyze the calibration target with the distortion. According to [37], the quality metric used in the developed extrinsic tool is the reprojection error, which can provide a more accurate result. We treat the calibration target as a corner feature and model the pixel difference on the equirectangular image as the the reprojection error, which has been illustrated in Fig. 4(b). The idea of the proposed extrinsic estimation is to iteratively update the extrinsic parameter $\mathbf{T}_{C,I}$ to get the minimum pixel error $\mathbf{y}_{j,k}$ of the calibration target j on the board at timestamp k . The noise $\mathbf{n}_{y,k} \sim \mathcal{N}(0, R_{j,k})$ is independent of each other.

The ${}^G\mathbf{T}_{\text{IMU}_c}^{-1}$ represents the continuous trajectory of the IMU sensor, and the ${}^G\mathbf{P}_{f,j}$ represents the target in the global coordinate.

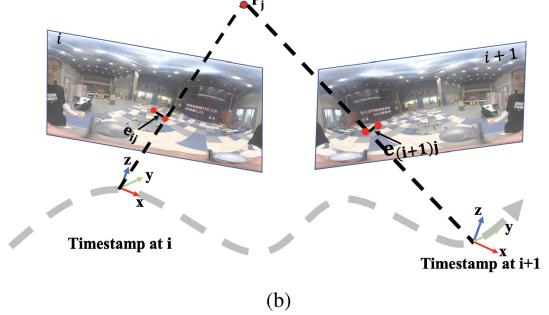
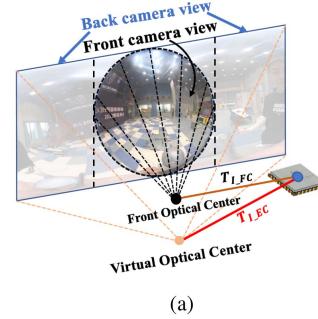


Fig. 4. Extension of the calibration toolbox. Fig. 4(a) shows the optical center's difference between the 360-image with the front fisheye. Fig. 4(b) shows the main procedure of the calibration process. (a) Optical center. (b) Extend Kalibr tool.

C. Propagation

The propagation module estimate the body state \mathbf{x}_k at time k through the gyroscope and accelerometer measurement. The kinematic model of the IMU under the continuous-time is given by the following equation:

$$\begin{aligned} {}^G\dot{\mathbf{R}}_I &= {}^G\mathbf{R}[\boldsymbol{\omega}_m - \mathbf{b}_g - \mathbf{n}_g] \times, \quad {}^G\dot{\mathbf{p}}_I = {}^G\mathbf{v}_I \\ {}^G\dot{\mathbf{v}}_I &= {}^G\mathbf{R}_I \cdot (\mathbf{T}_a \cdot \boldsymbol{\alpha}_m - \mathbf{b}_a - \mathbf{n}_a) + {}^G\mathbf{g} \\ \dot{\mathbf{b}}_g &= \mathbf{n}_{bg}, \quad \dot{\mathbf{b}}_a = \mathbf{n}_{ba}. \end{aligned} \quad (7)$$

The linearized error state $\delta\mathbf{x}$ at time $k+1$ is derived from the (7) as follows:

$$\delta\mathbf{x}_{k+1} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1}. \quad (8)$$

The linearized continuous dynamics of the error state is as follows:

$$\delta\mathbf{x}_{k+1} \simeq \mathbf{F}_k \delta\mathbf{x}_k + \mathbf{G}_k \mathbf{w} \quad (9)$$

where the matrix \mathbf{F}_k and \mathbf{G}_k is the Jacobian matrix of the kinematic model respect to the error state and the noise vector at timestamp k . The \mathbf{w} is the IMU Gaussian noise vector. In practical use, the IMU measurements are sampled discretely in time. We apply mid-term numerical integration to estimate the IMU-propagated state and the state-transition matrix $\Phi_k \simeq \mathbf{I} + \mathbf{F}_k$. So, the covariance matrix of IMU state is propagated as follows:

$$\mathbf{P}_{k+1} = \Phi_k \mathbf{P}_k \Phi_k^T + \mathbf{G}_k \mathbf{Q} \mathbf{G}_k^T \quad (10)$$

where the \mathbf{Q} is the matrix form of the process noise μ . The full uncertainty propagation is calculated as follows:

$$\mathbf{P}_{k+1} = \begin{bmatrix} \mathbf{P}_{\Pi_{k|k}} & \Phi_k \mathbf{P}_{\Pi_{k|k}} \\ \mathbf{P}_{\Pi_{k|k}}^\top \Phi_k^\top & \mathbf{P}_{\Pi_{k|k}} \end{bmatrix}. \quad (11)$$

D. Update

The main responsibility of the updation module is to use the 3-D visual landmark ${}^G\mathbf{P}_{f,j}$ with the corresponding visual feature \mathbf{z}_j to correct the drift of the predicted state \mathbf{x}_k at timestamp $k+1$, which can be summarized as follows:

$$\mathbf{r}_f = \mathbf{z}_f - \mathbf{h}(\mathbf{x}_{k+1}, {}^G\mathbf{P}_{f,j}) \simeq \mathbf{H}_x \delta \mathbf{x} + \mathbf{H}_f \delta {}^G\mathbf{P}_{f,j} + \mathbf{n}_f \quad (12)$$

where $\mathbf{h}(\cdot)$ is the nonlinear camera model and \mathbf{n}_f is the corresponding white Gaussian noise $\mathcal{N}(0, \sigma_v^2)$. With the consideration about the ideal 360-image, we model the residual on the tangent plane of the spherical surface plane instead of the image plane, which has been detailed in [7]. In order to save computational resources, following [36], we use the chain rule of the differentiation to derive the information matrix $\mathbf{H}_{f,j,x} = \frac{\mathbf{r}_f}{\sigma^2 \mathbf{P}_{f,j}} \cdot \frac{{}^G\mathbf{P}_{f,j}}{\mathbf{x}}$ with respect to the camera state as:

$$\mathbf{H}_j(\mathbf{x}) \approx \mathbf{H}_{f,j,x}(\mathbf{x}_0) \delta \mathbf{x} + \mathbf{H}_x(\mathbf{x}_0) \delta \mathbf{x} \quad (13)$$

where the j th row of the \mathbf{H} and the \mathbf{H}_f matrices correspond to the jacobians of the visual feature at the j th index.

Different from the classical perspective cameras with less feature observations, the 360° device brings the omnidirectional perception of the environment with the distortion. There are three strategies used in our algorithms. The first is the Fivept-Stewenius [38] random sample consensus (RANSAC) algorithm after the pyramidal Lucas–Kanade tracking. The second is the χ^2 outlier check before the update function and the third is the reprojection error check of the 3-D visual landmark according to our measurement noise model, which will be detailed in the following paragraph.

The update stage utilized in the 360-VIO algorithm involves three steps. First, extract the corner features in the image. Then, track these features and triangulate them. Third, measure the difference between the reprojection of the 3-D landmark with the new observations in the incoming images. Considering the efficiency, we extract the GFTT features [39] on the equirectangular image. As the aforementioned 360° camera model, the pixels on the u -axis are stretched due to the equirectangular projection. We model the observed features with additive Gaussian noise, which is derived as follows:

$$\mathbf{z}_f = \hat{\mathbf{z}}_f + \mathbf{n}_f, \mathbf{n}_f \sim \mathcal{N}\left(\frac{\sigma_f}{\cos(v)}, \sigma_f\right) \quad (14)$$

where v is the vertical value of each feature point and σ_f is the corresponding noise of pixel feature. After the tracking of the pyramidal Lucas–Kanade method and the filtering of the five points Stewenius RANSAC module, we triangulate the matched features using the simple DLT method. However, the omnidirectional observation make us drop the depth check, which is the common process for the classical camera module. Particularly, we use the inverse-depth to represent our 3-D visual landmark.

TABLE I
PARAMETER OF THE 360-VIO DATASET

Parameter	Value
IMU rate	500 [HZ]
Gyroscope noise density	4.6951×10^{-5} [rad/s/ $\sqrt{\text{Hz}}$]
Gyroscope random walk	1.7177×10^{-5} [rad/s $^2/\sqrt{\text{Hz}}$]
Accelerometer noise density	3.5558×10^{-4} [m/s $^2/\sqrt{\text{Hz}}$]
Accelerometer random walk	5.4589×10^{-5} [m/s $^3/\sqrt{\text{Hz}}$]
Camera rate	30 [HZ]
Magnitude of gravity	9.805 [m/s 2]



Fig. 5. Indoor testing environment equipped with a lighting control system and OptiTrack motion capture system.

At last, we use the (14) to relieve the filtering threshold in the 3-D visual landmark outlier reprojeciton, which can keep more important visual features for update and make the estimation more accurate and robust.

IV. EXPERIMENTAL EVALUATION

The experimental evaluation is designed and conducted to substantiate that: 1) the 360 camera model can improve the robustness of the VIO system, allowing it to maintain high accuracy under different challenging environments; 2) the quantity results of the estimation of the VIO trajectories support the efficiency of the proposed measurement module. For quantity and quality evaluation, we evaluate our 360-VIO system on the proposed 360-VIO datasets.

A. 360-VIO Dataset

To evaluate the proposed odometry method based on 360° cameras, we build a new dataset using an Insta-360 One X2 device, which provided high-resolution 4 K images captured at a rate of 30 frames per second and 500-Hz IMU measurements from its built-in IMU. The description corresponding to the dataset is listed at Table I. This dataset comprises fifteen distinct sequences, with ten designed for indoor environments and five tailored for outdoor settings.

We use the root-mean-squared error as a quantitative metric to compare our algorithm with other methods. As shown in the Fig. 5, the large indoor environment is within the OptiTrack motion capture system and the lighting control system. Different from previous datasets, we have set the illumination brightness and the camera movements as the condition variable. The sequence can be classified based on changes in illumination speed or camera movement intensity. The changes in illumination condition can be classified into normal light (N), illumination change light (IC), and low (L) light based on the brightness of the lights, which has been illustrated in Fig. 6. More specifically,

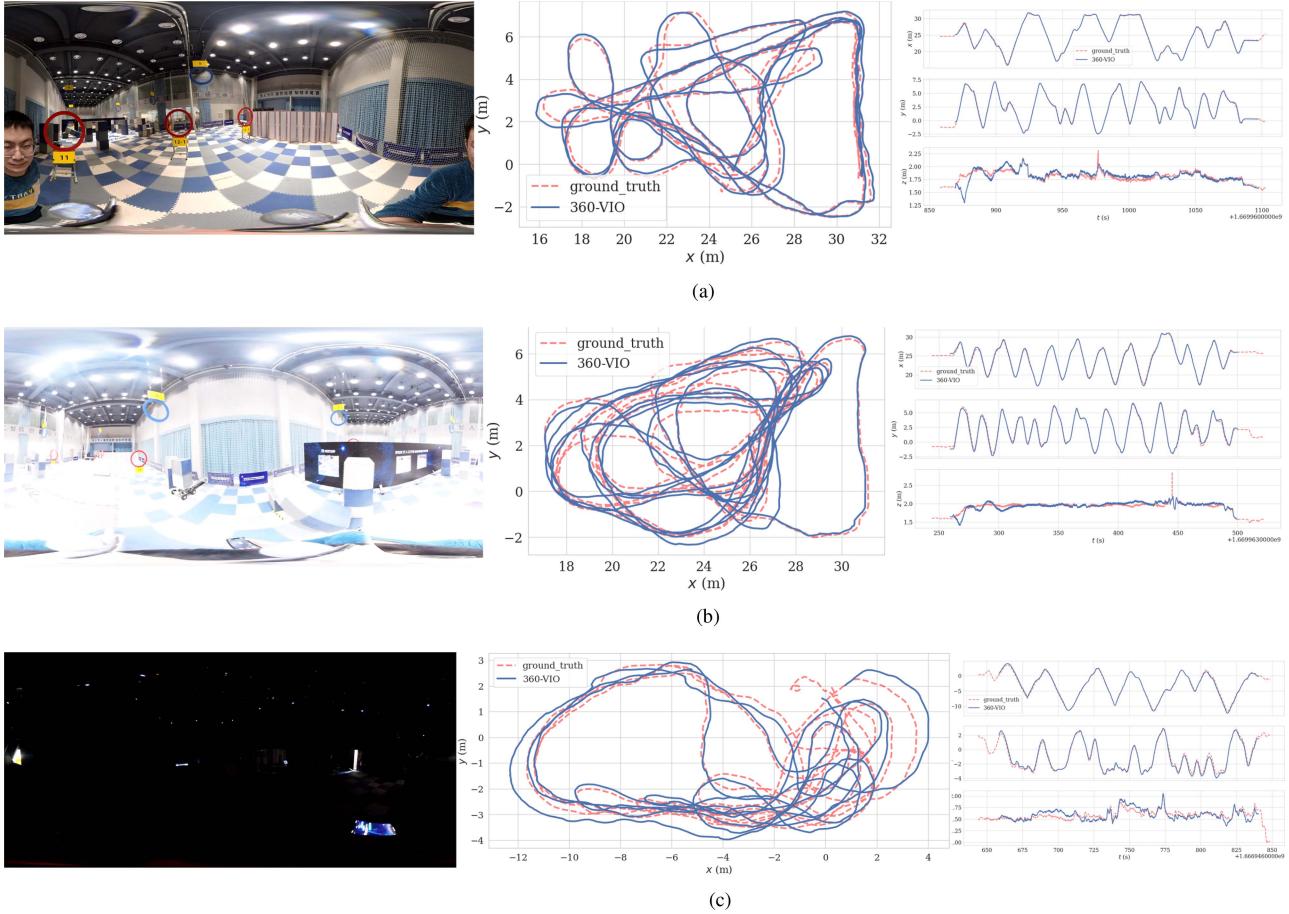


Fig. 6. Comparison of the estimated trajectories of our 360-VIO algorithm against the ground truth. (a) Normal light condition: there are no extreme brightness or darkness under this condition. (b) Illumination change light condition: the lights are rapid fluctuated between extreme brightness and darkness. (c) Low light condition: only some sparse lights is presented in the environment.

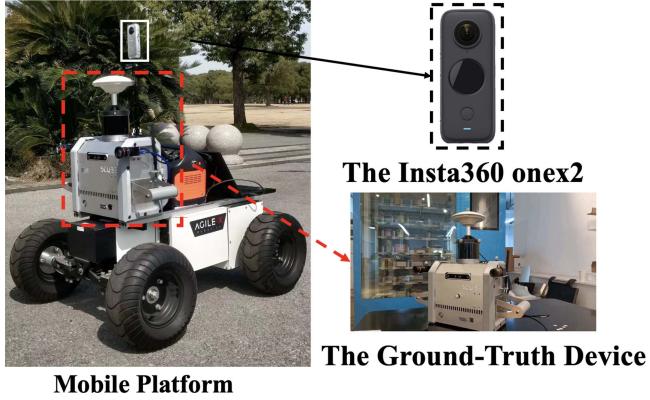


Fig. 7. Self-made mobile device for the outdoor comparison.

we noted the maximum linear speed greater than 4 m/s as fast speed (F), and lower than 2 m/s as slow speed (S). The speed between these two levels is referred to as the medium speed (M). Our outdoor experiments involved both qualitative and quantitative evaluations. To conduct the quantitative assessment, we equipped the Insta-360 with a self-developed mobile platform, as illustrated in Fig. 7. The mobile platform was also fitted

with an RTK-device for precise collection of ground-truth data. In addition, we performed a qualitative evaluation of outdoor environments, as depicted in Fig. 8. Each collection were last about 3 min, which were traveled about several hundred meters. All the data and the developed calibration tool in this study will be made publicly available to the research community.

B. Experiment Setup

We adopt the absolute trajectory error and the relative trajectory error as our evaluation metrics. Since there are not specialized 360-based VIO method, we choose the similar type camera-based VIO algorithms, such as the ORB-SLAM3 [4] and VINS-Mono [7], which supports the fisheye camera model and the MC-VIO [25] which is designed for the multiple nonoverlapped fisheye cameras. We also have tested some stereo-type fisheye system, e.g., VINS-Fisheye [29], however, the nonoverlapping stereo cameras drove the system easily to drift. We use the TartanCalib [37] to estimate the parameters between the IMU with the front/back fisheye cameras. As noted earlier, the extrinsic parameters of the equirectangular-image between the IMU were estimated using the extended Kalibr tool as is mentioned previously. All the algorithms were tested on a laptop equipped with an Intel i9 10885H@2.4GHZ with eight cores

TABLE II
VIO RESULTS COMPARISON ON THE 360-VIO BENCHMARK

Sequences (Playspeed – Metrics)	Traveling Dist.[m]	360-VIO				ORB-SLAM3 ¹				VINS-Mono ²				MC-VIO ³		Hyb-VIO ⁴	
		1.0		0.1		0.5		1.0		0.5		1.0		1.0		1.0	
		APE (m)	RPE (m)	APE (m)	RPE (m)	APE (m)	RPE (m)	APE (m)	RPE (m)	APE (m)	RPE (m)	APE (m)	RPE (m)	APE (m)	RPE (m)	APE (m)	RPE (m)
Seq 1(N-S)	245	0.456	0.058	0.186	0.108	–	–	1.406	0.082	0.933	0.134	0.776	0.045	–	–	–	–
Seq 2(N-M)	268	0.390	0.088	0.256	0.068	–	–	2.960	0.242	0.846	0.195	0.784	0.080	–	–	–	–
Seq 3(N-F)	412	0.632	0.122	0.186	0.086	–	–	–	–	26.96	2.279	3.755	0.127	–	–	–	–
Seq 4(IC-S)	273	0.515	0.064	–	–	–	–	–	–	–	–	5.730	0.154	1.158	0.052	–	–
Seq 5(IC-S)	270	0.722	0.069	–	–	–	–	–	–	–	–	–	–	–	0.507	0.061	–
Seq 6(IC-M)	391	0.874	0.142	5.245	0.281	–	–	–	–	53.401	0.247	3.079	0.089	–	–	–	–
Seq 7(L-S)	198	4.420	0.124	6.376	0.067	–	–	13.679	0.305	5.597	0.148	–	–	–	–	–	–
Seq 8(L-S)	225	0.420	0.066	–	–	–	–	–	–	–	–	–	–	–	1.312	0.608	–
Seq 9(L-S)	264	0.741	0.058	–	–	–	–	–	2.363	0.092	0.666	0.124	0.967	0.060	–	–	–
Seq 10(L-M)	198	0.646	0.064	–	–	–	–	–	–	–	–	–	–	–	3.617	0.100	–

Sequences are denoted by the indices with the respective labels of illumination conditions and camera movements. E.g. Seq 1(N-S) means the normal light condition with the slow moving speed. The bold text with underline represents the best result. The ‘-’ represents the system is drifted in this sequence. All the data(including the calibration data) will be open-sourced after paper accepted.

TABLE III
ODOMETRY COMPARISONS ON THE OUTDOOR ENVIRONMENT

Methods	sequences	Seq 1 [467 m]		Seq 2 [640 m]	
		APE (m)	RPE (m)	APE (m)	RPE (m)
360-VIO		5.220	0.074	7.377	0.076
MC-VIO*		5.260	0.158	–	–

The * symbol signifies that the play speed of the sequence or video has been reduced to 0.5 to facilitate improved algorithm performance.

The bold text with underline represents the best result.

with 128 GB RAM and an Nvidia Quadro RTX 3000 with 6 GB RAM.

C. Indoor Evaluations

In this section, we present a comparative evaluation of our proposed 360-VIO method with state-of-the-art VIO approaches, including VINS-Mono [7], ORB-SLAM3 [4], and MC-VIO [25], on the proposed 360-VIO datasets. The indoor dataset comprises ten sequences, each with varying illumination conditions and camera movements. Our experimental results demonstrate that the proposed 360-VIO method outperforms the existing VIO methods in terms of accuracy and robustness. Specifically, the 360-VIO-Reproj achieves consistently lower errors across all sequences, indicating the effectiveness of the proposed approach in diverse environmental conditions.

The Table II demonstrates the comparative evaluations with established algorithms, such as VINS-Mono and ORB-SLAM3. The 360-VIO consistently achieved competitive accuracy and maintained robustness in challenging environments. In addition, our algorithm exhibits greater efficiency, enabling real-time implementation. From the results presented in the table, we find that our proposed algorithm may perform lower than ORB-SLAM3 in some normal lighting sequences with smooth camera movements. However, this is mainly attributed to the loop closure module employed in the ORB-SLAM3 algorithm, which enables the system to repeatedly relocalize itself in the prebuilt visual map. While this module can improve the accuracy of the system, it is also time-consuming and computationally

demanding, preventing ORB-SLAM3 from being deployed in real-time applications. When we disabled the loop closure module in the ORB-SLAM3 system, it soon began to drift in all sequences. Although our approach may not always achieve the highest accuracy, it provides a more practical and efficient solution for real-world applications where both accuracy and efficiency are critical factors. This mainly due to the omniperceptive capability of our system, which allows the tracking module to capture and track visual keypoints even under challenging lighting conditions. While MC-VIO also shows robustness under some challenging illumination variations with movements, it is still vulnerable to fast camera movements due to its separately designed camera tracking module. Our approach, in contrast, utilizes the full 360° view to achieve robustness and accuracy in both steady and fast camera movements. The experimental results confirm the superiority of our proposed method in terms of accuracy and robustness in various real-world scenarios.

D. Outdoor Experiments

In this section, we conducted experiments under various outdoor conditions and different environments to test the algorithm's resilience. Our quantitative evaluation was conducted using two sequences that were recorded in campus-like environments. However, both the ORB-SLAM3 and VINS-Mono algorithms exhibited drift in these sequences. Therefore, we only recorded the results of our proposed MC-VIO algorithm in Table II for the comparison. In the qualitative evaluation, we were carrying the 360° device on different moving platforms, as shown in Fig. 8. Our experimental results demonstrate that the proposed algorithm exhibits robustness in diverse environmental conditions and illumination levels.

E. Ablation Study

In this section, we present an ablation study of the 360-VIO algorithm. As mentioned previously, the 360-VIO algorithm performs well in challenging environments. To further analyze our proposed algorithm, we specifically evaluate two part of

TABLE IV
ABLATION STUDY OF THE FOV IN THE ALGORITHM

Methods	sequences	Seq 1		Seq 2		Seq 3		Seq 4		Seq 6		Seq 8		Seq 10	
		APE (m)	RPE (m)												
360-VIO-front		0.776	0.045	0.784	0.080	1.175	0.120	0.978	0.064	1.009	0.096	0.628	0.067	3.029	0.070
360-VIO-back		0.643	0.061	0.765	0.089	0.748	0.090	1.21	0.065	1.538	0.096	0.668	0.068	1.607	0.068
360-VIO		0.456	0.058	0.390	0.088	0.632	0.121	0.515	0.064	0.874	0.142	0.420	0.066	0.646	0.064

The bold text with underline represents the best result.

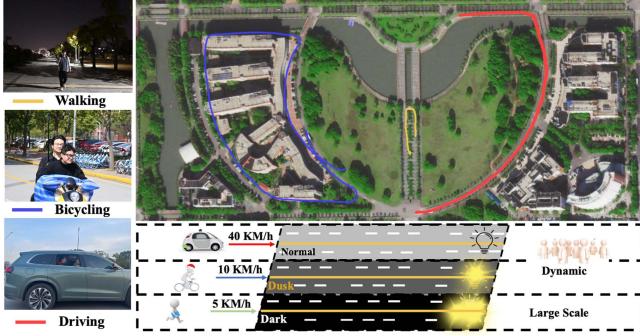


Fig. 8. Insta-360 camera was affixed to various platforms and subjected to diverse environmental conditions for a comprehensive qualitative assessment. The different color trajectories are aligned with the Google map for visualization.



Fig. 9. Equirectangular image only projected from the front fisheye.

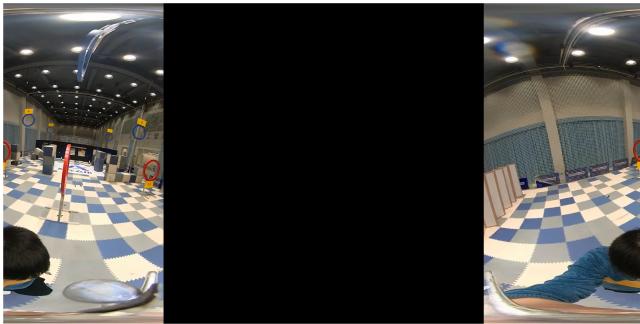


Fig. 10. Equirectangular image projected from the backend fisheye.

our system: the first is the FOV influence and the second is the constraint function in the updation module.

TABLE V
ABLATION STUDY OF THE UPDATION MODULE IN THE ALGORITHM

Methods	sequences	Seq 1		Seq 4	
		APE (m)	RPE (m)	APE (m)	RPE (m)
360-VIO-UV		3.904	0.043	4.814	0.054
360-VIO*		5.980	0.055	0.522	0.064
360-VIO		0.456	0.058	0.515	0.058

The bold text with underline represents the best result.

To investigate the impact of FOV on the algorithm, we added two additional camera configurations, using only the front camera and only the back camera, respectively. The representation of these two settings is shown in Figs. 9 and 10

The result is shown in the Table IV, Seq 1–3 provide that a single camera cannot achieve the best results due to feature losts and a larger FOV can brought a better accuracy result. The whole results prove that a larger FOV can make a more robust system.

As mentioned earlier, 360° images can be presented in the form of the UV-image or the 360-image. Therefore, we use three different constraint conditions to construct the update module. For the UV-image, we built the reprojection error on the image space, which we named it as the 360-VIO-UV; For the 360-image, we defined the reprojection error on the tangent space, which we named it as the 360-VIO*. The last is our proposed 360-VIO algorithm, which also considered the uncertainties of the features from the equirectangular projection. To demonstrate the performance, we have selected two sequences for comparison and the results are shown at Table V. As expected, the directly use of the reprojection error on the image space is more vulnerable than the reprojection error on the tangent plane. Our proposed measurement module improves both accuracy and the robustness of the system, showing a significant impact on the overall algorithm.

V. CONCLUSION

In this article, we presented a novel 360-VIO system that leverages equirectangular images to enhance the algorithm’s robustness. We evaluated the proposed approach on various datasets and compared its performance with existing methods. Our experimental results demonstrated that the proposed algorithm achieves state-of-the-art odometry accuracy and exhibits impressive robustness in challenging environments. Specifically, the algorithm showed consistent performance in various environmental conditions, indicating its effectiveness in real-world applications. In one word, the proposed 360-VIO

algorithm provides a valuable contribution to the field of VIO and has the potential to inspire further research in this domain and help advance the development of robust navigation and sensing systems for a wide range of applications.

REFERENCES

- [1] T.-j. Lee, C.-h. Kim, and D.-i. D. Cho, "A monocular vision sensor-based efficient SLAM method for indoor service robots," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 318–328, Jan. 2019.
- [2] C. Zhai, M. Wang, Y. Yang, and K. Shen, "Robust vision-aided inertial navigation system for protection against ego-motion uncertainty of unmanned ground vehicle," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12462–12471, Dec. 2021.
- [3] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 1280–1286.
- [4] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
- [5] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A research platform for visual-inertial estimation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 4666–4672.
- [6] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2007, pp. 3565–3572.
- [7] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.
- [8] L. Yu, J. Qin, S. Wang, Y. Wang, and S. Wang, "A tightly coupled feature-based visual-inertial odometry with stereo cameras," *IEEE Trans. Ind. Electron.*, vol. 70, no. 4, pp. 3944–3954, Apr. 2023.
- [9] D. Zou, Y. Wu, L. Pei, H. Ling, and W. Yu, "StructVIO: Visual-inertial odometry with structural regularity of man-made environments," *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 999–1013, Aug. 2019.
- [10] S. Chen, "Kalman filter for robot vision: A survey," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4409–4420, Nov. 2012.
- [11] G. Huang, "Visual-inertial navigation: A concise review," in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 9572–9582.
- [12] Y. Liu et al., "Stereo visual-inertial odometry with multiple Kalman filters ensemble," *IEEE Trans. Ind. Electron.*, vol. 63, no. 10, pp. 6205–6216, Oct. 2016.
- [13] D. He, W. Xu, and F. Zhang, "Symbolic representation and toolkit development of iterated error-state extended kalman filters on manifolds," *IEEE Trans. Ind. Electron.*, 2023.
- [14] L. Von Stumberg, V. Usenko, and D. Cremers, "Direct sparse visual-inertial odometry using dynamic marginalization," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 2510–2517.
- [15] J. Jiang, J. Yuan, X. Zhang, and X. Zhang, "DVIO: An optimization-based tightly coupled direct visual-inertial odometry," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11212–11222, Nov. 2021.
- [16] K. Shen, M. Wang, M. Fu, Y. Yang, and Z. Yin, "Observability analysis and adaptive information fusion for integrated navigation of unmanned ground vehicles," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7659–7668, Sep. 2020.
- [17] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," *Int. J. Robot. Res.*, vol. 41, no. 7, pp. 667–689, 2022.
- [18] S.-M. Lee, J. Jung, S. Kim, I.-J. Kim, and H. Myung, "DV-SLAM (dual-sensor-based vector-field SLAM) and observability analysis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 2, pp. 1101–1112, Feb. 2015.
- [19] M. Ramezani, K. Khoshelham, and C. Fraser, "Pose estimation by omnidirectional visual-inertial odometry," *Robot. Auton. Syst.*, vol. 105, pp. 26–37, 2018.
- [20] H. Seok and J. Lim, "ROVINS: Robust omnidirectional visual inertial navigation system," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 6225–6232, Oct. 2020.
- [21] Y. Yu, W. Gao, C. Liu, S. Shen, and M. Liu, "A GPS-aided omnidirectional visual-inertial state estimator in ubiquitous environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 7750–7755.
- [22] M. Ramezani, K. Khoshelham, and L. Kneip, "Omnidirectional visual-inertial odometry using multi-state constraint Kalman filter," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 1317–1323.
- [23] J.-P. Tardif, Y. Pavlidis, and K. Daniilidis, "Monocular visual odometry in urban environments using an omnidirectional camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 2531–2538.
- [24] D. Scaramuzza and R. Siegwart, "Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1015–1026, Oct. 2008.
- [25] Y. He, H. Yu, W. Yang, and S. Scherer, "Towards robust visual-inertial odometry with multiple non-overlapping monocular cameras," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 9452–9458.
- [26] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1335–1340, Aug. 2006.
- [27] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2007, pp. 3945–3950.
- [28] H. Matsuki, L. Von Stumberg, V. Usenko, J. Stückler, and D. Cremers, "Omnidirectional DSO: Direct sparse odometry with fisheye cameras," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 3693–3700, Oct. 2018.
- [29] W. Gao, K. Wang, W. Ding, F. Gao, T. Qin, and S. Shen, "Autonomous aerial robot using dual-fisheye cameras," *J. Field Robot.*, vol. 37, no. 4, pp. 497–514, 2020.
- [30] 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.
- [31] 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.
- [32] S. Sumikura, M. Shibuya, and K. Sakurada, "OpenvSLAM: A versatile visual SLAM framework," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 2292–2295.
- [33] H. Huang and S.-K. Yeung, "360VO: Visual odometry using a single 360 camera," in *Proc. Int. Conf. Robot. Automat.*, 2022, pp. 5594–5600.
- [34] H. Guan and W. A. Smith, "BRISKs: Binary features for spherical images on a geodesic grid," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4886–4894.
- [35] Q. Zhao, W. Feng, L. Wan, and J. Zhang, "SPHORB: A fast and robust binary feature on the sphere," *Int. J. Comput. Vis.*, vol. 113, pp. 143–159, 2015.
- [36] O. Seiskari et al., "HybVIO: Pushing the limits of real-time visual-inertial odometry," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, pp. 287–296.
- [37] B. P. Duisterhof, Y. Hu, S. H. Teng, M. Kaess, and S. Scherer, "TartanCalib: Iterative wide-angle lens calibration using adaptive subpixel refinement of apriltags," 2022, *arXiv:2210.02511*.
- [38] H. Stewenius, C. Engels, and D. Nistér, "Recent developments on direct relative orientation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 60, no. 4, pp. 284–294, 2006.
- [39] J. Shi et al., "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 593–600.



Qi Wu (Student Member, IEEE) received the bachelor's degree in electrical engineering and automation from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016, and the master's degree in electronic science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in electronic science and technology with Shanghai Jiao Tong University, Shanghai, China.



Xiangyu Xu received the bachelor's degree in geophysics and master's degree in navigation engineering from Wuhan University, Wuhan, China, in 2016 and 2019, respectively.

He is currently the Senior Algorithm Engineer with ByteDance, Beijing, China. His research interests include the areas of visual-inertial odometry, robotics, and reconstruction algorithms.

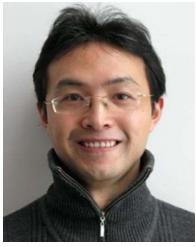


Xieyuanli Chen received the bachelor's degree in electrical engineering and automation from Hunan University, Changsha, China, in 2015, the master's degree in robotics from the National University of Defense Technology (NUDT), Changsha, China, in 2017, and the Ph.D. degree in robotics from Photogrammetry and Robotics Laboratory, University of Bonn, Bonn, Germany, in 2022.

He joined the College of Intelligence Science and Technology, NUDT, in 2022, where he is currently an Associate Professor.



Guoqing Liu received the bachelor's degree in automation from the School of Automation, Harbin University of Science and Technology, Harbin, China, in 2018, and the master's degree in robotics science and engineering with the Faculty of Robot Science and Engineering, Northeastern University, Shenyang, China, in 2021.



Ling Pei (Senior Member, IEEE) received the Ph.D. degree in testing and measurement technology and instrumentation from Southeast University, Nanjing, China, in 2007.

He is currently a Professor with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. From 2007 to 2013, he was a Specialist Research Scientist with Finnish Geospatial Research Institute, Masala, Finland. He has authored or coauthored more than 100 scientific papers. He is also an inventor of 25 patents and pending patents. His research interests include the areas of indoor/outdoor seamless positioning, ubiquitous computing, wireless positioning, bioinspired navigation, context-aware applications, location-based services, and navigation of unmanned systems.

Dr. Pei was a recipient of the Shanghai Pujiang Talent in 2014.



Sheng Yang received the Ph.D. degree in testing and measurement technology and instrumentation from Tsinghua University, Beijing, China, in 2019.

He is currently a Research Engineer leading the High-Definition Mapping team in the Autonomous Driving Lab of DAMO Academy, Alibaba Group, Hangzhou, China. His research interests include computer graphics, geometric modeling, and simultaneous localization and mapping for robotics.



Shilei Wen received the B.E. and M.E. degrees in computer software and theory from the Huazhong University of Science and Technology, Wuhan, China, in 2007 and 2011, respectively.

He is currently the Director of XR technology in Intelligent Creation Department, ByteDance, Beijing, China. His research interests include image/video understanding, virtual human, AR, computer vision, and machine learning.

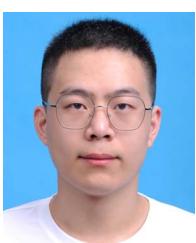


Chao Long received the bachelor's degree in detection, guidance, and control from Central South University, Changsha, China, in 2013, and the master's degree in aerospace engineering from Zhejiang University, Zhejiang, China, in 2016.

He is currently the Senior Algorithm Engineer with ByteDance, Beijing, China. His research interests include the areas of visual-inertial odometry, robotics, and reconstruction algorithms.



Junyuan Deng received the bachelor's degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2021, where he is currently working toward the master's degree in information engineering.



Wenxian Yu (Senior Member, IEEE) received the B.S. degree in radio measurement and control and data transmission, M.S. degree in communication and information systems, and Ph.D. degree in communication and information systems from the National University of Defense Technology, Changsha, China, in 1985, 1988, and 1993, respectively. He is currently a Yangtze River Scholar Distinguished Professor and the Head of the research part in the School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. From 1996 to 2008, he was a Professor with the College of Electronic Science and Engineering, National University of Defense Technology, where he was also the Deputy Head of the College and an Assistant Director of the National Key Laboratory of Automatic Target Recognition. From 2009 to 2011, he was the Executive Dean of the School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University. His research interests include remote sensing information processing, automatic target recognition, and multisensor data fusion.