MDPI

Article

Fast and Accurate Initialization for Monocular Vision/INS/GNSS Integrated System on Vehicle

Ronghe Jin [1,†,‡], Firstname Lastname [1,‡] and Firstname Lastname [2,*]

[1] School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; 773792173@qq.com
[2] Affiliation 2; e-mail@e-mail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials) +xx-xxxx-xxx-xxxx (F.L.)
† Current address: Affiliation 3
‡ These authors contributed equally to this work.

Version June 14, 2021 submitted to Journal Not Specified

Abstract: The integration of camera, inertial measurement units (IMU), and Global Navigation Satellite System (GNSS) has a great potential for localization and navigation. Due to the nonlinearity of the coupling system, the performance of the integrated system is highly dependent on the accuracy of its initial values including navigation states, gravity, IMU biases, and visual scale. In this paper, we propose a novel approach to achieve a fast and accurate initialization for the integrated system, and evaluate its capacities with six vehicle tests in different illumination and obstructions scenarios. With the outputs from GNSS RTK (real-time kinematic) algorithm, the rough roll, pitch, yaw, and the direction of gravity are calculated. Based on the known initial states and gravity, the GNSS/INS fusion and Visual-Inertial Odometry (VIO) are simultaneously launched whenever the observations are available. Subsequently, a coarse scale is retrieved by assigning the initial states of GNSS/INS fusion to VIO procedure. This scale, rather than the arbitrary scale set in most VIO or SLAM frameworks, can be utilized to get poses and points very close to the real world. Thus, the points that are too distant from the camera and may have a larger noise than nearby ones are treated as outliers and removed in our method. Once the VIO estimation converges, the transformation parameters between the poses of VIO and GNSS/INS fusion are estimated via non-linear optimization aiming to align these two trajectories and evaluate the scale error. Through the six vehicle tests in different scenarios, we compare the performance of the proposed approach with that of the state-of-the-art VIO algorithm VI ORB-SLAM2 and VINS-Mono.

Keywords: Initialization; Multi-sensor integration; GNSS/INS fusion; VIO; Convergence; Alignment.

1. Introduction

There has been increasing demands for accurate ego-motion estimation in the field of mobile robots, such as unmanned aerial vehicles (UAVs) and autonomous driving. In order to achieve 6 degrees-of-freedom (DOF) pose and three-dimensional (3D) velocity of a robot in real time, various sensors such as GNSS (including GPS, BDS, GLONASS and GALILEO), inertial measurement units(IMU, also known as Inertial Navigation System, INS), and cameras have been used [1]. GNSS, the only global high-precision and high-stability infrastructure providing Positioning, Navigation, and Timing (PNT) [2], can estimate the global coordinates in ECEF frame without error accumulation, IMU can get high-rate measurements of acceleration and angular velocity, images contain rich features, which can be used to construct the environment, and Visual SLAM may always be very precise. As many sensors are mutually complementary [3], multi-sensor integration such as

GNSS/INS coupled navigation, monocular visual-inertial odometry(VIO), GNSS/SLAM integration [4], Visual-Inertial-Magnetic combination [5], has been active research topics in the last two decades. The coupling of INS and GNSS offers an accurate and complete measurement of the robot's states, and thus has been applied in many scenarios of navigation [6]. Monocular visual–inertial SLAM systems have gained much attention in the field of robot navigation and augmented reality [7]. Some state-of-the-art optimization-based systems are VINS-Mono [8], OKVIS [9], and VI ORB-SLAM2 [10]. It is believed that camera, INS, GNSS module will become standard configuration for most UAVs and autonomous vehicles, since such sensors are getting cheaper, smaller, lower in power consumption. Motivated by the purpose of robust, drift-free pose estimation in large-scale autonomous navigation, Vision/INS/GNSS coupled system (also known as GNSS-aided VIO, or Vision-aided GNSS/INS integration) has been a trending topic in recent years [11]. Early works in [12–16] carry out the estimation through filter-based methods, which only update the last state, global position measurements were first aided for VIO with a pose-graph optimization estimator in [17]. Yu et al. [18] put forward a tightly-coupled sliding window optimization for visual and inertial measurements with loosely-coupled GPS refinement. Differently from [18], the work in [11] tightly couple the GPS using the IMU pre-integration algorithm to efficiently derive the global positional factors, which allows adding multiple global factors per keyframe in the sliding window with insignificant extra computational cost.

单个缺点，两两组合缺点

三个优点，精度、鲁棒性、无漂移、低成本

三组合

Due to the nonlinearity of the integrated systems, a good initialization can provide linearization points close to the optimal solution, resulting in faster convergence of the optimization and low risk of a local minimum [19]. The initialization of GNSS/INS integration is the alignment of the computational navigation frame and the local level navigation frame, and the core part of the initial alignment is to determine the rotation and translation between the body frame and reference navigation frame. For the low-cost INS, the noise threshold of gyros is near or higher than the Earth's rotation rate($\approx 15^\circ/h$). As a result, roll and pitch determination can be obtained with stationary accelerometer measurements, however, yaw can not be determined. In this respect, the GNSS-aided velocity matching alignment is applicable to determine the initial state coupling GNSS and low-cost imu while in motion [20], but it will take tens of seconds to converge and may suffer from large initial attitude estimation error [21]. The initial value of accelerometer and gyroscope bias are set zeros, and they will converge with a long period of time, which depends on the INS noise level and the GNSS accuracy. For the widely researched VIO, the initialization task includes estimation of the gravity direction, imu bias and monocular scale, which makes it a more challenging problem. Since pose estimation problem for visual-inertial systems may not have a unique solution depending on the types of motion [22], Bootstrapping a VIO system requires careful treatment, to avoid incorrect system parameters breaking the system. The work in [23] presents a deterministic closed-form solution for computing the gravity orientation and the visual scale of a VIO system, but the system is unstable due to the lacks of estimation for accelerometer and gyroscope biases. Similar to [23], the gyroscope bias in the initialization procedure of [22,24] is neglected, this kind of works above will lead to inaccurate initial states. The researchers of [25] present a re-initialization and failure recovery algorithm assumes that the UAVs should keep horizontally at the beginning, to accomplish the initialization process. The method proposed in [26] requires that the initial attitude to be aligned with the gravity direction. Based on these early studies, pioneering works are proposed in [10] and [27]. The initial estimation in [10] compute the scale, gravity direction, velocity and IMU biases for the visual-inertial full BA with a prior processing of keyframes by a monocular SLAM module. The work in [27] also estimate the scale, gravity direction, velocity and gyroscope bias, but it ignores the accelerometer bias to ensure fast convergence, resulting in a lot of resources cost to refine the gravity and scale. As pointed out in [22], the gravity and accelerometer bias are difficult to be perfectly distinguished without sufficient excited motion, the initialization systems in [10,27] may easily get ill-conditioned, leading to incorrect state solutions. To avoid mixing gravity and accelerometer bias, Li et al. [28] propose to use the detected vertical edges to estimate a better gravity. However, in order to obtain effective edges which are parallel to gravity, some human-made structures

are needed. A disjoint visual-inertial initialization approach proposed in [29], which regards the problem as an inertial-only one, and takes the probabilistic model of IMU noises, while the inertial-only optimization relies on a precise vision-only MAP estimation up-to-scale. In terms of Vision/INS/GNSS integration, researches on initialization are rare. Initialization methods of both monocular and stereo visual-inertial system aided by GPS are introduced in [18], the VIO initialization procedure follows the same one as [27], which ignores accelerometer bias. The initialization method of [16] is similar to [18]. The alignment aim to compute the rotation from world frame(defined by VIO) to global frame based on the assumption that there is only yaw rotation between such two frames.

In this paper, a novel initialization scheme for Mono Vision/INS/GNSS integration is designed. Firstly, GNSS RTK measurements are used to calculate the initial navigation states, thus, the direction of gravity is determined by initial altitude. Then, GNSS/INS fusion thread can be launched to estimate the navigation states under global frame. Hence, we can assign initial pose of VIO with GNSS/INS fusion, to constrain the orientation of VIO under global frame as well. Since vision SFM retrieves poses with arbitrary scale, a coarse VIO scale can be calculated by its assigned poses. Once VIO estimation converges, the alignment between trajectories of VIO and GNSS/INS fusion is conducted, and the scale error is evaluated as well.

The remaining of the research is organized as follows. In Section 2, the mathematical preliminaries of Vision/INS/GNSS integration and the proposed initialization approach are presented. In Section 3, the tests in various scenarios and the experimental results are described in details. At last, we draw a conclusion and present the future prospects in Section 4.

2. Methods

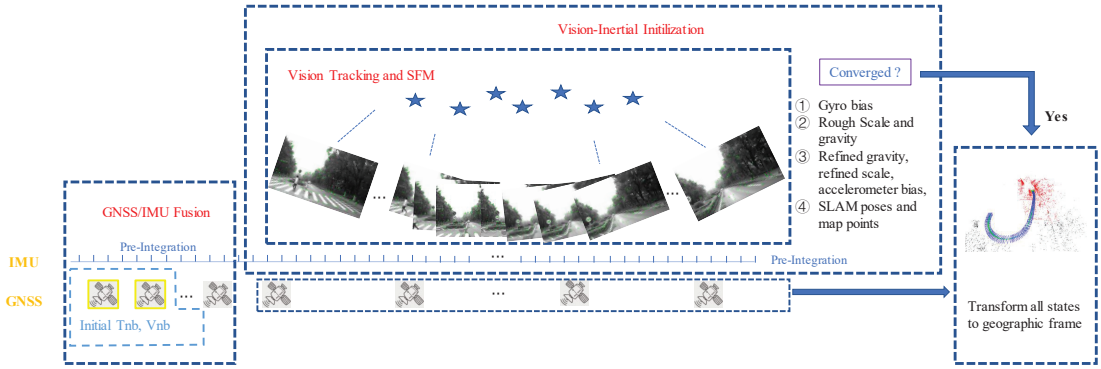


Figure 1. Algorithm.

This part starts with the definition of the notations. The inertial frame has its origin at the center of Earth and axes which are non-rotating with respect to the fixed stars, it is the measurement reference of IMU. We denote $(\cdot)_N$ as the navigation frame which has its origin coinciding with that of the sensor frame, and x-axis pointing towards geodetic north, z-axis parallel to the gravity direction, and y-axis completing a right-handed orthogonal frame, i.e. the north-east-down (NED) coordinate system. Therefore, the gravity vector can be written as $\boldsymbol{g} = \begin{bmatrix} 0 & 0 & g \end{bmatrix}^T$. GNSS observations are commonly converted to NED coordinate frame, from latitude, longitude and altitude in geodetic coordinate system, by setting a datum point, e.g., the first measurement. Correspondingly, VIO system estimates its states under the VIO frame, also called World frame $(\cdot)_W$, which can be defined at arbitrary origin and directions. In general, the navigation states include pose, velocity, and the features positions are represented in navigation frame, thus we can define x_k as the state at time stamp k, instead of x_k^n. The IMU frame is an orthogonal axis set which is aligned to the vehicle body frame $(\cdot)_b$, hence we treat IMU frame as the body frame. We consider $(\cdot)_c$ as the camera frame, in which the normalized features can be expressed. The matrix $\boldsymbol{T}_{XY} = \begin{bmatrix} \boldsymbol{R}_{XY} & \boldsymbol{t}_{XY} \end{bmatrix}$ means the transformation from frame Y to frame X,

where $\boldsymbol{R}_{XY} \in SO(3)$ and $\boldsymbol{t}_{XY} \in \mathbb{R}^3$ are the rotation and translation respectively, also can be used to transform the point coordinate in frame Y to frame X, i.e., $\boldsymbol{p}_X = \boldsymbol{T}_{XY} \cdot \boldsymbol{p}_Y$. We regard $\boldsymbol{T}_{NB}$ as the body pose, $\boldsymbol{T}_{BC}$ as the transformation from camera frame C to body frame B, also known as the extrinsic calibration. We use $\widetilde{(\cdot)}$to denote sensor measurements, which may be affected by noise and bias. With the assumption that the intrinsic parameter of the camera and extrinsic parameter between the camera and IMU has been calibrated before the algorithm, we provide some preliminary knowledge about the sensor measurements, models.

2.1. System Preliminaries

2.1.1. Factor Graph Optimization Model

We utilize the factor graph model as the optimization framework [30,31], which is one of most widely used graph models for the maximum a posteriori(MAP) estimation. Considering the general case, the inference problem $\boldsymbol{x}^{MAP} = \arg\max p(\boldsymbol{x}|\boldsymbol{z})$ can be factorized as:

$$p(\boldsymbol{x}|\boldsymbol{z}) = \prod_{i=1}^{k} p(x_k|z_k) = \prod_{i=1}^{k} f_i(x_i) \tag{1}$$

where $\boldsymbol{z}$ and $\boldsymbol{x}$ are measurements and states respectively, and with gaussian assumption we have $f_i(x_i) \propto exp\left(-\frac{1}{2}\|h_i(x_i) - z_i\|^2\right)$, each factor fi represents an error function that should be minimized, it is also a measurement unit in a graphical model known as a factor graph. Computing the MAP estimation is equivalent to running inference over the factor graph [32]. To avoid the involved computational complexity, we use the incremental smoothing developed in [33].

2.1.2. GNSS Positional Measurements

The GNSS measurement at epoch k is given by the equation:

$$\boldsymbol{z}_k^G = \boldsymbol{h}(\boldsymbol{x}_k)^G + \boldsymbol{n}^G \tag{2}$$

where h^G is the function linking the vehicle's position and the GNSS measurement z_k^G, and n^G is the measurement noise. Given the lever-arm between GNSS antenna center and the IMU measurement center, h^G will contain the rotation[34]. Therefore, if the lever-arm is a zero vector, i.e., GNSS antenna and IMU share the same center, GNSS measurements have no contribution for rotation estimation.

2.1.3. Monocular Vision

We use a monocular pinhole camera model with a known calibration matrix $\boldsymbol{K}$ to predict the image observation, via the non-linear function π:

$$\boldsymbol{\pi}(\boldsymbol{x}, \boldsymbol{l}) = \boldsymbol{K}\begin{bmatrix}\boldsymbol{R}_{CN} & \boldsymbol{t}_{CN}\end{bmatrix}{}_N\boldsymbol{l} \tag{3}$$

where $\boldsymbol{K}$ is the camera intrinsic calibration. $\boldsymbol{x}$ is the inverse of camera pose, $\boldsymbol{l}$ represents the landmark in the navigation frame, they are states to be estimated. Both landmark and predicted image observation are expressed in homogeneous coordinates. Given the known image measurement $\boldsymbol{z}(u, v)$, the re-projection error is:

$$\boldsymbol{e}(\boldsymbol{x}, \boldsymbol{l}) = \boldsymbol{z} - \boldsymbol{\pi}(\boldsymbol{x}, \boldsymbol{l}) \tag{4}$$

By minimizing the re-projection error, we can recover the relative pose up to an unknown scale within multiple frames. With the known calibration between body and camera frame, we can transform the camera pose to the navigation state. Such problem is known as full SLAM or BA, requires adding the unknown landmarks into the optimization. Alternatively, one can apply the multi-view geometry, instead of the re-projection equation, to avoid the including the unknown landmarks.

142 2.1.4. IMU Kinematic Model and Pre-integration on Manifold

An IMU sensor usually contains a 3-axis gyroscope sensor and a 3-axis accelerometer, which can measure the angular velocity and the acceleration of the inertial sensor, i.e., the body frame, with respect to inertial frame. IMU provides outputs at a much higher rate than GNSS and visual measurements. IMU measurements are polluted by two error sources: a white noise $\boldsymbol{\eta}(t)$, and a random walk slow time-varying bias $\boldsymbol{b}(t)$. Therefore, the IMU measurement model can be constructed as:

$$ {}_B\widetilde{\boldsymbol{\omega}}_{NB}(t) = {}_B\boldsymbol{\omega}_N B(t) + \boldsymbol{b}_g(t) + \boldsymbol{\eta}_g(t) \tag{5} $$

$$ {}_B\widetilde{\boldsymbol{a}}(t) = \boldsymbol{R}_{NB}^T(t)({}_N\boldsymbol{a}(t) - {}_N\boldsymbol{g}) + \boldsymbol{b}_a(t) + \boldsymbol{\eta}_a(t) \tag{6} $$

where ${}_B\widetilde{\boldsymbol{\omega}}_{NB}(t)$ and ${}_B\widetilde{\boldsymbol{a}}(t)$ are respectively the measured angular velocity and linear acceleration in the body frame b, the real angular velocity $\boldsymbol{\omega}(t)$ and linear acceleration $\boldsymbol{a}(t)$ are what we need $\boldsymbol{R}_{NB}$ maps a point or vector from body frame to navigation frame N, ${}_N\boldsymbol{g}$ is gravity under N frame. We ignore effects due to earth's rotation, which means we can assume N is an inertial frame. The time-varying bias $\boldsymbol{b}_g(t)$ and $\boldsymbol{b}_a(t)$ are modeled as:

$$ \dot{\boldsymbol{b}}_g(t) = \boldsymbol{\eta}_{b_g}(t), \dot{\boldsymbol{b}}_a(t) = \boldsymbol{\eta}_{b_a}(t) \tag{7} $$

To infer the motion state, we utilize the kinematic model in [35]:

$$ \begin{cases} {}_N\dot{\boldsymbol{R}}_{NB} = \boldsymbol{R}_{NB} \cdot ({}_B\boldsymbol{\omega})\wedge \\ {}_N\dot{\boldsymbol{v}}^n = {}_N\boldsymbol{a} \\ {}_N\dot{\boldsymbol{p}} = {}_N\boldsymbol{v} \end{cases}, \quad where \quad \boldsymbol{\omega}^\wedge = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{8} $$

During the short time interval $[t, t+\Delta t]$, we can assume that ${}_N\boldsymbol{a}$ and ${}_B\boldsymbol{\omega}_{NB}$ stay constant. We can obtain the pose and velocity at time $t+\Delta t$ via integrating Eq. (8):

$$ \boldsymbol{R}(t+\Delta t) = \boldsymbol{R}(t) Exp\left((\widetilde{\boldsymbol{\omega}}(t) - \boldsymbol{b}_g(t) - \boldsymbol{\eta}_g(t))\Delta t\right) \tag{9} $$

$$ \boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t) + \boldsymbol{g}\Delta t + \boldsymbol{R}(t)\left(\widetilde{\boldsymbol{a}}(t) - \boldsymbol{b}_a(t) - \boldsymbol{\eta}_a(t)\right)\Delta t \tag{10} $$

$$ \boldsymbol{p}(t+\Delta t) = \boldsymbol{p}(t) + \boldsymbol{v}(t)\Delta t + \frac{1}{2}\boldsymbol{g}\Delta t^2 + \frac{1}{2}\boldsymbol{R}(t)\left(\widetilde{{}_N\boldsymbol{a}}(t) - \boldsymbol{b}_a(t) - \boldsymbol{\eta}_a(t)\right)\Delta t^2 \tag{11} $$

143 We have substituted IMU measurements into the equations, and we dropped the coordinate frame
144 subscripts for readability.

Assuming that IMU is synchronized with the camera and provides measurements at discrete times k, we can iterate the integration Eqs. (9-11). To avoid recompute the integration whenever the linearization point changes, we follow [35] and apply the following relative motion increments between two consecutive keyframes that are independent of the pose and velocity at t_i:

$$ \Delta\boldsymbol{R}_{ij} \doteq \boldsymbol{R}_i^T\boldsymbol{R}_j = \prod_{k=i}^{j-1} Exp\left((\widetilde{\boldsymbol{\omega}}_k - \boldsymbol{b}_{g_k} - \boldsymbol{\eta}_{g_k})\Delta t\right) \tag{12} $$

$$ \Delta\boldsymbol{v}_{ij} \doteq \boldsymbol{R}_i^T(\boldsymbol{v}_j - \boldsymbol{v}_i - \boldsymbol{g}\Delta t_{ij}) = \sum_{k=i}^{j-1} \Delta\boldsymbol{R}_{ik}\left(\widetilde{\boldsymbol{a}}_k - \boldsymbol{b}_{a_k} - \boldsymbol{\eta}_{a_k}\right)\Delta t \tag{13} $$

$$ \Delta\boldsymbol{p}_{ij} \doteq \boldsymbol{R}_i^T(\boldsymbol{p}_j - \boldsymbol{p}_i - \boldsymbol{v}_i\Delta t_{ij} - \frac{1}{2}\boldsymbol{g}\Delta t_{ij}^2) = \sum_{k=i}^{j-1}\left[\Delta\boldsymbol{v}_{ik}\Delta t + \frac{1}{2}\Delta\boldsymbol{R}_{ik}(\widetilde{\boldsymbol{a}}_k - \boldsymbol{b}_{a_k} - \boldsymbol{\eta}_{a_k})\Delta t^2\right] \tag{14} $$

As the noise covariance has a strong influence on the MAP estimator, it is of paramount importance to accurately model the noise covariance. Defining the noise vector as $\boldsymbol{\eta}_{ik}^{\Delta} \doteq \begin{bmatrix} \delta\boldsymbol{\phi}_{ij}^T & \delta\boldsymbol{v}_{ij}^T & \delta\boldsymbol{p}_{ij}^T \end{bmatrix}^T$, the covariance propagation can be written in iterative form:

$$\delta\boldsymbol{\phi}_{ij} = \Delta\widetilde{\boldsymbol{R}}_{j-1,j}\delta\boldsymbol{\phi}_{i,j-1} + \boldsymbol{J}_r^{j-1}\boldsymbol{\eta}_{j-1}^{gd}\Delta t \tag{15}$$

$$\delta\boldsymbol{v}_{ij} = \delta\boldsymbol{v}_{i,j-1} - \Delta\widetilde{\boldsymbol{R}}_{j-1,j}(\widetilde{\boldsymbol{a}}_{j-1} - \boldsymbol{b}_i^a)^{\wedge}\delta\boldsymbol{\phi}_{i,j-1}\Delta t + \Delta\widetilde{\boldsymbol{R}}_{j-1,j}\boldsymbol{\eta}_{j-1}^{ad}\Delta t \tag{16}$$

$$\delta\boldsymbol{p}_{ij} = \delta\boldsymbol{p}_{i,j-1} + \delta\boldsymbol{v}_{i,j-1}\Delta t - \frac{1}{2}\Delta\widetilde{\boldsymbol{R}}_{j-1,j}(\widetilde{\boldsymbol{a}}_{j-1} - \boldsymbol{b}_i^a)^{\wedge}\delta\boldsymbol{\phi}_{i,j-1}\Delta t^2 + \frac{1}{2}\Delta\widetilde{\boldsymbol{R}}_{j-1,j}\boldsymbol{\eta}_{j-1}^{ad}\Delta t^2 \tag{17}$$

where $\boldsymbol{J}_r$ is the right Jacobian of $SO(3)$, $\boldsymbol{\eta}_k^d$ is the discrete-time noise of IMU, whose covariance is a function of the sampling rate. Defining the noise as $\boldsymbol{\eta}_k^d = \begin{bmatrix} \boldsymbol{\eta}_k^{gd} & \boldsymbol{\eta}_k^{ad} \end{bmatrix}$, the corresponding covariance as $\boldsymbol{P}_k$, we can write the noise equations (15-17) in compact matrix form, and the preintegrated measurement covariance iteratively:

$$\boldsymbol{\eta}_{ij}^{\Delta} = \boldsymbol{A}_{j-1}\boldsymbol{\eta}_{i,j-1}^{\Delta} + \boldsymbol{B}_{j-1}\boldsymbol{\eta}_{j-1}^{d} \tag{18}$$

$$\boldsymbol{P}_{ij} = \boldsymbol{A}_{j-1}\boldsymbol{P}_{i,j-1}\boldsymbol{A}_{j-1}^T + \boldsymbol{B}_{j-1}\boldsymbol{Q}_{\eta}\boldsymbol{B}_{j-1}^T \tag{19}$$

We consider that the biases during the pre-integration period between the interval Δt is constant. However, the biases change slightly by a small amount $\delta\boldsymbol{b}$ in the optimization window To avoid the time -expensive recomputation when the bias changes, we can update the states using a first-order expansion:

$$\boldsymbol{R}_{NB}^{i+1} = \boldsymbol{R}_{NB}^{i}\Delta\boldsymbol{R}_{i,i+1}Exp\left(\boldsymbol{J}_{\Delta\boldsymbol{R}}^{\boldsymbol{b}_g}\delta\boldsymbol{b}_g^i\right) \tag{20}$$

$${}_N\boldsymbol{v}_B^{i+1} = {}_N\boldsymbol{v}_B^{i} + {}_N\boldsymbol{g}\Delta t_{i,i+1} + \boldsymbol{R}_{NB}^{i}(\Delta\boldsymbol{v}_{i,i+1} + \boldsymbol{J}_{\Delta\boldsymbol{v}}^{\boldsymbol{b}_g}\delta\boldsymbol{b}_g^i + \boldsymbol{J}_{\Delta\boldsymbol{v}}^{\boldsymbol{b}_a}\delta\boldsymbol{b}_a^i) \tag{21}$$

$${}_N\boldsymbol{p}_B^{i+1} = {}_N\boldsymbol{p}_B^{i} + {}_N\boldsymbol{v}_B^{i}\Delta t_{i,i+1} + \frac{1}{2}{}_N\boldsymbol{g}\Delta t_{i,i+1}^2 + \boldsymbol{R}_{NB}^{i}(\Delta\boldsymbol{p}_{i,i+1} + \boldsymbol{J}_{\Delta\boldsymbol{p}}^{\boldsymbol{b}_g}\delta\boldsymbol{b}_g^i + \boldsymbol{J}_{\Delta\boldsymbol{p}}^{\boldsymbol{b}_a}\delta\boldsymbol{b}_a^i) \tag{22}$$

Here, the Jacobians can be precomputed during the pre-integration.

2.2. Vision/INS/GNSS Initialization

We use monocular vision, low-cost INS and GNSS for navigation, the methods in [27] and [10] offer the use for reference to our algorithm, and both are state-of-the-art VIO systems. In general, the initialization for VIO estimate the gyroscope bias as the first step, and then formulate system equations including states such as vision scale, accelerometer biases, gravity and velocity. Since the system may easily get ill-conditioned due to the mixture of gravity and accelerometer bias. Such problem will result incorrect solutions, which may break the system. In this paper, using the GNSS observations will decouple vision scale, accelerometer bias and gravity with respective computation.

2.2.1. Initial State by GNSS

To estimate the global state in N frame, we can directly utilize the initial position and velocity by GNSS-only measurements. Since GNSS sensor is 3-DOF without rotation information, for the in-motion vehicle, we can use GNSS velocity to obtain the approximate attitude angle to form the rotation matrix. However, we can not achieve roll via GNSS velocity, which is always roughly set zero, due to the mounting structure of INS relative to the body. Pitch and yaw angle can be computed as following:

$$\theta = tan^{-1}\left(\frac{v_u}{\sqrt{v_e^2 + v_n^2}}\right) \tag{23}$$

$$\phi = tan^{-1}\left(\frac{v_n}{v_e}\right) \tag{24}$$

where θ and ϕ is pitch and yaw angle respectively, whose accuracy depends on the accuracy of GNSS velocity. Note that the calculation rule depends on the definition of the Euler Angles. Another initial attitude determination method is introduced in [36] by accelerometer levelling, which uses knowledge of gravity sensed by each accelerometer, under static conditions. If the GNSS algorithms fail to provide velocities due the lack of Doppler measurements, one can use the position difference in N frame as following:

$$v_N = \frac{NED_i - NED_{i-1}}{t_i - t_{i-1}} \tag{25}$$

2.2.2. Short term GNSS/INS Fusion

The estimation and refinement of the gravity orientation is the essential step for current VIO system, due to the unknown initial attitude of the body. The refined gravity orientation is used to determine the attitude [27]. As a result, the mixing bias error may lead to an incorrect gravity direction, which will result in attitude error. For most vehicle applications, a heavy dynamic of the body may only occur along z axis, resulting a large change just on yaw angle, which means the z may always be approximately parallel to gravity vector. In this paper, given known axes of the IMU frame, one can correct the gravity directly with the optimized initial state.

2.3. Vision Structure with Initial States

The initialization is based on loosely-coupled system, vision structure is needed to estimate a graph of camera poses and feature positions. Our method borrows idea from [27], which maintains several spacial-separate frames selected by enough parallax near the neighbor. We set rotation via Five-point method, to the chosen two frames which contain sufficient feature parallax. Differently from [27], the translation between such two frames is computed by IMU pre-integration retraction [31] based on GNSS/INS optimization, by the following transformation:

$$\boldsymbol{R}_{WB} = \boldsymbol{R}_{WC}\boldsymbol{R}_{CB} \tag{26}$$

$$_W\boldsymbol{p}_B = s_W\boldsymbol{p}_C + \boldsymbol{R}_{WC} \cdot {}_C\boldsymbol{p}_B \tag{27}$$

Such setting allows us to assign the structure a rough scale from GNSS/INS integration, whose metric measurements has a high accuracy. With the triangulation, Perspective-n-Point (PnP) estimation, and a global full bundle adjustment successively, we get all frame poses and feature positions under N frame.

2.3.1. Gyroscope Bias Estimation

With the relationship of (26), the relationship between two consecutive frames relative rotation and pre-integration rotation can be described as:

$$\Delta\boldsymbol{R}_{i,i+1}Exp(\boldsymbol{J}^g_{\Delta\boldsymbol{R}}\boldsymbol{b}_g) = \boldsymbol{R}_{BC}\boldsymbol{R}^i_{CW}\boldsymbol{R}^{i+1}_{WC}\boldsymbol{R}_{CB} \tag{28}$$

where $\boldsymbol{R}_{BC}$ is the extrinsic rotation calibrated in the previous iteration.

$$\arg\min_{\boldsymbol{b}_g} = \sum_{i=1}^{n-1} \| Log\left((\Delta\boldsymbol{R}_{i,i+1}Exp(\boldsymbol{J}^g_{\Delta\boldsymbol{R}}\boldsymbol{b}_g))^T\boldsymbol{R}^i_{BW}\boldsymbol{R}^{i+1}_{WB}\right)\|^2 \tag{29}$$

where n is the number of the keyframes, $\boldsymbol{R}^{(\cdot)}_{WB} = \boldsymbol{R}^{(\cdot)}_{WC}\boldsymbol{R}_{CB}$ is the body rotation by the operation of the computed vision structure rotation $\boldsymbol{R}^{(\cdot)}_{WC}$ and the calibration $\boldsymbol{R}_{CB}$. Solving the equation (29) by

least-square with a zero bias seed, we can update the pre-integration $\Delta\boldsymbol{R}_{ij}$, $\Delta\boldsymbol{v}_{ij}$, $\Delta\boldsymbol{p}_{ij}$ with respect to the estimated $\boldsymbol{b}_g$.

2.4. Accelerometer Bias, Gravity and Scale Estimation

Although assigned an initial scale based on short-term GNSS/INS fusion, the accuracy of ${}_N\boldsymbol{p}_C$ is polluted by the IMU biases which are not converged, and other random noises. With the estimated gyroscope bias and initial states, we can substitute (27) to (22), and use the estimated gyroscope bias in (29) to modify $\Delta p_{i,i+1}$, it follows:

$$\begin{aligned} s\,{}_W\boldsymbol{p}_C^{i+1} =& s\,{}_W\boldsymbol{p}_C^{i} + {}_W\boldsymbol{v}_B^{i}\Delta t_{i,i+1} + \frac{1}{2}{}_W\boldsymbol{g}\Delta t_{i,i+1}^2 + (\boldsymbol{R}_{WC}^{i} - \boldsymbol{R}_{WC}^{i+1})\,{}_C\boldsymbol{p}_B + \\ & \boldsymbol{R}_{WB}^{i}(\Delta p_{i,i+1} + \boldsymbol{J}_{\Delta\boldsymbol{p}}^{b_a}\delta\boldsymbol{b}_a^{i}) \end{aligned} \tag{30}$$

Our goal is to compute the accelerometer bias, extrinsic translation and the scale by solving a linear system of equations. To avoid noises of N velocities, and reduce complexity, we consider two relations (30) between three consecutive keyframes and utilize velocity relation in (21), resulting in the following expression:

$$\begin{bmatrix} \boldsymbol{A}_1(i) & \boldsymbol{A}_2(i) & \boldsymbol{A}_3(i) \end{bmatrix} \cdot \begin{bmatrix} s \\ \boldsymbol{b}_a \\ {}_C\boldsymbol{p}_B \end{bmatrix} = \boldsymbol{B}(i) \tag{31}$$

Writing keyframes i, $i+1$, $i+2$ as 1,2,3 for clarity of notation, we have:

$$\begin{aligned} \boldsymbol{A}_1(i) =& ({}_W\boldsymbol{p}_C^2 - {}_W\boldsymbol{p}_C^3)\Delta t_{12} - ({}_W\boldsymbol{p}_C^1 - {}_W\boldsymbol{p}_C^2)\Delta t_{23} \\ \boldsymbol{A}_2(i) =& \boldsymbol{R}_{WB}^2\boldsymbol{J}_{\Delta\boldsymbol{p}_{23}}^{b_a}\Delta t_{12} - \boldsymbol{R}_{WB}^1\boldsymbol{J}_{\Delta\boldsymbol{p}_{12}}^{b_a}\Delta t_{23} + \boldsymbol{R}_{WB}^1\boldsymbol{J}_{\Delta\boldsymbol{v}_{12}}^{b_a}\Delta t_{12}\Delta t_{23} \\ \boldsymbol{A}_3(i) =& -(\boldsymbol{R}_{WC}^1 - \boldsymbol{R}_{WC}^2)\boldsymbol{t}_{CB}\Delta t_{23} + (\boldsymbol{R}_{WC}^2 - \boldsymbol{R}_{WC}^3)\boldsymbol{t}_{CB}\Delta t_{12} \\ \boldsymbol{B}(i) =& \boldsymbol{R}_{WB}^1\Delta\boldsymbol{p}_{12}\Delta t_{23} - \boldsymbol{R}_{WB}^1\Delta\boldsymbol{v}_{12}\Delta t_{12}\Delta t_{23} - \boldsymbol{R}_{WB}^2\Delta\boldsymbol{p}_{23}\Delta t_{12} \\ & - \frac{1}{2}{}_W\boldsymbol{g}\Delta t_{12}\Delta t_{23}(\Delta t_{12} + \Delta t_{23}) \\ & + ({}_W\boldsymbol{p}_C^1 - {}_W\boldsymbol{p}_C^2)\Delta t_{23} - ({}_W\boldsymbol{p}_C^2 - {}_W\boldsymbol{p}_C^3)\Delta t_{12} \end{aligned} \tag{32}$$

Stacking all relations between three consecutive keyframes (31) we form an overdetermined linear system of equations $\boldsymbol{A}_{3(n-2)\times 3}\boldsymbol{b}_{3\times 1} = \boldsymbol{B}_{3(n-2)\times 1}$ which can be solved via least-square. We can calculate the condition number to check whether the problem is well-conditioned.

Criterions are necessary for the termination when all the parameters are converged. Since gyroscope bias and scale usually converge fast to stable values, more attention should be paid to gravity, accelerometer bias. We set an empirical threshold 0.02 m/s^2 for the std(standard deviation) of gravity and accelerometer bias estimated by the last 10 consecutive keyframes, to determine their convergence. Another widely-used criterion is condition number, which can suggest whether the equation is well conditioned. However, no quantitative value of such indicator is set, since an absolute threshold is not reliable enough[10].

2.5. GNSS-VIO Initialization

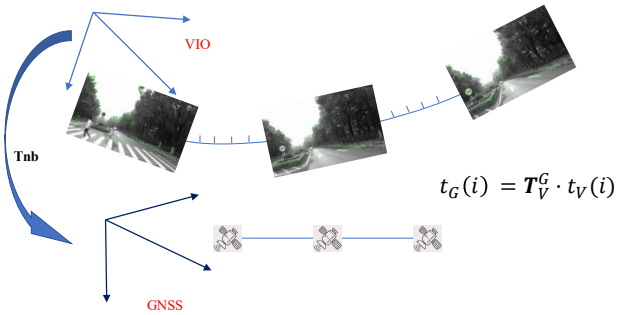


Figure 2. Algorithm.

As described earlier, to fuse GNSS measurements p_N and VIO position estimation p_W, the transformation or alignment parameters to relate them:

$$p_N = T_{NW} \oplus p_W, T \in Sim(3) \tag{33}$$

where T_{NW} consists of $\{s, R_{NW}, t_{NW}\}$. Horn alignment [37] is a widely used method to calculate T_{NW}, with a set of accurate ground truth. However, it is difficult to obtain true trajectories in some sheltered or GNSS-denied scenarios, and the position results are estimated with different accuracy. To apply the GNSS noise model, we perform a non-linear optimization to compute T_{NW}, with n point pairs:

$$\underset{s,R_{NW},t_{NW}}{\arg\min} = \sum_{i=1}^{n} \|z - f(s, R_{NW}, t_{NW})\|^2_{z_{noise}} \tag{34}$$

We can transform the estimated VIO positions to global frame with T_{NW}, and evaluate scale error as well.

3. Experiments

3.1. Implementation details

Our system is equipped with a basler aca1300-60gm monocular grayscale camera captures 1280×1024 images at 20Hz, and a GNSS/INS integrated system M40, which consists of a NovAtel multi-constellation GNSS card works at 1Hz and a ADIS-16460 IMU runs at 200Hz. All sensors are rigidly mounted on a plate fixed on the roof of a sport utility vehicle(SUV), we regard IMU roughly aligned with the vehicle frame.

Figure 3. Test Platform.

In our system, all the time systems are aligned via hard-ware trigger, generated by the GNSS card, is fed into camera and IMU. The extrinsic rotation and translation between camera and IMU are calibrated by the state-of-the-art software Kalibr[38]. The position of GNSS antenna center under IMU frame, also known as lever arm, is often measured with a metric ruler.

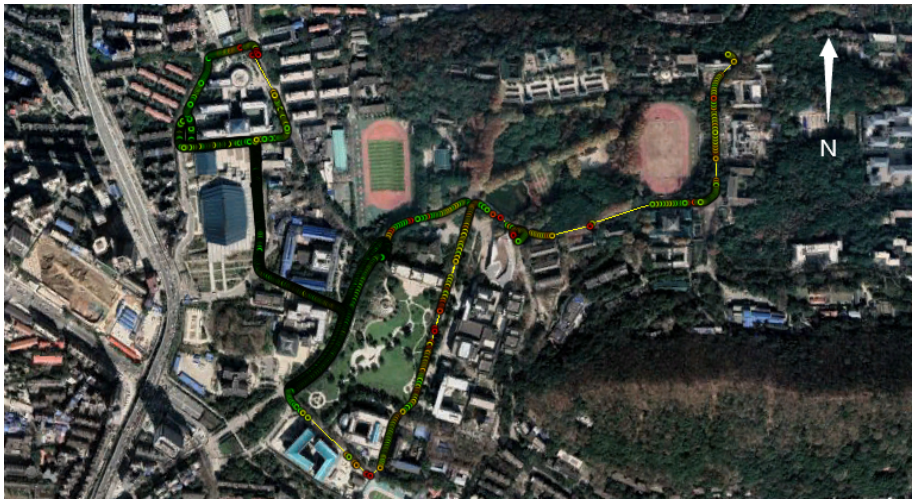


Figure 4. GNSS RTK Trajectories of the Tests.

Six tests were implemented in the campus of Wuhan University, whose RTK trajectories are illustrated in Fig. 4. As built under a hill, the roads of the campus are rarely plain or straight, hence the car could perform lots of dynamic motion. And some parts of the area are full of trees, where GNSS signal may be denied, while the objects and features can be easily detected. In order to avoid long-term GNSS signal blocking, a part of open-sky environment in the campus is selected to conduct all tests. Thus, some segments of the trajectories are overlapped. To perform a comprehensive evaluation for the initialization, different types of scenarios are selected, a short summary of which is given in Table 1:

Table 1. Initialization scenarios.

t1	t2	t3	t4	t5	t6
lots of trees dim scence turning corner 8.1~11.1km/h	a few tress bright scence turning corner 6.1~8.9km/h	a few buildings dim scence turning corner 5.5~6.2km/h	open-sky bright scence turning corner 9.0~15.8km/h	open-sky bright scence straight line 14.1~16.3km/h	a few tress bright scence straight line 15.0~19.9km/h

We focus on the initial navigation states by GNSS, and the estimated VIO parameters including gyroscope bias, accelerometer bias, gravity, visual scale, and the optimized GNSS-VIO transformation. To evaluate the initial navigation states, the GNSS/INS integrated navigation system is carried out to be taken as the reference values. The VIO parameters estimation will be compared with the state-of-the-art VIO algorithm VI ORB-SLAM2[10]. At last, an optimization of the transformation between GNSS and VIO trajectories is performed, to align the two sets of navigation states, and to assess the scale error.

3.2. Initial Altitude Estimation by GNSS

GNSS velocities to estimate pitch and yaw, are provided by RTK algorithm. Table 2 shows the error statistics of all the six test datasets. We can see that the error of pitch and yaw is about 1~2 deg. Yaw angle has a smaller error than pitch angle, this is because only pitch variable involves v_d, whose relative noise is often larger than v_e and v_n. The GNSS-denied environments (test1, test3, test4, note this is about the whole trajectories, thus it can be different from the initialization scenarios in Table 1) reduced the accuracy of GNSS velocities, and thus of the pitch and yaw.

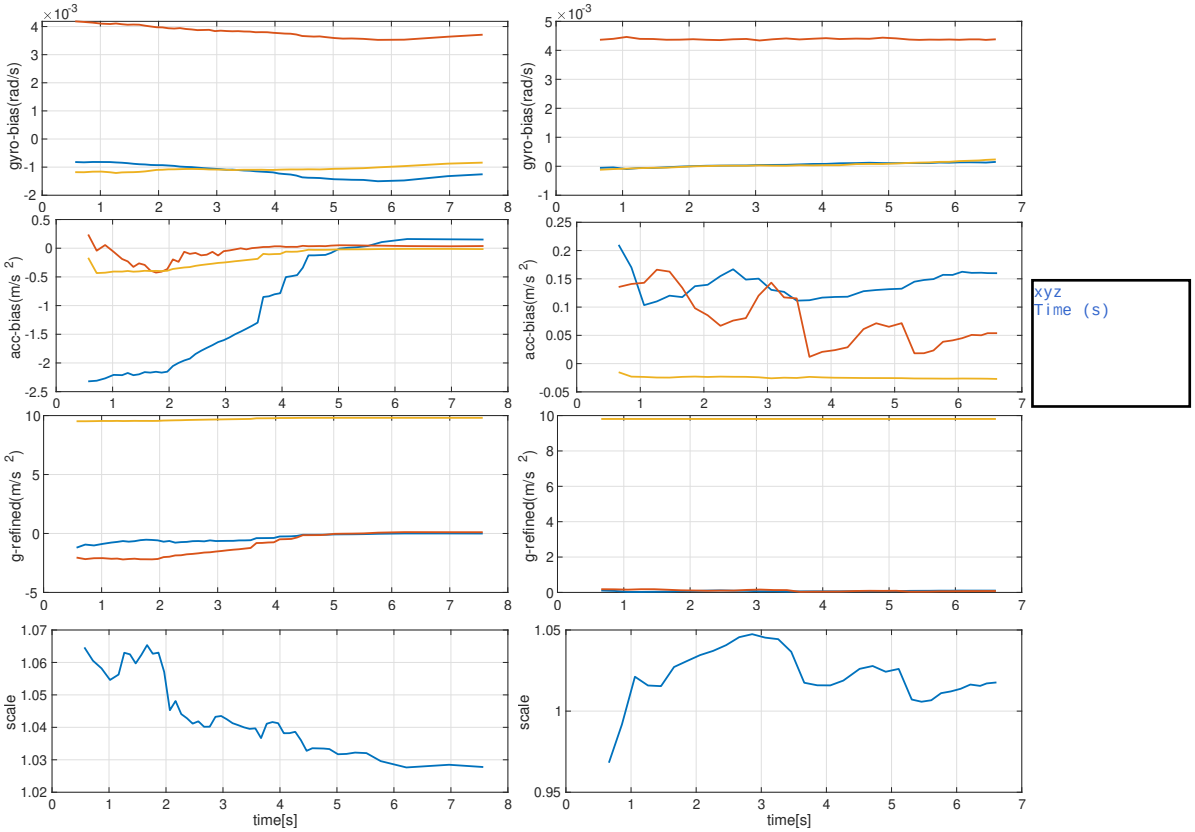
Table 2. statistics of pitch, yaw error(deg).

		t1	t2	t3	t4	t5	t6	mean
Pitch	rms	2.127	1.663	2.185	2.048	1.773	1.897	1.949
	std	1.896	1.310	1.720	1.863	1.671	1.740	1.700
Yaw	rms	1.667	1.626	1.566	1.329	1.753	1.633	1.596
	std	1.397	1.101	0.957	0.936	1.351	1.312	1.176

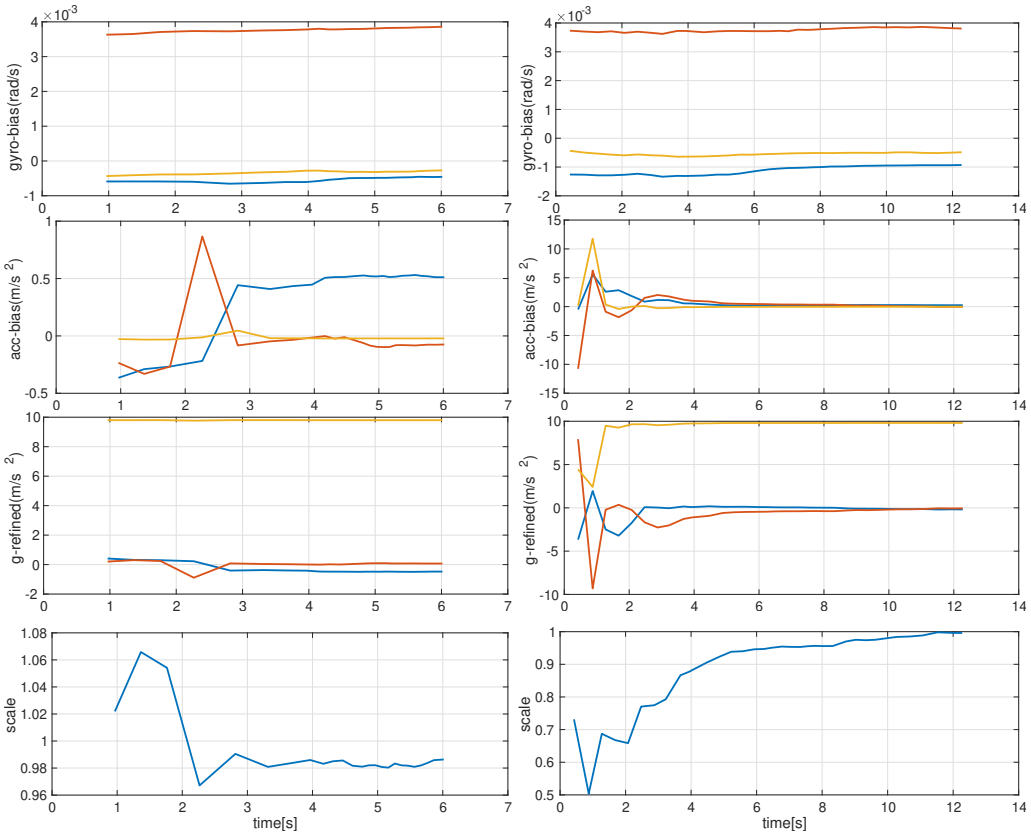
3.3. The Performance of VIO Parameters Estimation

In order to analyze the performance of the estimation of IMU bias, gravity and camera scale, test1, test2, test4 and test6 are expected to evaluate the proposed algorithm. The time varied characteristic curves of the four tests' VIO parameters by our method are shown in Fig. 5. It can be seen that all the variables by our estimation start to converge between 6 and 8 seconds on test1, test2 and test4, while it takes a longer time on test6. This is because the parallax between consecutive frames changes slowly on a straight line, thus enough parallax for keyframe selection needs larger time duration than other motion patterns, despite the highest speed.

The curves of gyroscope bias can achieve stable values in a very short time, while accelerometer bias and gravity suffer oscillation in the beginning, for the difficulty of distinguishing them. Comparing with test1, the variables especially the accelerometer bias and gravity change dramatically in about 2 second on test1 and test4. It might be because the bright scene can construct a better quality of SLAM model.

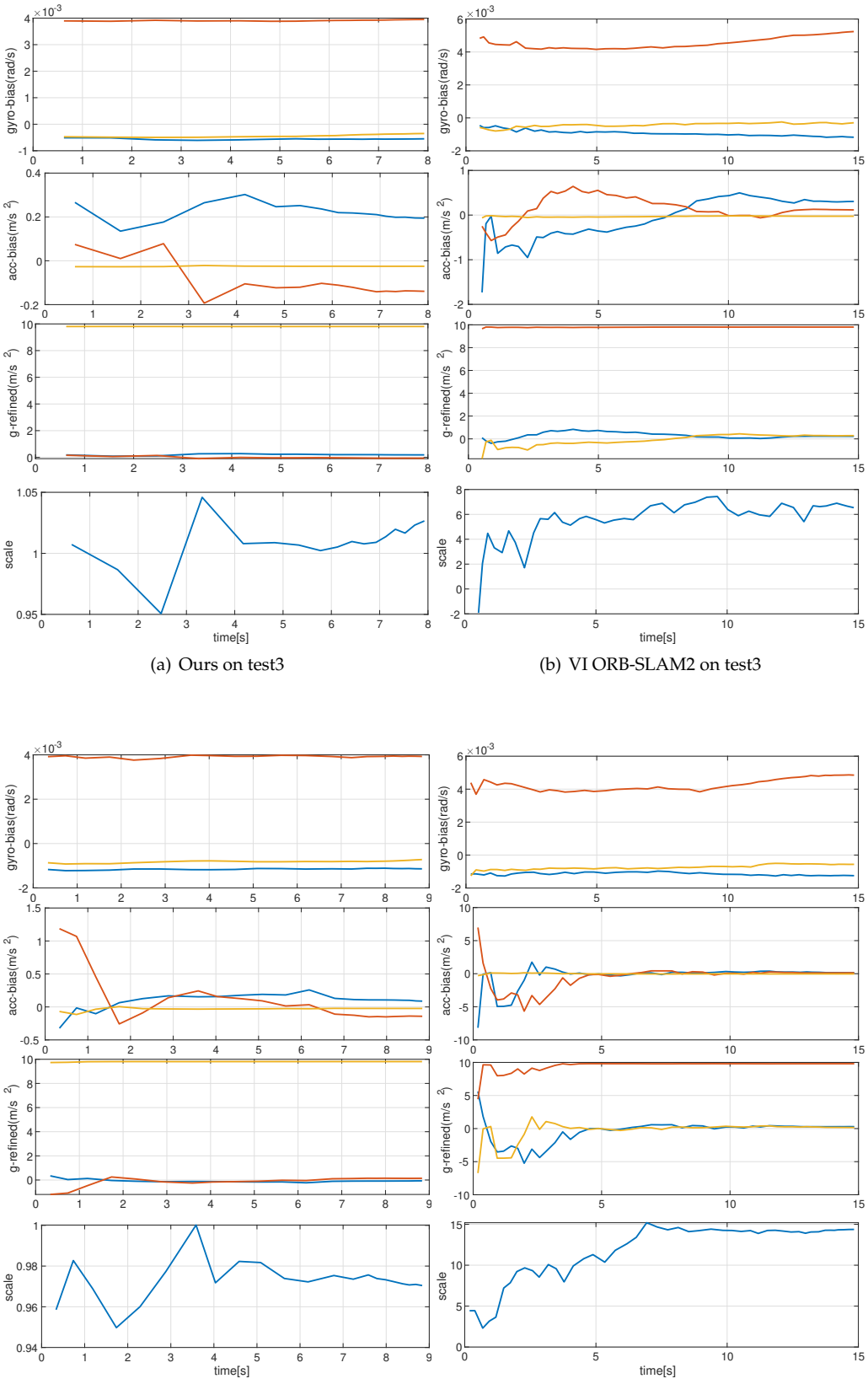


(a) test1 (b) test2



(c) test4 (d) test6

Figure 5. VIO parameters estimation results.



(a) Ours on test3 (b) VI ORB-SLAM2 on test3

(c) Ours on test5 (d) VI ORB-SLAM2 on test5

Figure 6. The comparison of Our method with VI ORB-SLAM2 on test3 and test5.

To analyze the consistency of the proposed initialization method, our method is compared with the state-of-the-art VI ORB-SLAM2. Fig. 6 shows comparison of VIO parameters results of the proposed

method and VI ORB-SLAM2 on tets3 and test5. It can be seen that our method beats VI ORB-SLAM2 on time cost of convergence, which shows that a coarse initial states especially the attitude and scale fed to VIO initialization, can achieve good performance. The gyroscope bias of our method are very close to VI ORB-SLAM2, while such consistency does not apply to accelerometer bias and gravity. This phenomenon confirms that the coupling of such two variables. Since the motion of a car is almost on a plane, at least two axes of the IMU may not be excited enough. As a result, the accelerometer bias and gravity estimation of both methods could be inaccurate.

3.4. GNSS-VIO Alignment

Applying the scale to keyframe translations and feature positions in Visual-Inertial estimation, we can calculate the keyframe poses and points very close to the real world. To measure the scale error, we optimize the transformation parameters between the initialization and GNSS/INS fusion, which is taken as the ground truth. To compare, we run the same exhaustive tests with VI ORB-SLAM2 and another state-of-the-art VINS-Mono[8]. As the initialization of VINS-Mono use only a limited count of keyframes, we set the window size 20 to balance the keyframe quantity and the co-visual field of the whole window. Hence, some initialization tests may fail to get the correct values for the insufficient number of keyframes.

The results are summarized in Table 3. It can be seen that scale error of all tests by the proposed initialization is within 10%, while VI ORB-SLAM2 and VINS-Mono has some errors larger than 10% or higher. The proposed method achieves the best scale estimation on most tests. Our method outperforms VI ORB-SLAM2 and VINS-Mono in the accuracy of alignment error on each test. Most alignment error of our method are within a decimeter level, while that of VI ORB-SLAM2 and VINS-Mono can be several meters.

Table 3. The Comparison of Our method with VI ORB-SLAM2 and VINS-Mono on scale error and alignment error. The estimation of VINS-Mono on test2 failed to get a correct scale value.

		t1	t2	t3	t4	t5	t6
Ours	scale error	8.24%	**0.08%**	**5.19%**	6.18%	**1.32%**	**0.18%**
	alignment error(m)	0.05	0.06	0.04	0.09	0.03	0.26
VI ORB-SLAM2	scale error	**1.13%**	6.91%	46%	**1.50%**	5.16%	2.8%
	alignment error(m)	1.40	0.52	2.27	0.51	0.97	0.76
VINS	scale error	85.9%	-	31%	20.4%	10.03%	7.4%
	alignment error(m)	3.21	-	1.06	2.35	1.06	1.37

4. Conclusions

In this paper, a fast and accurate initialization algorithm by the integration of Mono Vision, INS and GNSS is proposed. We focus on vehicle scene, whose motion pattern is different from the wide-used Euroc dataset[39] recorded by Micro Aerial Vehicle(MAV). Firstly, the GNSS RTK measurements are used to estimate the roughly initial navigation state, especially the attitude, which can provide a coarse direction of the gravity. However, most state-of-the-art algorithms set an arbitrary initial pose for the lacking of observations to calculate the direction. Secondly, with a roughly determined gravity, we can estimate the initial scale, which can be an initial value for the joint refinement for gravity, accelerometer bias, and scale. At the same time, GNSS/INS fusion is performed to provide trajectories of ground truth. Once the Visual-Inertial parameters converged, the transformation between the estimated keyframe poses and the ground truth is optimized to align two trajectories, and calculate the scale error as well. We verify the effectiveness on six tests with various scenarios. The results show that good brightness and non-straight line motion could improve the performance of initialization. The comparison shows that the proposed methods outperforms the VI ORB-SLAM2 and VINS-Mono on both gravity, accelerometer bias convergence and scale estimation.

The proposed method is still a kind of joint initialization of vision and IMU, which contains some steps based on non-linear optimization, which may be sensitive to the accuracy of the initial values. As a future work, we will try to split the system in an up-to-scale visual problem, followed by and inertial-only optimization phase.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing–original draft preparation, X.X.; writing–review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER grant number XXX." and and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

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

References

1. Hong, E.; Lim, J. Visual-Inertial Odometry with Robust Initialization and Online Scale Estimation. *Sensors* **2018**, *18*, 4287. doi:10.3390/s18124287.
2. Liu, J.; Gao, K.; Guo, W.; Cui, J.; Guo, C. Role, Path, and Vision of "5G + BDS/GNSS". *Satellite Navigation* **2020**, *1*, 23. doi:10.1186/s43020-020-00024-w.
3. Huang, W.; Liu, H.; Wan, W. Online Initialization and Extrinsic Spatial-Temporal Calibration for Monocular Visual-Inertial Odometry. *arXiv:2004.05534 [cs]* **2020**, [arXiv:cs/2004.05534].
4. Chen, X.; Hu, W.; Zhang, L.; Shi, Z.; Li, M. Integration of Low-Cost GNSS and Monocular Cameras for Simultaneous Localization and Mapping. *Sensors (Switzerland)* **2018**, *18*, 1–18. doi:10.3390/s18072193.
5. Wang, J.; Li, L.; Yu, H.; Gui, X.; Li, Z. VIMO: A Visual-Inertial-Magnetic Navigation System Based on Non-Linear Optimization. *Sensors* **2020**, *20*, 4386. doi:10.3390/s20164386.
6. Zhong, M.; Guo, J.; Zhou, D. Adaptive In-Flight Alignment of INS/GPS Systems for Aerial Mapping. *IEEE Transactions on Aerospace and Electronic Systems* **2018**, *54*, 1184–1196. doi:10.1109/TAES.2017.2776058.
7. Xufu Mu.; Jing Chen.; Zixiang Zhou.; Zhen Leng.; Lei Fan. Accurate Initial State Estimation in a Monocular Visual–Inertial SLAM System. *Sensors* **2018**, *18*, 506. doi:10.3390/s18020506.
8. Qin, T.; Li, P.; Shen, S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator **2017**. pp. 1–17. doi:10.1002/rob.21732.
9. Leutenegger, S.; Furgale, P.; Rabaud, V.; Chli, M.; Konolige, K.; Siegwart, R. Keyframe-Based Visual-Inertial SLAM Using Nonlinear Optimization. *Robotics: Science and Systems* **2013**. doi:10.15607/RSS.2013.IX.037.
10. Mur-Artal, R.; Tardos, J.D. Visual-Inertial Monocular SLAM With Map Reuse. *IEEE Robotics and Automation Letters* **2017**, *2*, 796–803. doi:10.1109/LRA.2017.2653359.
11. Cioffi, G.; Scaramuzza, D. Tightly-Coupled Fusion of Global Positional Measurements in Optimization-Based Visual-Inertial Odometry. *arXiv:2003.04159 [cs]* **2020**, [arXiv:cs/2003.04159].
12. Shunsuke, K.; Yanlei, G.; Hsu, L.T. GNSS/INS/On-Board Camera Integration for Vehicle Self-Localization in Urban Canyon. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* **2015**, *2015-Octob*, 2533–2538. doi:10.1109/ITSC.2015.407.
13. Sahmoudi, M.; Ramuni, N. Analysis of a Navigation System Based on Partially Tight Integration of IMU -Visual Odometry with Loosely Coupled GPS. ION, 2016, pp. 1322–1329.
14. Kim, J.; Cheng, J.; Guivant, J. Tightly-Coupled Integration of GPS/INS and Simultaneous Localisation and Mapping. *Australasian Conference on Robotics and Automation, ACRA* **2016**, *2016-Decem*, 194–199.
15. Adeel, M.; Gong, Z.; Liu, P.; Wang, Y.; Chen, X. Research and Performance Analysis of Tightly Coupled Vision, INS and GNSS System for Land Vehicle Applications. 30th International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS 2017, 2017, Vol. 5, pp. 3321–3330.
16. Lee, W.; Eckenhoff, K.; Geneva, P.; Huang, G. Intermittent GPS-Aided VIO: Online Initialization and Calibration. p. 9.

17. Mascaro, R.; Teixeira, L.; Hinzmann, T.; Siegwart, R.; Chli, M. GOMSF: Graph-Optimization Based Multi-Sensor Fusion for Robust UAV Pose Estimation. 2018 IEEE International Conference on Robotics and Automation (ICRA); IEEE: Brisbane, QLD, 2018; pp. 1421–1428. doi:10.1109/ICRA.2018.8460193.
18. Yu, Y.; Gao, W.; Liu, C.; Shen, S.; Liu, M. A GPS-Aided Omnidirectional Visual-Inertial State Estimator in Ubiquitous Environments. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); IEEE: Macau, China, 2019; pp. 7750–7755. doi:10.1109/IROS40897.2019.8968519.
19. Harsányi, K.; Kiss, A.; Szirányi, T.; Majdik, A. MASAT: A Fast and Robust Algorithm for Pose-Graph Initialization. *Pattern Recognition Letters* **2019**, p. S0167865519303241. doi:10.1016/j.patrec.2019.11.010.
20. Shin, E.H. Accuracy Improvement of Low Cost INS/GPS for Land Applications. PhD thesis.
21. Huang, Y.; Zhang, Y.; Chang, L. A New Fast In-Motion Coarse Alignment Method for GPS-Aided Low-Cost SINS. *IEEE/ASME Transactions on Mechatronics* **2018**, *23*, 1303–1313. doi:10.1109/TMECH.2018.2835486.
22. Martinelli, A. Closed-Form Solution of Visual-Inertial Structure from Motion. *International Journal of Computer Vision* **2014**, *106*, 138–152. doi:10.1007/s11263-013-0647-7.
23. Kneip, L.; Weiss, S.; Siegwart, R. Deterministic Initialization of Metric State Estimation Filters for Loosely-Coupled Monocular Vision-Inertial Systems. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems; IEEE: San Francisco, CA, 2011; pp. 2235–2241. doi:10.1109/IROS.2011.6094699.
24. Yang, Z.; Shen, S. Monocular Visual–Inertial State Estimation With Online Initialization and Camera–IMU Extrinsic Calibration. *IEEE Transactions on Automation Science and Engineering* **2017**, *14*, 39–51. doi:10.1109/TASE.2016.2550621.
25. Faessler, M.; Fontana, F.; Forster, C.; Scaramuzza, D. Automatic Re-Initialization and Failure Recovery for Aggressive Flight with a Monocular Vision-Based Quadrotor. 2015 IEEE International Conference on Robotics and Automation (ICRA); IEEE: Seattle, WA, USA, 2015; pp. 1722–1729. doi:10.1109/ICRA.2015.7139420.
26. Weiss, S.; Brockers, R.; Albrektsen, S.; Matthies, L. Inertial Optical Flow for Throw-and-Go Micro Air Vehicles. 2015 IEEE Winter Conference on Applications of Computer Vision; IEEE: Waikoloa, HI, USA, 2015; pp. 262–269. doi:10.1109/WACV.2015.42.
27. Qin, T.; Shen, S. Robust Initialization of Monocular Visual-Inertial Estimation on Aerial Robots. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); IEEE: Vancouver, BC, 2017; pp. 4225–4232. doi:10.1109/IROS.2017.8206284.
28. Li, J.; Bao, H.; Zhang, G. Rapid and Robust Monocular Visual-Inertial Initialization with Gravity Estimation via Vertical Edges. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); IEEE: Macau, China, 2019; pp. 6230–6236. doi:10.1109/IROS40897.2019.8968456.
29. Campos, C.; Montiel, J.M.M.; Tardós, J.D. Inertial-Only Optimization for Visual-Inertial Initialization. *arXiv:2003.05766 [cs]* **2020**, [arXiv:cs/2003.05766]. Comment: 2020 International Conference on Robotics and Automation.
30. Loeliger, H.A. An Introduction to Factor Graphs. *Teacher* **2008**, pp. 245–246. doi:10.1109/MSP.2004.1267047.
31. Dellaert, F.; Kaess, M. Factor Graphs for Robot Perception. *Foundations and Trends in Robotics* **2017**, *6*, 1–139. doi:10.1561/2300000043.
32. Indelman, V.; Williams, S.; Kaess, M.; Dellaert, F. Information Fusion in Navigation Systems via Factor Graph Based Incremental Smoothing. *Robotics and Autonomous Systems* **2013**, *61*, 721–738. doi:10.1016/j.robot.2013.05.001.
33. Kaess, M.; Johannsson, H.; Roberts, R.; Ila, V.; Leonard, J.J.; Dellaert, F. iSAM2: Incremental Smoothing and Mapping Using the Bayes Tree. *The International Journal of Robotics Research* **2012**, *31*, 216–235. doi:10.1177/0278364911430419.
34. Farrell, J. *Aided Navigation: GPS with High Rate Sensors*; Electronic Engineering, McGraw-Hill: New York, 2008. OCLC: ocn212908814.
35. Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D. On-Manifold Preintegration for Real-Time Visual-Inertial Odometry. *IEEE Transactions on Robotics* **2017**, *33*, 1–21. doi:10.1109/TRO.2016.2597321.
36. Godha, S. Performance evaluation of low cost MEMS-based IMU integrated with GPS for land vehicle navigation application. 2006.
37. Horn, B.K.P. Closed-Form Solution of Absolute Orientation Using Unit Quaternions. *Journal of the Optical Society of America A* **1987**, *4*, 629. doi:10.1364/JOSAA.4.000629.

38. Rehder, J.; Nikolic, J.; Schneider, T.; Hinzmann, T.; Siegwart, R. Extending Kalibr: Calibrating the Extrinsics of Multiple IMUs and of Individual Axes. 2016 IEEE International Conference on Robotics and Automation (ICRA); IEEE: Stockholm, Sweden, 2016; pp. 4304–4311. doi:10.1109/ICRA.2016.7487628.
39. Burri, M.; Nikolic, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Achtelik, M.W.; Siegwart, R. The EuRoC Micro Aerial Vehicle Datasets. *The International Journal of Robotics Research* **2016**, *35*, 1157–1163. doi:10.1177/0278364915620033.

Sample Availability: Samples of the compounds ..... are available from the authors.