micromachines

MDPI

Article

A Novel Fault-Tolerant Navigation and Positioning Method with Stereo-Camera/Micro Electro Mechanical Systems Inertial Measurement Unit (MEMS-IMU) in Hostile Environment

Cheng Yuan [1,2], Jizhou Lai [1,2,*], Pin Lyu [1,2], Peng Shi [1,2], Wei Zhao [1,2] and Kai Huang [3]

[1] Navigation Research Center, College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China; ycauto@nuaa.edu.cn (C.Y.); lvpin@nuaa.edu.cn (P.L.); ship@nuaa.edu.cn (P.S.); zhwac@nuaa.edu.cn (W.Z.)

[2] Key Laboratory of Internet of Things and Control Technology in Jiangsu province, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China

[3] Shanxi Baocheng Aviation Instrument Co., Ltd., AVIC, Baoji 721006, China; bj212hk@163.com

* Correspondence: laijz@nuaa.edu.cn; Tel.: +86-138-5147-5429

Received: 13 September 2018; Accepted: 22 November 2018; Published: 27 November 2018

Abstract: Visual odometry (VO) is a new navigation and positioning method that estimates the ego-motion of vehicles from images. However, VO with unsatisfactory performance can fail severely in hostile environment because of the less feature, fast angular motions, or illumination change. Thus, enhancing the robustness of VO in hostile environment has become a popular research topic. In this paper, a novel fault-tolerant visual-inertial odometry (VIO) navigation and positioning method framework is presented. The micro electro mechanical systems inertial measurement unit (MEMS-IMU) is used to aid the stereo-camera, for a robust pose estimation in hostile environment. In the algorithm, the MEMS-IMU pre-integration is deployed to improve the motion estimation accuracy and robustness in the cases of similar or few feature points. Besides, a dramatic change detector and an adaptive observation noise factor are introduced, tolerating and decreasing the estimation error that is caused by large angular motion or wrong matching. Experiments in hostile environment showing that the presented method can achieve better position estimation when compared with the traditional VO and VIO method.

Keywords: stereo visual-inertial odometry; fault tolerant; hostile environment; MEMS-IMU

1. Introduction

Visual navigation is an emerging technology that uses camera to capture images of the surrounding environment and processes these images to estimate ego-motion, recognize path, and make navigation decisions. The visual sensor is mature, low-cost and widely-used in robotics. Given that visual sensor is a passive sensor and does not rely on any external equipment except ambient light, one of the most important features of visual navigation is the autonomy. With the improvement of computational capabilities, visual navigation can be applied to many important applications in various fields, for instance, robot navigation [1], unmanned aerial vehicles [2], and virtual or augmented reality.

Visual odometry (VO) was first raised by Nister et al. [3] and it has become a widely-used pose estimation method. Typical VO detects and extracts feature points from a series of images that were captured by camera, then matches feature points and calculates relative pose to estimate the relative ego-motion of camera. VO can be classified based on the number of cameras into monocular VO, stereo (binocular) VO [4], and multi-camera VO [5]. The main difference is that stereo and multi-camera

VO can get absolute scale information in application while monocular VO dose not, and therefore requires a more complex initial process. Thus, the stereo VO is usually the preferable choice in practical navigation

Micro electro mechanical systems inertial measurement unit (MEMS-IMU) is also a common sensor in robots, unmanned aerial vehicles, and other moving carriers to estimate ego-motion [6,7]. It is mainly composed of accelerometers and gyroscopes, which are respectively used to obtain the acceleration and angular velocity of the carrier. Its high frequency provides precious motion information filling the interval gap of lower frequency associated vision sensors. Through using the two integrals of the acceleration and angular velocity, the attitude of the carrier can be measured. It also does not rely on any external information, can work in all conditions at any time, and has high data update rate, short-term accuracy and stability.

In recent years, visual and inertial information are usually combined to estimate the six degrees of freedom (6DOF) pose. When compared to VO, visual inertial odometry (VIO) [4,8–10] makes good use of the visual sensors and the inertial sensors, thereby acquiring more precise and robust 6DOF pose estimation. That also makes VIO play an essential role in autonomous navigation, especially in GPS-denied environment. Besides, more and more mobile robots are navigating through VIO, owing to the recent hardware improvements in mobile central processing units (CPUs) and graphics processing units (GPUs) (e.g., NVIDIA Jetson TX2 (NVIDIA corporation, Santa Clara, CA, USA)).

The mainstream of existing VIO approaches can be classified into loose coupling and tight coupling [2,5,9–11] by type of information fusion shown in Figure 1. When the system is loosely-coupled, both inertial and visual information are seen as independent measurements. The process of visual pose estimation, regarded as a black box, is only used to update a filter to restrain the inertial measurement unit (IMU) covariance propagation. By contrast, tight coupling considers the interaction of all measurements of sensors information before pose estimation, thereby achieving higher accuracy than loose coupling.

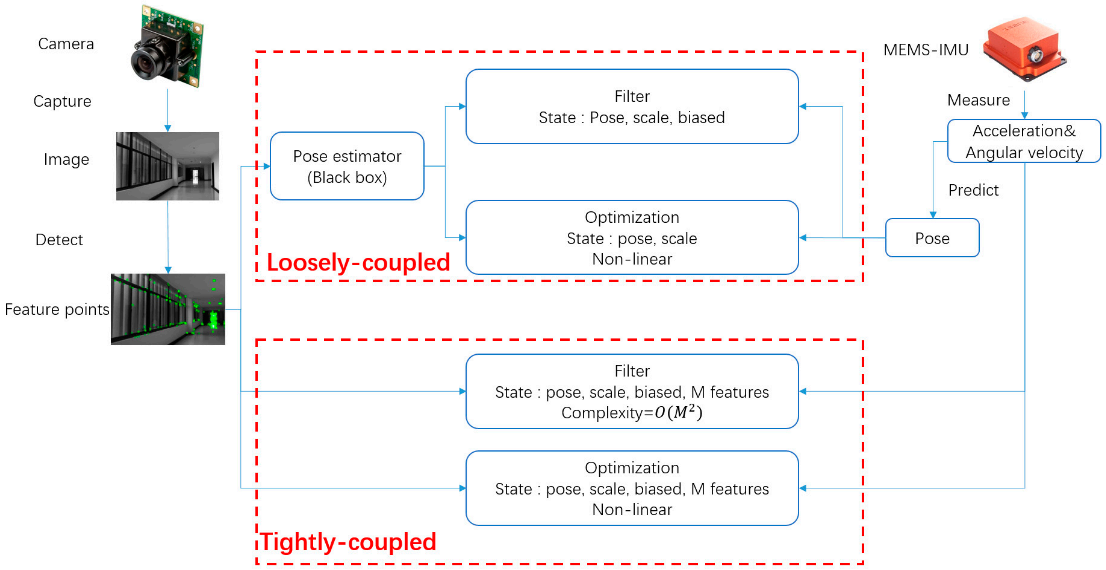


Figure 1. Loosely and tightly coupled visual inertial odometry (VIO).

Recently loosely-coupled stereo VIO systems are mostly based on Kalman filter and its derivatives. Tardif, et al. [12] proposed an EKF-based stereo VIO deployed on a moving vehicle. It used inertial information to predict the state and the stereo VO motion estimation as observations to get high frequency positioning information. Nevertheless, all of the states forecasted by inertial information, the covariance is sensitive to the IMU's bias and drift. Liu, et al. [13] presented a stereo VIO that carried out the orientation and position estimation with three filters. It fused the accelerometer and gyroscope to estimate a drift-free pitch and roll angle then fused VO and IMU to estimate motion. Nevertheless, its filtering architecture was complex and not in real-time. Schmid, et al. [14] proposed a real-time stereo VIO. It computed high quality depth images and estimated the ego-motion by key-frame based VO and fused with the data of inertial information. However, it did not take the stereo VO's failure into account. All loosely-coupled stereo VIO systems share the disadvantage that the stereo VO's and IMU's covariance were independent and cannot reflect the entire error.

Recently tightly-coupled stereo VIO systems mainly use a filtering-based [15] or optimization-based [16] approach. Filtering-based methods propagated the mean and covariance in kalman-filtering framework, together with feature points and IMU's error. Sun, et al. [11] presented a filter-based stereo VIO system using the multi-state constraint kalman filter (MSCKF) [15] applied on an unmanned aerial vehicle. The system focused on lower computation costs. Ramezani, et al. [17] presented a stereo VIO system that was based on MSCKF and applied on vehicle, focusing on highly precise positioning. However, approaches above had high dimensional states vector and lack of robustness. The target of the optimization-based approach target was to minimize an energy function with a non-linear optimization by gauss-newton algorithm through frameworks, such as g2o [18] and ceres [19]. Usenko, et al. [4] presented a direct stereo VIO system estimated motion by minimizing a combined photometric and inertial energy function. It employed semi-dense depth maps instead of sparse feature points. Nevertheless, the inertial stability easily influenced by visual error and fault-tolerant method is simple consideration.

Subject to visual limitation, visual navigation is easily influenced when facing large scene changes that are caused by fast angular motion and low or dynamic light. To avoid positioning interruption, a fatal failure in robot navigation, current research mainly focuses on changing the feature descriptor to enhance the robustness of VO. Alismail, et al. [20] proposed new binary descriptors to achieve robust and efficient visual odometry with applications to poorly lit subterranean environments. However, the descriptors utilized information just from the images. When fast angular motion causes an image to be blurred or the environment is dark, the VO is doomed to fail. That will result in serious consequences.

To achieve satisfactory performance of VO withstanding all the limitations mentioned above, a fault-tolerant adaptive extended kalman filter (FTAEKF) framework integrated with a stereo-camera and a MEMS-IMU is proposed in this paper. The use of an EKF or one of its variants has been favored and extensively employed to fuse inertial and vision data, essentially to resolve pose estimation problem. When compared to traditional loose and tight VIO framework, both robustness and accuracy are under orders. Our main contributions are as follows:

- A stereo VIO with MEMS-IMU aided method is proposed in the framework. MEMS-IMU pre-integration constraint from prediction model is used to constrain a range of candidate feature points searching and matching. The constraint also set as to optimize the initial iterator pose to avoid local optimum instead of adding MEMS-IMU measurements error joint optimization.
- An adaptive method is introduced to adjust measurement covariance according to motion characteristic. Besides, a novel fault-tolerant mechanism is used to decide whether stereo VIO pose estimation is reliable by comparing it with MEMS-IMU measurements.

An improved stereo VIO method based on ORB-SLAM2 [21] (a visual-only stereo SLAM system demonstrated with its superior performance) is proposed in the framework. The framework can be easily integrated with any other stereo VO method. Because the computation process of MEMS-IMU pre-integration and initial iteration point prediction are mostly independent with the stereo VO.

The remainder of this paper is structured as follows: The definitions of coordinates and some symbols are presented in Section 2.1. The stereo VIO system aided by MEMS-IMU is introduced in Section 2.2. The FTAEKF is presented in Section 2.2.3. Experiment and evaluation of the proposed method are shown in Section 3, followed by discussion in Section 4.

2. Materials and Methods

2.1. Coordinates and Notations

The four coordinates that were used in our framework are shown in Figure 2, The world frame W is defined as ENU (east-north-up) by axes X_W, Y_W, and Z_W, with Z_W opposite to gravity, Y_W points forward. The IMU frame, coincided with the body frame B also defined as ENU is attached to the center of MEMS-IMU with Z_B pointing upward and Y_B points forward. The camera frame C is set at the coordinate of left camera with Z_C forward and Y_C points downward. C is rigid relative pose with B. The relative pose is calibrated in advance.

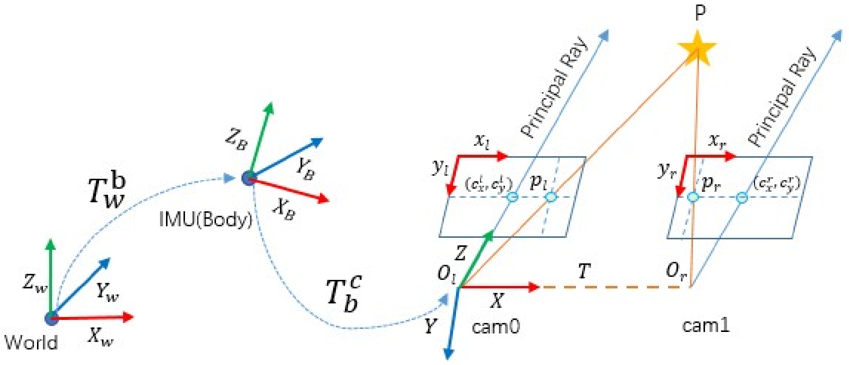


Figure 2. An illustration of coordinate system.

The rotation matrix of framework is modeled by ZYX Euler angles. To get from w to b, rotates about Z_W, Y_W, and X_W axes in turn, by the yaw angle ψ the pitch angle γ and the roll angle θ, respectively. The transformation matrix $\mathbf{T}$ is $\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$, where $\mathbf{R} \in SO(3)$ denotes the rotation matrix, and the rotation matrix $\mathbf{R}_w^c$ represents from w to c. $\mathbf{t} = \left(p_x, p_y, p_z\right)^T$ denotes the translation vector. Vectors in the camera, body and world frames are defined as $(\cdot)^c$, $(\cdot)^b$ and $(\cdot)^w$, respectively. The transformation matrix from w to b is $\mathbf{T}_w^b$, b to c is $\mathbf{T}_b^c$.

2.2. Framework of Fault-Tolerant with Stereo-Camera and MEMS-IMU

The pipeline of the proposed framework is illustrated in Figure 3. The aim of the proposed framework is to get robust and precise motion estimation in a hostile environment. The loop closing and full bundle adjustment in ORB-SLAM2 are not involved in this paper. Our contributions are mainly on the dark red block and the red arrow.

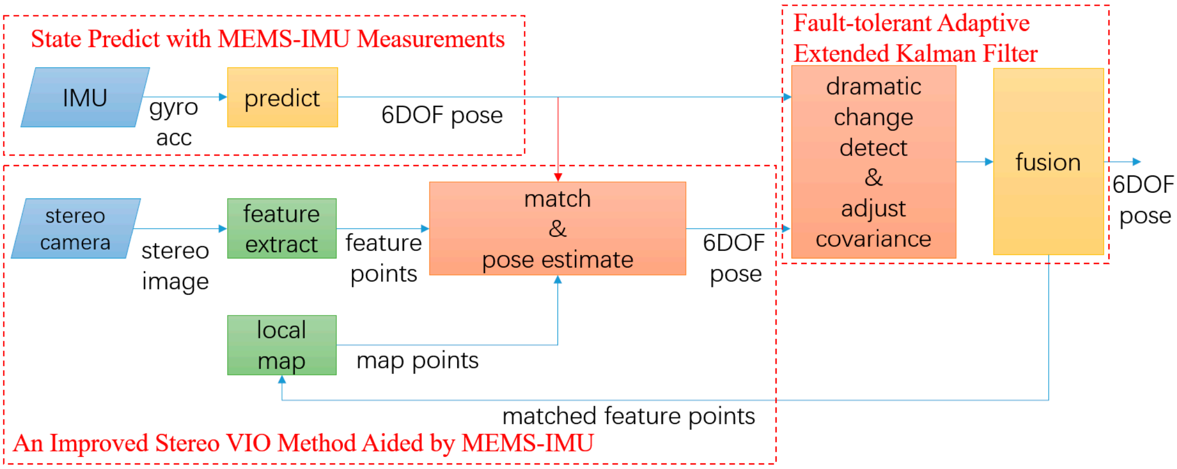

State Predict with MEMS-IMU Measurements
IMU
gyro
acc
predict
6DOF pose
Fault-tolerant Adaptive Extended Kalman Filter
dramatic change detect & adjust covariance
fusion
6DOF pose
stereo camera
stereo image
feature extract
feature points
match & pose estimate
6DOF pose
local map
map points
matched feature points
An Improved Stereo VIO Method Aided by MEMS-IMU

Figure 3. Framework of the proposed method. (the dark red blocks and red line are the difference between traditional VIO framework. the blue blocks represent the source from micro electro mechanical systems inertial measurement unit (MEMS-IMU) and stereo-camera. the green blocks represent traditional VO and dark yellow blocks represent MEMS-IMU measurements aided).

The stereo-camera and MEMS-IMU are tightly-coupled based on FTAEKF. The pre-integration of MEMS-IMU measurement confines the range of searching and matching feature points, and fault tolerance. Different from the traditional VIO method, the pre-integration of MEMS-IMU measurements is used to optimize the initial iterate point of pose estimation. It is also used to decide whether the result of pose estimation is credible to detect fault. Besides, to reflect the accumulated drift error, the observation covariance is adaptive according to motion characteristics. It combines the good properties of both loosely-coupled and tightly-coupled approaches. In this framework, the independence of stereo VO maximized. The framework has a good level of fault tolerance. It can function properly, even under stereo VIO failure, and then recover the whole system. This is because the framework allows a limited amount of independence and stereo VIO system avoids scale ambiguity in the monocular VO system. The details are described below.

2.2.1. State Predict with MEMS-IMU Measurements

The framework of FTAEKF is based on an iterated EKF where the state prediction is driven by IMU measurements. The system states $x \in \mathbb{R}^{16\times 1}$ of VIO consists of number of states:

$$\mathbf{x} = \left(\mathbf{q}^w, \mathbf{p}^w, \mathbf{v}^w, \boldsymbol{\beta}_g^b, \boldsymbol{\beta}_a^b\right)^T \tag{1}$$

Namely, $\mathbf{q}^w = (q_0, q_1, q_2, q_3)^T$ is the attitude in quaternions, reflecting the world frame (W) to the body frame (B). $\mathbf{p}^w = (px^w, py^w, pz^w)^T$ is the position and $\mathbf{v}^w = \left(v_x^w, v_y^w, v_z^w\right)$ is the velocity expressed in the world frame, $\boldsymbol{\beta}_g^b$ and $\boldsymbol{\beta}_a^b$ are the biases of three-axis gyroscopes and three-axis accelerometers, respectively. The measurements from gyroscope and accelerometer are denoted as $\boldsymbol{\eta}_{wb}^b$ and $\mathbf{a}_{wb}^b$, respectively.

The prediction model vector $\dot{x} = (\dot{\mathbf{q}}^w, \dot{\mathbf{p}}^w, \dot{\mathbf{v}}^w, \dot{\boldsymbol{\beta}}_g^b, \dot{\boldsymbol{\beta}}_a^b)^T$ is defined as:

$$\begin{aligned}
\dot{\mathbf{q}}^w &= \tfrac{1}{2}\Omega(\hat{\boldsymbol{\eta}}_{wb}^b)\mathbf{q}^w \\
\dot{\mathbf{p}}^w &= \mathbf{v}^w \\
\dot{\mathbf{v}}^w &= \mathbf{C}_b^w\left(\mathbf{a}_{wb}^b - \boldsymbol{\beta}_a^b\right) + \mathbf{g}^w \\
\dot{\boldsymbol{\beta}}_g^b &= 0 \\
\dot{\boldsymbol{\beta}}_a^b &= 0
\end{aligned} \tag{2}$$

with $\mathbf{C}_b^w$ representing the rotation matrix from B to W, the instantaneous angular velocity of B relative to W expressed in coordinate frame B $\hat{\boldsymbol{\eta}}_{wb}^b$ and the quaternion update matrix $\Omega(\hat{\boldsymbol{\eta}}_{wb}^b)$ are defined as: $\hat{\boldsymbol{\eta}}_{wb}^b = \boldsymbol{\eta}_{wb}^b - \boldsymbol{\beta}_g^b$,

$$\Omega\left(\hat{\boldsymbol{\eta}}_{wb}^b\right) = \begin{bmatrix} 0 & -\hat{\eta}_{wbx}^b & -\hat{\eta}_{wby}^b & -\hat{\eta}_{wbz}^b \\ \hat{\eta}_{wbx}^b & 0 & -\hat{\eta}_{wbz}^b & \hat{\eta}_{wby}^b \\ \hat{\eta}_{wby}^b & \hat{\eta}_{wbz}^b & 0 & -\hat{\eta}_{wbx}^b \\ \hat{\eta}_{wbz}^b & -\hat{\eta}_{wby}^b & \hat{\eta}_{wbx}^b & 0 \end{bmatrix} \tag{3}$$

In proposed framework, the pre-integration of MEMS-IMU measurements is obtained through the prediction model.

2.2.2. An Improved Stereo VIO Method Aided by MEMS-IMU

In this part, the pre-integration of MEMS-IMU measurements is used to aid the stereo VO system. The stereo VIO system that was employed in this paper is based on ORB-SLAM2 with good performance.

Both original feature based VO and VIO use brute-force or bag of words (BOW) matchers to match extracted feature points within reference frame and current frame These matchers take the descriptor of one feature in current frame and are matched to all other features in reference frame using hamming distance calculation. The closest one is returned. As a result, the pose estimation produced error when false matching occurred frequently in a hostile environment due to the close hamming distance of similar descriptor. In our approach, the MEMS-IMU measurements are pre-integrated to aid stereo VIO through constraining matching and predicting initial iteration pose. The process of this part shown in Figure 4.

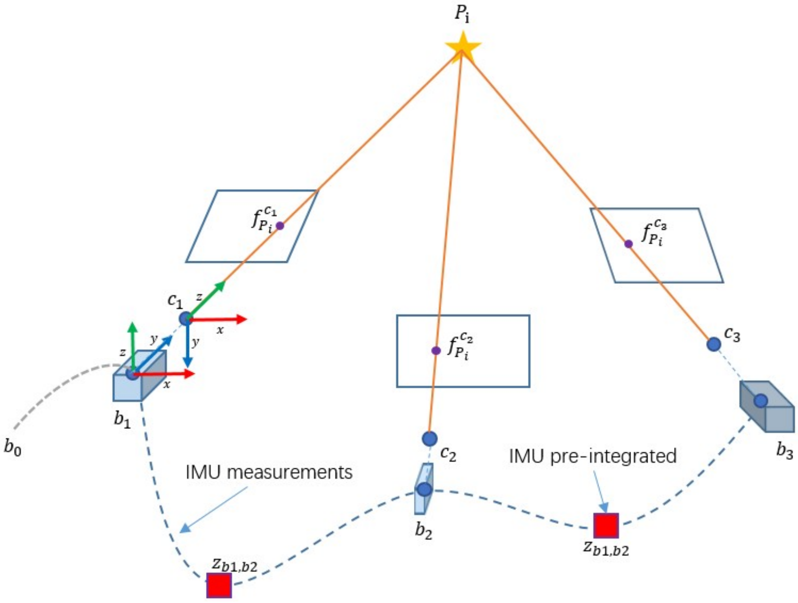


Figure 4. The process of improved stereo VIO method aided by MEMS-IMU. The inertial measurement unit (IMU) measurements are pre-integrated to predict position of feature points.

Traditionally, the initial frame pose of stereo VO is configured as world frame. However, it hardly reflects physical truth. As shown in Figure 4, VIO initialized coordinate with MEMS-IMU forward as initial heading and aligns geographic coordinate system through gravity. The stereo VIO pose is compensated by $\mathbf{T}_w^{b_1}$ from the MEMES-IMU measurement.

$$\mathbf{T}_w^{b_1} = \begin{bmatrix} \mathbf{R}_w^{b_1} & \mathbf{t}_w^{b_1} \\ 0 & 1 \end{bmatrix}, \mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^{3\times 1} \tag{4}$$

where $\mathbf{R}_w^{b_1}$ is the rotation matrix and $\mathbf{t}_w^{b_1}$ are the translation matrix from w to b_1 when VIO obtains the first image. The time interval between the image and closest MEMS-IMU measurement can be ignored due to high frequency of MEMS-IMU and low dynamic condition in beginning.

When the first stereo image is retrieved from camera, ORB feature points are extracted and matched with left and right image to estimate the depth through epipolar and disparity constraints. Then initial three-dimensional (3D) feature points in C are generated and projected based on initial pose. When a new frame was obtained from the stereo-camera, the 3D feature points are reconstructed then matched to the reference frame 3D feature points with ORB descriptors. In order to avoid the false matching caused by similar descriptors in a hostile environment. We introduce MEMS-IMU pre-integration constraint, which confined the searching and matching region to get more correct matching.

As shown in Figure 5, a point P_i is observed by two consequent frames that obtain two feature points $f^{c_1}_{P_i}$, $f^{c_2}_{P_i}$. The feature point in current frame can be project to last frame with MEMS-IMU pre-integration. The coordinates in the pixel coordinates of both feature points $f^{c_1}_{P_i}$ and $f^{c_2c_1}_{P_i}$ are close after reprojection. We can match within bounds to decrease the workload and possibility of error.

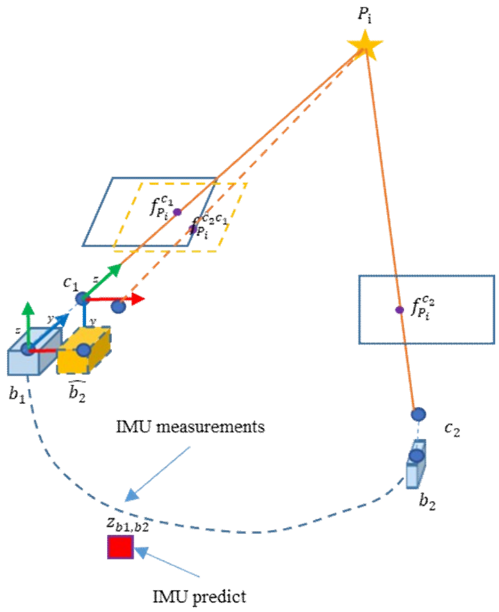


Figure 5. An illustration of predicting searching region with pre-integrating measurements of MEMS-IMU.

In our approach, the MEMS-IMU pre-integration is obtained with the prediction model. MEMS-IMU measurements between two consequent frames at discrete time $k-m$, k predict MEMS-IMU pre-integration $\Delta\zeta^{imu}_{k-m,k} = \left(\Delta\mathbf{q}^w_{k-m,k}, \Delta\mathbf{p}^w_{k-m.k}\right)^T$:

$$\Delta\boldsymbol{\xi}^{imu}_{k-m,k} = \sum_{i=k-m}^{k} \begin{bmatrix} \frac{1}{2}\Omega\left(\hat{\boldsymbol{\eta}}^b_{wb}\right)q^w_i & \frac{1}{2}(v^w_{i-1}+v^w_i) \end{bmatrix}^T \Delta t \tag{5}$$

where $\mathbf{v}^w_i$ denotes the velocity in w at time i, $\hat{\boldsymbol{\eta}}^b_{wb}$ denotes the instantaneous angular velocity of B and $\mathbf{q}^w_i$ denotes the quaternions from w to b at time i.

To reflect the motion of the camera, the pre-integration $\Delta\boldsymbol{\xi}^{imu}_{k-m,k}$ needs to align with C:

$$\begin{aligned} &\mathbf{T}(\Delta\boldsymbol{\xi}^{cam}_{k-m,k}) = \mathbf{T}^c_b\mathbf{T}(\Delta\boldsymbol{\xi}^{imu}_{k-m,k})\mathbf{T}^{c\,-1}_b \\ &\mathbf{T}(\Delta\boldsymbol{\xi}^{imu}_{k-m,k}) = \begin{bmatrix} \mathbf{R}^b_{k-m,k} & \mathbf{t}^b_{k-m,k} \\ 0 & 1 \end{bmatrix}, \mathbf{T}(\Delta\boldsymbol{\xi}^{cam}_{k-m,k}) = \begin{bmatrix} \mathbf{R}^c_{k-m,k} & \mathbf{t}^c_{k-m,k} \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{6}$$

where $\mathbf{T}(\Delta\boldsymbol{\xi}^{cam}_{k-m,k})$ denotes the transformation matrix from time $k-m$ to k in c, $\mathbf{T}^c_b$ is the transformation matrix from b to c. $\mathbf{R}^b_{k-m,k}$ is the quaternions $\Delta\mathbf{q}^w_{k-m,k}$ expressed in rotation matrix, $\mathbf{t}^b_{k-m,k} = C^b_w\Delta\mathbf{p}^w_{k-m.k}$ is the translation vector in B, where C^b_w is the rotation matrix from w to b.

After getting the coarse pose estimation of camera $\hat{\mathbf{T}}(\Delta\boldsymbol{\xi}^{cam}_{k-m,k})$, we can predict the camera pose by equation:

$$\hat{\mathbf{T}}(\xi^{cam}_k) = \mathbf{T}(\Delta\boldsymbol{\xi}^{cam}_{k-m,k})\mathbf{T}(\xi^{cam}_{k-m}) = \begin{bmatrix} \hat{\mathbf{R}}(\xi^{cam}_k) & \hat{\mathbf{t}}(\xi^{cam}_k) \\ 0 & 1 \end{bmatrix} \tag{7}$$

For each 3D feature point of current frame, the matched feature points should near it. After predicting the coarse pose estimation, we project each feature point of current frame into the initial camera frame. The search for candidates only in a small range of each 3D feature points in local map. The range depends on the bias and noise of the MEMS-IMU. We do BOW matching between each feature point and its candidates to get matched feature point. Due to the confinement of the region, the error and the time consuming in searching and matching will reduce.

After getting the matched result, bundle adjustment optimization is performed to optimize the camera pose by minimizing the reprojection error between the matched 3D feature points $\mathbf{F}^i \in \mathbb{R}^3$ in map and feature points $\mathbf{f}^i \in \mathbb{R}^3$ in current frame. The $i \in \chi$ is a set of matched points:

$$\{\mathbf{R}, \mathbf{t}\} = \underset{\mathbf{R},\mathbf{t}}{\operatorname{argmin}} \sum_{i \in \chi} \rho(\left\| f^i_{(.)} - \pi_{(.)}(RF^i + t) \right\|^2_{\Sigma}) \tag{8}$$

where the ρ is the robust Huber cost function and $\sum$ is the covariance matrix associated to the scale of feature points, which is one when with stereo-camera. $\pi_{(.)}$ is the projection functions monocular π_m, rectified stereo π_s are defined, as follows:

$$\pi_m\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \begin{pmatrix} f_x\frac{X}{Z} + c_x \\ f_y\frac{X}{Z} + c_y \end{pmatrix}, \pi_s\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \begin{pmatrix} f_x\frac{X}{Z} + c_x \\ f_y\frac{X}{Z} + c_y \\ f_x\frac{X - b}{Z} + c_x \end{pmatrix} \tag{9}$$

where (f_x, f_y) is focal length, (c_x, c_y) is the principal point and b is the baseline, all is known in advanced.

However, the bundle adjustment to minimize the reprojection error is nonlinear. It cannot always get a global optimal point. As shown in Figure 6, VO falls into local optimum easily because the initial iteration point is last frame pose.

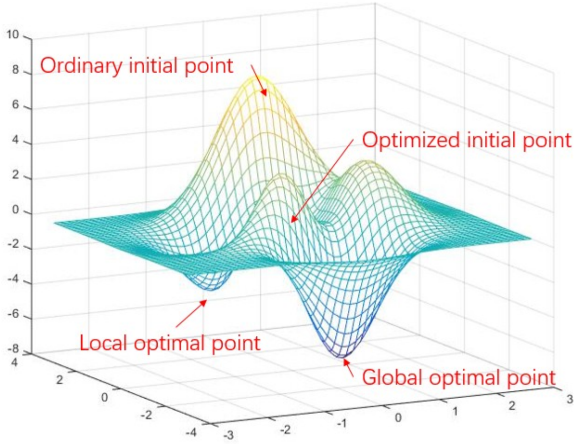


Figure 6. An illustration of association between initial point and result of optimization

In our approach, the initial iteration pose is set as prediction of MEMS-IMU pre-integration $\mathbf{R} = \hat{\mathbf{R}}(\xi^{cam}_k)$ and $\mathbf{t} = \hat{\mathbf{t}}(\xi^{cam}_k)$ to get close to global optimal point. Then, stereo VIO 6DOF pose estimation is optimized in order to avoid local optimum.

2.2.3. Fault-Tolerant Adaptive Extended Kalman Filtering

In this part, the FTAEKF is introduced to tolerant wrong stereo VIO pose estimation limited by the visual principle in a hostile environment.

1. Fault-tolerance with dramatic change detection

In some extreme cases, with fast motion in hostile environment, a large error of VIO pose estimation occurs because of the limited number in matched feature points or similar descriptor. The matcher matches feature points simply depending on the hamming distance.

Therefore, a fault-tolerant method with MEMS-IMU measurements is introduced through dramatic change detection.

One way to detect the sudden step change, by comparing the number of matched points with threshold after eliminating exterior point in bundle adjustment, has been proposed before. However, this is an indirect technique. In some scenario, the number of matched points is large enough, but they mostly matched with wrong feature points and significant estimation error still occurs in this direction. Sudden step change detecting in VIO mostly consider setting a transformation threshold between two consequent frames. They all only detected faults without isolation lead to failure of the system.

In this paper, a new approach using the detection function to detect and isolate dramatic change was proposed. As an accurate pose can be estimated from MEMS-IMU during a short period, the framework considered the MEMS-IMU pre-integration $\hat{\mathbf{T}}(\Delta\boldsymbol{\xi}_{k-m,k}^{cam})$ as a reference. It compares to final relative VIO pose estimation $\mathbf{T}\left(\Delta\xi_{k-m,k}^{cam}\right) = \mathbf{T}\left(\xi_k^{cam}\right)\mathbf{T}\left(\Delta\xi_{k-m}^{cam}\right)^{-1}$ between time k and $k-1$ to detect dramatic change. If the value of detection function $f_{\mathrm{d}} \geq 1$, then the dramatic change detection is deemed to occur. The detection function f_{d} is defined as:

$$\begin{aligned} \Delta\mathbf{T}\left(\Delta\xi_{k-m,k}^{cam}\right) &= \mathbf{T}\left(\Delta\xi_{k-m,k}^{cam}\right)\hat{\mathbf{T}}(\Delta\boldsymbol{\xi}_{k-m,k}^{cam})^{-1} = \begin{bmatrix} \Delta\mathbf{R}_{k-m,k}^{cam} & \Delta\mathbf{t}_{k-m,k}^{cam} \\ 0 & 1 \end{bmatrix} \\ f_{\mathrm{d}} &= \sqrt{\frac{\left(\Delta\mathbf{t}_{k-m,k}^{cam} - \mathbf{t}_{k-m,k}^{imu}\right)^T\left(\Delta\mathbf{t}_{k-m,k}^{cam} - \mathbf{t}_{k-m,k}^{imu}\right)}{E_{\varepsilon t}{}^2} \cdot \frac{\varepsilon\psi_{k-m,k}{}^2 + \varepsilon\theta_{k-m,k}{}^2 + \varepsilon\gamma_{k-m,k}{}^2}{E_{\varepsilon\psi}{}^2 + E_{\varepsilon\theta}{}^2 + E_{\varepsilon\gamma}{}^2}} \end{aligned} \tag{10}$$

where the $\Delta\mathbf{T}\left(\Delta\xi_{k-m,k}^{cam}\right)$ is the transformation difference estimation between pre-integration of MEMS-IMU measurements and VIO. $\varepsilon\psi_{k-m,k}$, $\varepsilon\theta_{k-m,k}$, and $\varepsilon\gamma_{k-m,k}$ are defined as: $\varepsilon\gamma_{k-m,k} = \Delta\gamma_{k-m,k}^{imu} - \Delta\gamma_{k-m,k}^{cam}$, $\varepsilon\theta_{k-m,k} = \Delta\theta_{k-m,k}^{imu} - \Delta\theta_{k-m,k}^{cam}$, and $\varepsilon\psi_{k-m,k} = \Delta\psi_{k-m,k}^{imu} - \Delta\psi_{k-m,k}^{cam}$. Where $\Delta\gamma_{k-m,k}^{imu}$, $\Delta\theta_{k-m,k}^{imu}$, and $\Delta\psi_{k-m,k}^{imu}$ are the incremental relative attitude change estimated by MEMS-IMU measurements, $\Delta\gamma_{k-m,k}^{cam}$, $\Delta\theta_{k-m,k}^{cam}$, and $\Delta\psi_{k-m,k}^{cam}$ are the incremental relative attitude change estimated by VIO.

The threshold $E_{\varepsilon t}$, $E_{\varepsilon\psi}$, $E_{\varepsilon\theta}$, and $E_{\varepsilon\gamma}$ are set up according to the drift of motion estimation by prediction using MEMS-IMU during one period of slam procedure, which is from discrete time $k-m$ to k. As a more reliable pose can be estimated from MEMS-IMU during a short period of time, the transformation difference estimation between MEMS-IMU prediction and stereo VIO system estimation should be within this range.

In consideration of the drift of estimation by MEMS-IMU, the threshold $E_{\varepsilon t}$, $E_{\varepsilon\psi}$, $E_{\varepsilon\theta}$, and $E_{\varepsilon\gamma}$ change adaptively. As continuous change detected in hostile environment increases, $E_{\varepsilon t}$, $E_{\varepsilon\psi}$, $E_{\varepsilon\theta}$, and $E_{\varepsilon\gamma}$ are growing. $E_{\varepsilon t}$, $E_{\varepsilon\psi}$, $E_{\varepsilon\theta}$, and $E_{\varepsilon\gamma}$ are to be reinitialized with the original value if no environmental transition is detected.

2. Covariance adaptive filtering

Due to the change and accumulation of error in each process of pose estimation from VIO, the observation covariance from VIO is set to dynamic dependent upon the distance and motion characteristics to achieve better positioning accuracy. The observation covariance is adjusted to better represent practical situations.

VIO is a dead-reckon algorithm in which the error of stereo VIO pose estimation is accumulated by distance. A factor λ_d, related to the distance of stereo VIO d^{cam} reflect the error accumulating is introduced:

$$\begin{aligned} d^{cam} &= \sum_{i=1}^{k-1}\sqrt{t\left(\Delta\xi_{i,i+1}^{cam}\right)^T t\left(\Delta\xi_{i,i+1}^{cam}\right)} \\ \lambda_d &= \sigma d^{cam} \end{aligned} \tag{11}$$

where $\mathbf{t}\left(\xi_{i,i+1}^{cam}\right)$ is the camera translation vector between time k and $k+1$ in C, σ is dependent on characteristics of the stereo VIO system.

Besides, the precision of stereo VIO pose estimation is also influenced obviously by motion characteristics. The field of view changes fast and the same feature points are reduced speedily when great angular change is made in a short time. MEMS-IMU measurements are more suitable and precise for the estimation and VIO is no longer reliable. Thus, a factor λ_a is introduced to adapt the specialties of MEMS-IMU and stereo VIO.

$$\lambda_a = \sum_{i=k-n}^{k} \sqrt{\hat{\boldsymbol{\eta}}_{wb,i}^{b}{}^{T}\hat{\boldsymbol{\eta}}_{wb,i}^{b}} \tag{12}$$

where $\hat{\boldsymbol{\eta}}_{wb,i}^{b}$ is $\hat{\boldsymbol{\eta}}_{wb}^{b}$ at time i, n is the size of the slide window.

When filtering, the error state vector used to correct the predicted state in filter is defined as follows:

$$\delta\mathbf{X} = \left(\delta\mathbf{q}^{w}, \delta\mathbf{p}^{w}, \delta\mathbf{v}^{w}, \delta\boldsymbol{\beta}_{g}^{b}, \delta\boldsymbol{\beta}_{a}^{b}\right)^{T} \tag{13}$$

where, $\delta\mathbf{X}$ is the state vector composed by quaternions, position, velocity, and bias error.

With no dramatic change detecting in perceived environment, the predicted states are corrected by measurements information obtained from stereo VIO pose estimation. As no drift pitch or roll angle can be obtained through gravity correction, the observation model in proposed FTAEKF is as follows:

$$\begin{aligned}
&\mathbf{Z}_k = \mathbf{H}_k\boldsymbol{\delta}\mathbf{X}_k + \boldsymbol{\mu}_k \\
&\boldsymbol{\mu}_k = \left[\begin{array}{cccc} \lambda_d\varepsilon_{px}^{r} & \lambda_d\varepsilon_{py}^{r} & \lambda_d\varepsilon_{pz}^{r} & \lambda_a\varepsilon_{\psi}^{r} \end{array}\right]^{T} \\
&\mathbf{Z}_k = \left(\widetilde{x}_k^{w} - \overline{x}_k^{w}, \widetilde{y}_k^{w} - \overline{y}_k^{w}, \widetilde{z}_k^{w} - \overline{z}_k^{w}, \widetilde{\psi}_k^{w} - \overline{\psi}_k^{w}\right)^{T} \\
&\overline{\psi}_k^{w} = \tan^{-1}\left(\frac{2\left(q_{1,k}^{w} * q_{2,k}^{w} + q_{0,k}^{w} * q_{3,k}^{w}\right)}{1 - 2\left(q_{2,k}^{w} * q_{2,k}^{w} + q_{3,k}^{w} * q_{3,k}^{w}\right)}\right) \\
&\mathbf{H}_k = \left[\begin{array}{cccccc} \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times9} \\ \dfrac{\partial\overline{\psi}_k^{w}}{\partial q_{o,k}^{w}} & \dfrac{\partial\overline{\psi}_k^{w}}{\partial q_{1,k}^{w}} & \dfrac{\partial\overline{\psi}_k^{w}}{\partial q_{2,k}^{w}} & \dfrac{\partial\overline{\psi}_k^{w}}{\partial q_{3,k}^{w}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times9} \end{array}\right]
\end{aligned} \tag{14}$$

where $\mathbf{Z}_k$ is the observation, $\widetilde{x}_k^{w}$, $\widetilde{y}_k^{w}$, $\widetilde{z}_k^{w}$, and $\widetilde{\psi}_k^{w}$ are the observation position and yaw in the world frame from the stereo VIO pose estimation, respectively, $\overline{x}_k^{w}$,$\overline{y}_k^{w}$, $\overline{z}_k^{w}$, and $\overline{\psi}_k^{w}$ are the predicted position and yaw in the world frame from IMEMS-MU mechanization, respectively, $\mathbf{H}_k$ is the observation matrix and $\boldsymbol{\mu}_k$ is the observation noise, which is adaptive.

When dramatic change occurred, MEMS-IMU measurements pre-integration will be used as pose estimation to isolate and tolerate fault. Since the pose estimated with MEMS-IMU during a short period of time is with sufficient accuracy, the stereo VIO system is reinitialized based on the MEMS-IMU pose in W at the closest time. The λ_a and λ_d is also reinitialized. That makes the framework with the ability to navigate even when stereo VIO system failed.

After filtering, the new matched feature points are projected to initial c to update the local map. The position of the same feature is represented using the average of position value.

When the dramatic change is detected, the local map points are cleared and the initial pose is set to MEMS-IMU pose in w with the closest time.

3. Results

3.1. Experiment Setup

3.1.1. Equipment

The equipment that we employed was based on commercial off the shelf shown in Figure 7. It consists of a ZED stereo camera, a Xsens MTI-G-710 MEMS-IMU, and a NVIDIA Jetson TX2. The ZED stereo camera resolution is set to 1280 × 720, baseline is 12cm and the frame rate at 15 HZ.

The Xsens MTI-G-710 can measure the acceleration and angular velocity in body frame running at 200 HZ. The MEMS-IMU was mounted on left camera of ZED that was calibrated in advanced. The processing platform is NVIDIA Jetson TX2 with dual-core NVIDIA Denver2 and quad-core ARM Cortex-A57 running on Ubuntu 16.04. The Novatel OEM6 GPS receiver worked with GPS-RTK running at 1HZ as outdoor reference. All of the sensors were connected with TX2 through USB cable and the implementation is based on C++ with Robot Operating System (ROS) Kinetic. The sensors are mounted on a tripod with three rollers.

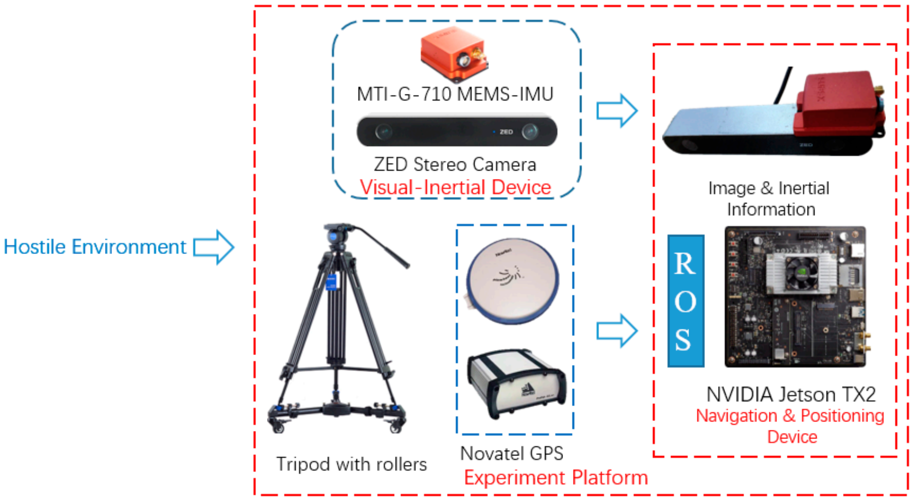


Figure 7. An illustration of platform. It consisted of ZED camera, MTI MEMS-IMU, Novatel GPS and Jetson TX2.

3.1.2. Experiment Environment Description

In order to evaluate the performance of the proposed method under a hostile environment, the experiments were carried out in the corridor outside the laboratory and a tennis court in campus, as shown in Figures 8 and 9. For the corridor, the wall of the corridor was sparse-feature. The make part of descriptors were similar. Ambient lighting in the corridor is unsatisfactory in some places, as it is bright near the window but is considerably darker elsewhere. The corridor plan is known in advance with the floor that consisted of fixed size tiles. Each tile is a square with sides of 60 cm. We pushed the tripod along the tile edge and obtained the ideal trajectory reference through a corridor plan. Some artificial mark points located at door and corner have been set in advance to evaluate the performance more comprehensively. It is regarded as the ideal path to evaluate the performance of the proposed framework. The yaw angle of MTI that was fused with magnetic is regarded as yaw angle reference. For the tennis court, the color of the ground was also simple and surrounded by similar meshes. The outdoor distance of feature was far beyond indoor environment. The reference of pose was obtained through GPS-RTK. Both environments can be considered as the hostile environment.

3.2. Experiments Results

We carried out a semi-physical simulation experiment to verify the performance of our proposed framework. The data was collected with the equipment and processed in platform. The proposed framework is compared against ORB-SLAM2, MSF-EKF [22], and VINS-Mono [23] in the experiments. The MSF-EKF based on the modular-sensor fusion framework by the University of Zurich is widely used to loosely couple inertial information and visual information. Moreover, the tightly-coupled VINS-Mono is high-performance and robust by the Hong Kong University of Science and Technology. Because the methods was multi-threaded and contained some random processing, the data took the 3σ bounds of results to eradicate any discrepancies.

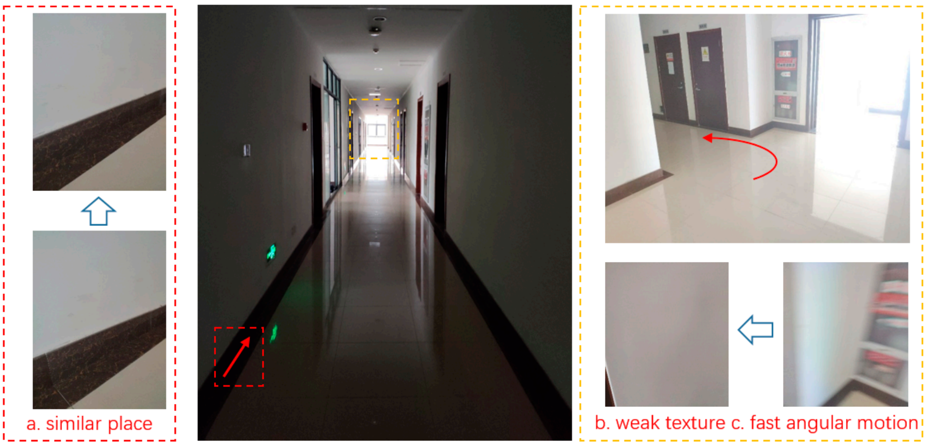


Figure 8. An illustration of the corridor where experiment carried on.

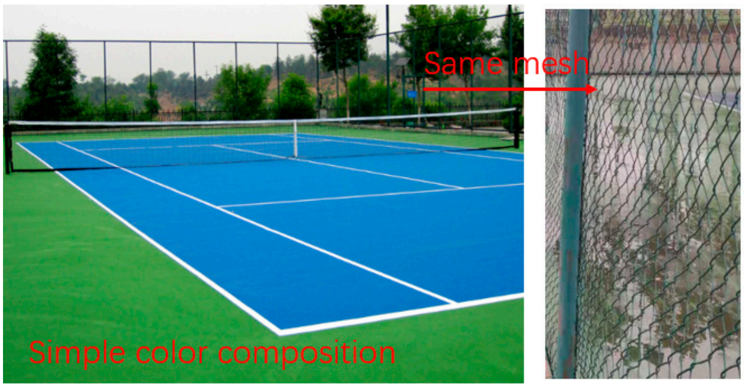


Figure 9. An illustration of the tennis court where experiment carried on.

3.2.1. Experiment I: In Corridor

In experiment I, we pushed the tripod along the tile edge in the corridor. The experiment intended to assess the comprehensive performance of the proposed framework in an indoor hostile environment.

The red line is the ideal trajectory, as shown in Figure 10. The time at passing the mark points was recorded. The estimation of motion and yaw angle from different methods shown in Figure 11a,b. The position is projected onto X-Y plane. It was clear to see our proposed method achieved more accurate pose estimation. In addition, the value of fault illustrated seven dramatic changes that were detected by FTAKF in the experiment I in Figure 12a and the adaptive observation covariance is shown in Figure 12b. Moreover, the value of mean error and root mean square error (RMSE) of yaw angle and motion estimation from different methods, as shown in Figures 13 and 14.

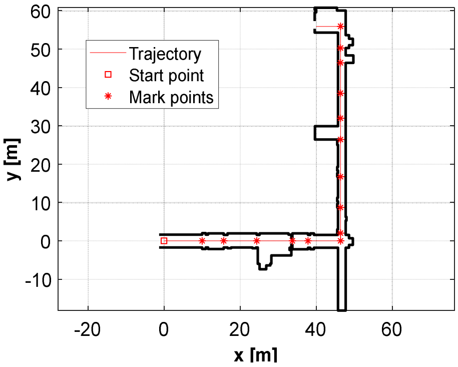
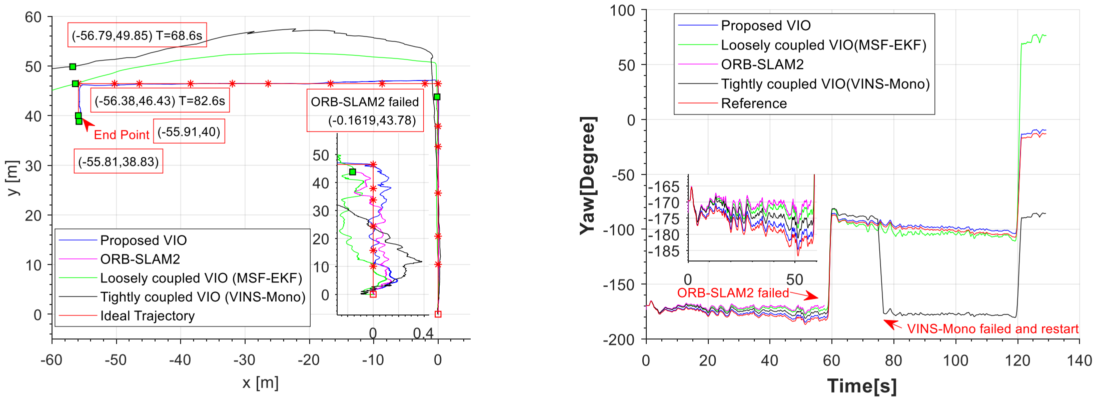
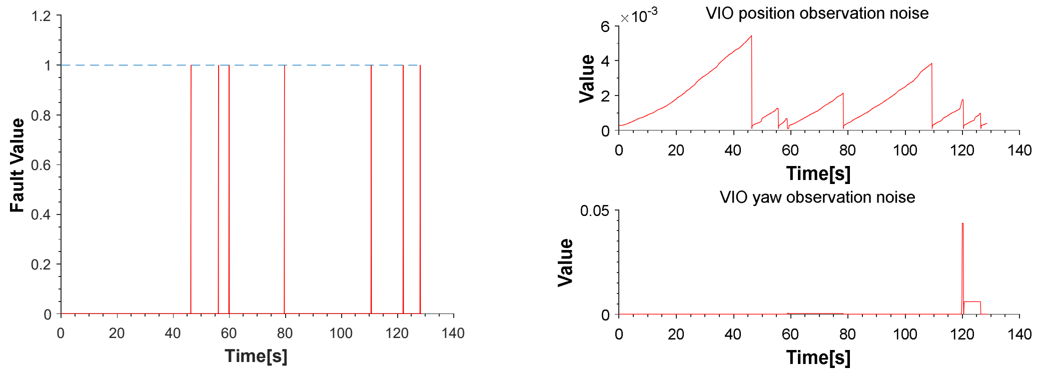


Figure 10. An illustration on the corridor plan, the ideal trajectory, and markers.



(**a**) (**b**)

Figure 11. (**a**) An illustration of motion estimation results from different methods. (**b**) An illustration of Yaw angle estimated by different methods. ORB-SLAM2 failed due to few feature points and the noise of MEMS-IMU propagated speedily without measurements. The MEMS-IMU was meeting a corner causing fast angular velocity at 120 s. The noise of the gyroscopes propagated more speedily that causing sudden change in yaw angle difference with MSF-EKF.



(**a**) (**b**)

Figure 12. (**a**) The value of fault detect function demonstrates the dramatic change. (**b**) An illustration of the value of position and yaw observation noise.

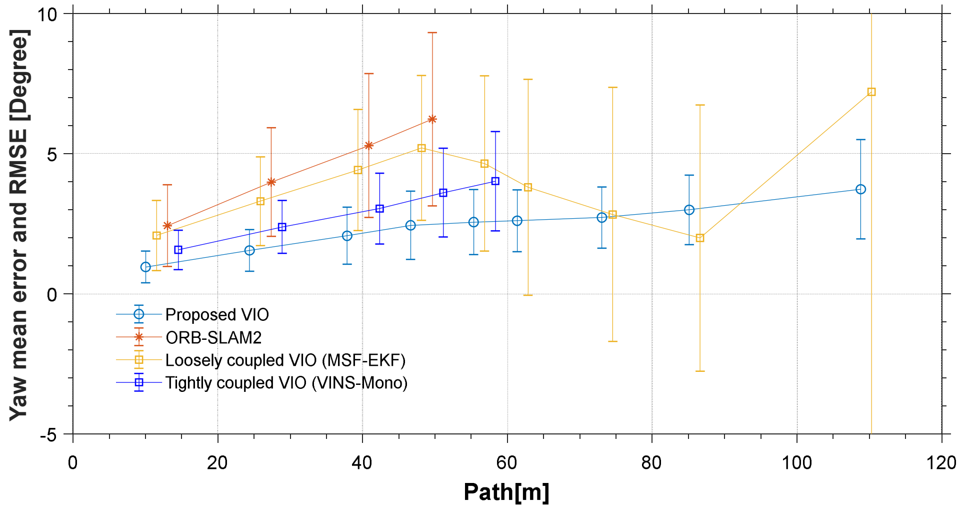


Figure 13. An illustration of value of yaw angle mean and RMSE from different methods. (VINS-Mono without output before initialization and value after system failing are ignored).

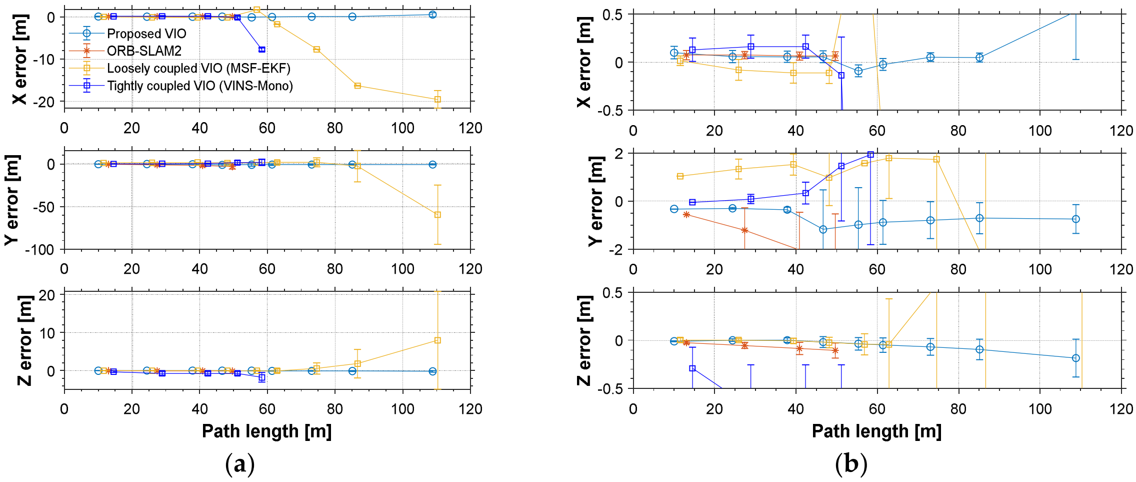


Figure 14. (**a**) The value of mean error and root mean square error (RMSE) of motion estimation from different methods. (**b**) An illustration of partial enlargement.

3.2.2. Experiment II: In tennis court

In experiment II, we pushed the tripod along the edge of the tennis court. The experiment intended to evaluate the performance of the proposed framework in an outdoor hostile environment under the RTK position and heading reference.

The red line is RTK trajectory as shown in Figure 15 with time synchronized through ROS. The estimation of motion and yaw angle from different methods shown in Figure 15a,b. Our proposed method achieved more accurate pose estimation. The value of fault illustrated six dramatic changes was detected by FTAKF in the experiment II in Figure 16a and the adaptive observation covariance is shown in Figure 16b. The value of mean error and RMSE of yaw angle and motion estimation from different methods shown in Figures 17 and 18.

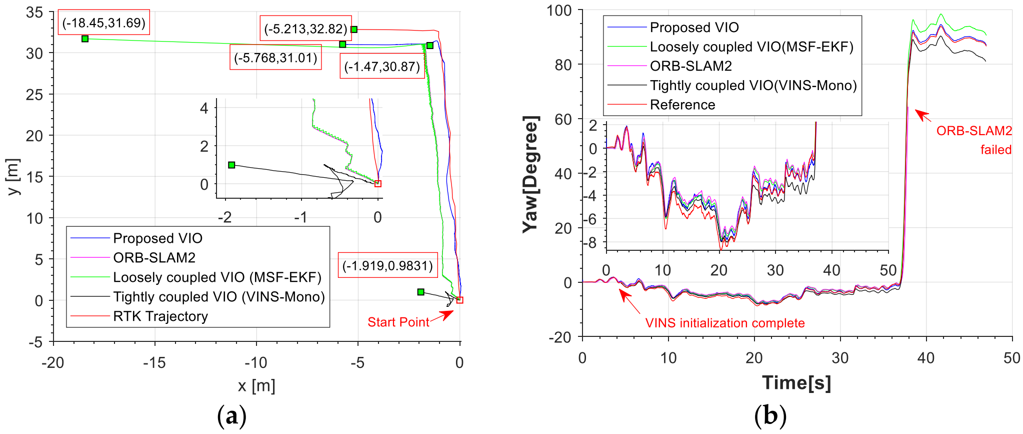


Figure 15. (**a**) An illustration of motion estimation results from different methods (**b**) An illustration of yaw angle estimated by different methods. (VINS-Mono without output before initialization).

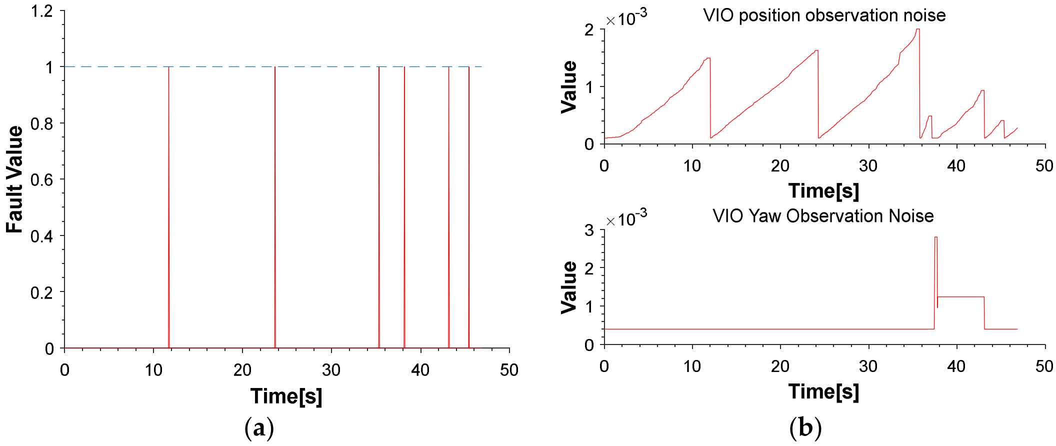


Figure 16. (**a**) The value of fault detect function demonstrates the dramatic change. (**b**) An illustration of the value of position and yaw observation noise.

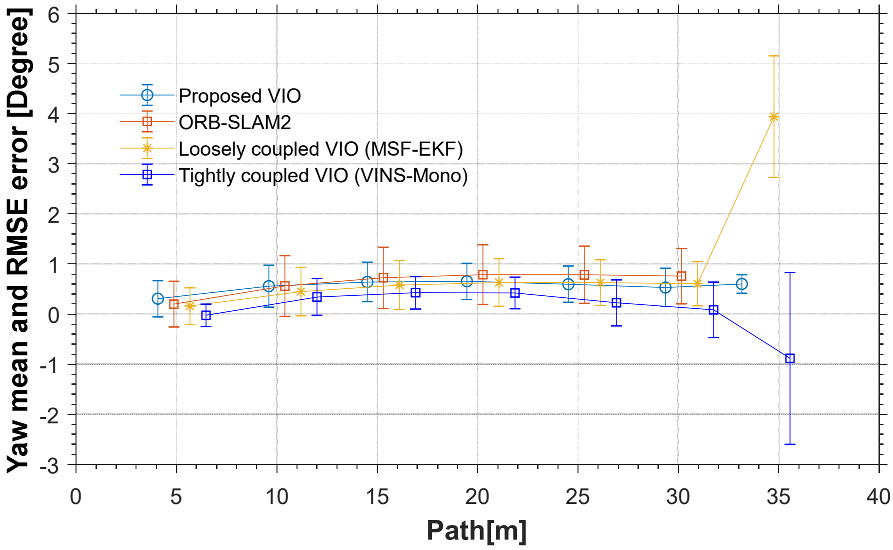


Figure 17. An illustration of value of yaw angle mean and RMSE from different methods. VINS-Mono without output before initialization.

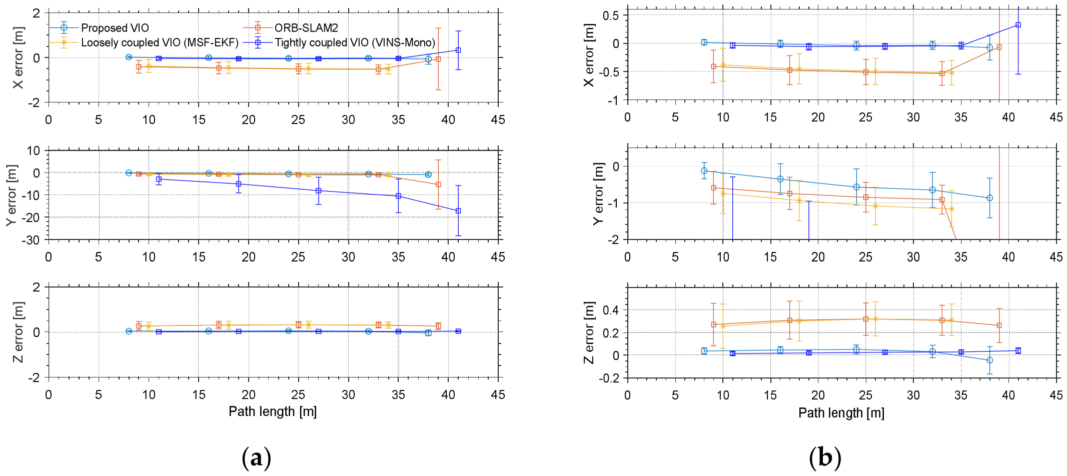


Figure 18. (**a**) The value of mean error and RMSE error of motion estimation from different methods. (**b**) An illustration of partial enlargement.

3.3. Experimental Analysis

3.3.1. Accuracy Analysis

In the experiments, the accuracy of the proposed algorithm in the reconstructed trajectory is calculated as the RMSE with mark points and RTK references in Tables 1 and 2. Moreover, the Euclidean distance between the last position of the estimated camera trajectory and the expected end point were calculated in Tables 3 and 4. Value marked with an asterisk (*) was obtained before failure.

Table 1. RMSE (m) of motion estimation in different methods. (Value marked with an asterisk (*) was obtained before VO failure.)

Length (m)	Proposed Error	ORB-SLAM2 Error	MSF-EKF Error	VINS-Mono Error
Experiment I: 108.8	0.43(0.58 *)	0.94 *	16.57 (0.90 *)	1.80 *
Experiment II: 38	0.6(0.53 *)	0.75 *	3.94 (0.6 *)	0.88 (0.08 *)

Table 2. RMSE (°) of yaw angle estimation in different methods. (Value marked with an asterisk (*) was obtained before VO failure.)

Yaw Angle Change (°)	Proposed Error	ORB-SLAM2 Error	MSF-EKF Error	VINS-Mono Error
Experiment I: 180	4.52 (2.9 *)	3.13 *	21.84 (3.10 *)	3.0 *
Experiment II: 90	0.19 (0.38 *)	0.55 *	1.21 (0.44 *)	1.72 (0.56 *)

Table 3. Length accuracy (m).

Length (m)	Proposed Error	ORB-SLAM2 Error	MSF-EKF Error	VINS-Mono Error
Experiment I: 108.8	0.92, 0.8%	194.3, 179.9%	55.88, 51.4%	67.36, 67.4%
Experiment II: 38	1.89, 4.98%	4.22, 11.1%	13.3, 35.0%	32.0, 84.2%

Table 4. Yaw angle accuracy (°).

Yaw Angle Change (°)	Proposed Error	ORB-SLAM2 Error	MSF-EKF Error	VINS-Mono Error
Experiment I: 180	1.8, 1%	176.3, 97.9%	68.5, 38.1%	62.3, 34.6%
Experiment II: 90	0.37, 0.4%	22.17, 25%	3.98, 4.04%	5.95, 6.61%

The accuracy for the experiments was depicted in above tables. The true length of different trajectories is, respectively, 108.8 m and 38 m, and the changes of reference yaw angle are 180° and 90°. As shown in Figures 11 and 15, the stereo-camera and MEMS-IMU experienced different motions with smooth motion, fast rotational, and translational motion of indoor and outdoor. As both mean error and root mean square error of ORB-SLAM2, MSF-EKF, and VINS-Mono were larger than the proposed method in hostile environment. It is clearly seen that the estimated results from the proposed method in Experiment I and II were more accurate and robust than those from ORB-SLAM2, MSF-EKF, and VINS-Mono in Figures 13, 14, 17 and 18. Pose estimation of both VO and VIO without fault tolerance were failed or divergent, which may cause fatal problems in robot navigation.

3.3.2. Inertial Aided Matching and Fault Tolerance Analysis

Figures 11 and 15 shows the pose estimation of two experiments from four different methods. ORB-SLAM2, MSF-EKF, and VINS-Mono produced large error in both position and yaw angle estimation under hostile environments. During experiments, systems including ORB-SLAM2 and VINS-Mono were in poor performance due to few feature or similar feature in hostile environment.

Moreover, ORB-SLAM2 failed because the number of feature points at corner lower than threshold. The failure of ORB-SLAM2 also caused divergence of MSF-EKF without VO output as measurement.

With the number of feature points decreasing, the part of cost function occupied by each feature points was increasing. In addition, the influence of mismatch was increased, resulting in the divergence of a system. VINS-Mono failed by detecting much large translation between two frames in experiment I. For experiment II, the feature points in starting position of tennis court were too similar and far to produce enough disparity between two consequent frames. This situation caused the error in direction of x axis with ORB-SLAM2 and false initialization with VINS-Mono which tracking feature points through optical flow method.

The pre-integration of measurements of MEMS-IMU could constrain the region of matching to reduce incorrect candidate points that achieve better match result, as shown in Figure 19. Besides, the dramatic changes was detected shown in Figures 12 and 16, were isolated in the proposed framework that able to navigate properly in hostile environment. In addition, the adaptive noise of measurements shown in Figures 12 and 16 make the proposed framework obtained more accurate pose estimation than traditional loosely-coupled VIO, such as MSF-EKF.

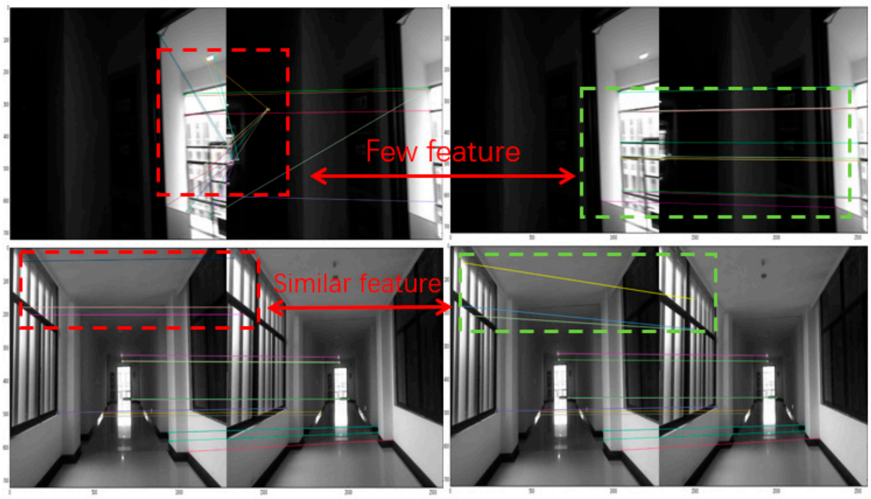


Figure 19. An illustration of wrong matching in hostile situation. Left image represented matching with all feature points in references frame and right confined matching by pre-integration.

4. Conclusions

In this work, a novel fault-tolerant framework with stereo-camera and MEMS-IMU was proposed to obtain robust and precise positioning information in a hostile environment. MEMS-IMU measurements predict the camera motion and adaptive observation covariance noise are taken in the framework. It makes stereo VO motion estimation more precise when meeting hostile environment. A fault-tolerant mechanism is also introduced to detect and isolate the dramatic change in order to achieve more robust positioning information.

When comparing to traditionally loosely-coupled VIO systems that are not considered to detect the wrong measurements, our proposed method introduced an adaptive noise according to motion characteristics that obtain more precise positional information. For the tightly-coupled VIO systems, which introduced inertial error to obtain more robust and accurate positioning results, the relation between inertial error and visual error is not considered, which leads to the influence of inertial error estimation after the error of visual matching, resulting in the instability of the whole system. Our proposed framework isolated visual error, which was detected by comparing with more reliable inertial error, made the whole system more reliable and stable. The framework also maintains a certain degree of independence between framework and stereo VO system that can be easily integrated with other stereo VO system. By evaluating the results of experiments, the proposed VIO system has achieved a satisfactory performance in state estimation in a hostile environment.

In our future work, we hope to apply the inertial information to graph-pose optimization in order to realize the function of loop detection and optimization in hostile environment. We also hope to employ the method in more challenging environments.

Author Contributions: C.Y. and P.S. proposed the original idea and wrote this paper; W.Z. and P.L. performed the experiments, analyzed the data; J.L. and K.H. participated in design of the experimental demonstration, revised the paper and gave some valuable suggestions.

Funding: This research was funded by [Jiangsu provincial SixTalent Peaks] grant number [2015-XXRJ-005], [Jiangsu Province Qing Lan Project], [National Natural Science Foundation of China] grant number [61703207], [Jiangsu Provincial Natural Science Foundation of China] grant number [BK20170801], and [Aeronautical Science Foundation of China] grant number [2017ZC52017].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Z.; El-Sheimy, N.; Yu, C.; Qin, Y.; Liu, Z.; El-Sheimy, N.; Yu, C.; Qin, Y. Motion Constraints and Vanishing Point Aided Land Vehicle Navigation. *Micromachines* **2018**, *9*, 249. [CrossRef] [PubMed]
2. Weiss, S.; Achtelik, M.W.; Lynen, S.; Chli, M.; Siegwart, R. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments. In Proceedings of the IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 957–964.
3. Nister, D.; Naroditsky, O.; Bergen, J. Visual odometry. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 27 June–2 July 2004; Volume 1, pp. I-652–I-659.
4. Usenko, V.; Engel, J.; Stückler, J.; Cremers, D. Direct visual-inertial odometry with stereo cameras. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 1885–1892.
5. Vidal, A.R.; Rebecq, H.; Horstschaefer, T.; Scaramuzza, D. Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. *IEEE Robot. Autom. Lett.* **2018**, *3*, 994–1001. [CrossRef]
6. Corrêa, D.; Santos, D.; Contini, L.; Balbinot, A. MEMS Accelerometers Sensors: An Application in Virtual Reality. *Sens. Transducers Tor.* **2010**, *120*, 13–26.
7. Wang, J.; Zeng, Q.; Liu, J.; Meng, Q.; Chen, R.; Zeng, S.; Huang, H. Realization of Pedestrian Seamless Positioning Based on the Multi-Sensor of the Smartphone. *Navig. Position. Timing* **2018**, *1*, 28–34. [CrossRef]
8. Tian, Y.; Chen, Z.; Lu, S.; Tan, J. Adaptive Absolute Ego-Motion Estimation Using Wearable Visual-Inertial Sensors for Indoor Positioning. *Micromachines* **2018**, *9*, 113. [CrossRef] [PubMed]
9. Mur-Artal, R.; Tardos, J.D. Visual-Inertial Monocular SLAM with Map Reuse. *IEEE Robot. Autom. Lett.* **2017**, *2*, 796–803. [CrossRef]
10. He, Y.; Zhao, J.; Guo, Y.; He, W.; Yuan, K.; He, Y.; Zhao, J.; Guo, Y.; He, W.; Yuan, K. PL-VIO: Tightly-Coupled Monocular Visual–Inertial Odometry Using Point and Line Features. *Sensors* **2018**, *18*, 1159. [CrossRef] [PubMed]
11. Sun, K.; Mohta, K.; Pfrommer, B.; Watterson, M.; Liu, S.; Mulgaonkar, Y.; Taylor, C.J.; Kumar, V. Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight. *IEEE Robot. Autom. Lett.* **2018**, *3*, 965–972. [CrossRef]
12. Tardif, J.P.; George, M.; Laverne, M.; Kelly, A.; Stentz, A. A new approach to vision-aided inertial navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 4161–4168.
13. Liu, Y.; Xiong, R.; Wang, Y.; Huang, H.; Xie, X.; Liu, X.; Zhang, G. Stereo Visual-Inertial Odometry With Multiple Kalman Filters Ensemble. *IEEE Trans. Ind. Electron.* **2016**, *63*, 6205–6216. [CrossRef]
14. Schmid, K.; Lutz, P.; Tomić, T.; Mair, E.; Hirschmüller, H. Autonomous Vision-based Micro Air Vehicle for Indoor and Outdoor Navigation. *J. Field Robot.* **2014**, *31*, 537–570. [CrossRef]
15. Mourikis, A.I.; Roumeliotis, S.I. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigatio. Proceedings IEEE International Conference on Robotics and Automation, Roma, Italy, 10–14 April 2007; pp. 3565–3572.
16. Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D. On-Manifold Preintegration for Real-Time Visual–Inertial Odometry. *IEEE Trans. Robot.* **2017**, *33*, 1–21. [CrossRef]

17. Ramezani, M.; Khoshelham, K. Vehicle Positioning in GNSS-Deprived Urban Areas by Stereo Visual-Inertial Odometry. *IEEE Trans. Intell. Veh.* **2018**, *3*, 208–217. [CrossRef]
18. Kümmerle, R.; Grisetti, G.; Strasdat, H.; Konolige, K.; Burgard, W. G2o: A general framework for graph optimization. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3607–3613.
19. Wielicki, B.A.; Barkstrom, B.R.; Harrison, E.F.; Lee, R.B.; Smith, G.L.; Cooper, J.E. Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bull. Am. Meteor. Soc.* **1996**, *77*, 853–868. [CrossRef]
20. Alismail, H.; Kaess, M.; Browning, B.; Lucey, S. Direct Visual Odometry in Low Light Using Binary Descriptors. *IEEE Robot. Autom. Lett.* **2017**, *2*, 444–451. [CrossRef]
21. Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [CrossRef]
22. Lynen, S.; Achtelik, M.W.; Weiss, S.; Chli, M.; Siegwart, R. A robust and modular multi-sensor fusion approach applied to MAV navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 3923–3929.
23. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *arXiv* **2017**, arXiv:1708.03852. [CrossRef]