ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XI-1-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission I, 4–11 July 2026, Toronto, Canada

GNSS-Constrained Motion Estimation for Robust Visual-Inertial-Odometry Initialization

Chunqi Dai, Sagi Filin

Mapping and Geo-Information Engineering, Technion - Israel Institute of Technology, Haifa, Israel
chunqi.dai@campus.technion.ac.il, filin@technion.ac.il

Keywords: Visual-Inertial Odometry, Global Navigation Satellite System, State Initialization, Multi-Sensor Fusion, Pose Estimation

Abstract

Visual-inertial odometry (VIO) plays a key role in modern navigation and mapping systems. For their successful integration, an initialization phase, in which IMU-related bias factors are estimated, becomes a fundamental step. Without one, the subsequent nonlinear estimation of the platform pose may fail to converge or completely diverge. As reliance on visual and inertial information may exhibit instability due to error accumulation with time, incorporating absolute positioning information through global navigation satellite system (GNSS) measurements, may enhance its robustness and accuracy. Accordingly, GNSS and visual-inertial initialization frameworks have been receiving growing attention in recent years where current strategies tend to follow a loosely-coupled formulation that first initializes the VIO trajectory, and then aligns it with GNSS measurements. Such strategies are multi-stage, nonlinear, and computationally expensive, motivating us to introduce an alternative framework in which GNSS position is integrated with the raw visual-inertial measurements to form absolute translation constraints. Accordingly, we achieve a closed-form, linear and globally consistent drift-free solution which is computationally efficient and requires neither 3D reconstruction nor nonlinear refinement, as common approaches do. Testing our initialization formulation on benchmark multi-sensor datasets, results show that we outperform current baselines while exhibiting robustness in challenging scenarios.

1. Introduction

Advances in imaging and processing technologies are leading to an increased focus on real-time pose estimation and mapping for a broad range of applications, including autonomous driving, robotics, and augmented reality (Huai and Huang, 2022; Cheng et al., 2023; Li et al., 2024). Such frameworks usually involve multi-sensor fusion that combines visual, inertial, and global navigation satellite system (GNSS) measurements. Inertial sensors provide high-rate motion measurements but tend to drift over time. Visual sensors provide rich environmental information but are sensitive to changes in the illumination conditions, texture, and motion blur, while GNSS measurements provide global position information, but their signals can be unreliable or unavailable in certain environments. To compensate for their shortcomings and leverage the complementary strengths of each, sensor fusion strategies have been proposed over the years. Among them, the integration of GNSS with visual inertial odometry (VIO) has been widely investigated as an effective solution for robust and accurate pose estimation. In such systems, VIO provides continuous high-frequency motion estimates, while GNSS offers globally referenced position information, enabling reliable navigation in outdoor and partially degraded environments such as dense forests or urban canyons. Nevertheless, the performance of these integrated systems strongly depends on a proper initialization. An inaccurate one is likely to lead to poor convergence or even divergence in the subsequent nonlinear optimization stage, making reliable initialization a critical component of GNSS/VIO integration.

VIO initialization strategies can be partitioned into loose- and tight-coupling categories (Dong-Si and Mourikis, 2012; Mur-Artal and Tardós, 2017b; Qin et al., 2018; Domínguez-Conti et al., 2018; Campos et al., 2021). Loosely-coupled strategies compute first the camera motion from structure-from-motion (SfM) or visual odometry, and then use inertial measurement unit (IMU) preintegration to align the visual and inertial trajectories. They are simple, but sensitive to the presence of low-texture or fast-rotation scenarios, where SfM often fails. In contrast, tightly-coupled formulations jointly exploit both visual and inertial measurements to compute the initial motion states. With the inclusion of GNSS measurements, additional initialization strategies arise due to the newly introduced absolute positional constraints (Cao et al., 2022). This additional information can align the local VIO frame to a global one either via the computed global positions or by fusing the raw GNSS and visual-inertial measurements (Lee et al., 2020; Jin et al., 2021; Niu et al., 2022). Though promising, current fusion methods predominantly involve nonlinear and complex multi-stage optimization, which can be computationally expensive (Jin et al., 2021; Cao et al., 2022).

To achieve efficient and accurate initialization, this paper introduces a new tightly coupled framework that integrates GNSS and visual-inertial measurements and improves the sensor fusion. Here, camera and IMU rotations alignment sets the IMU gyro bias estimate, while GNSS positions are formulated to constrain the initial velocity and gravity vectors. As we demonstrate, the resulting initialization is both globally consistent and computationally efficient, requiring neither 3D reconstruction nor nonlinear refinement. Our main contributions are as follows: *i*) a tightly coupled integration of global GNSS positions and a rotation-translation-decoupled VIO framework, enabling metric-consistent and drift-free initialization; *ii*) the establishment of a globally referenced initialization for pose estimation, jointly estimating velocity and gravity in a linear, closed-form manner; and *iii*) robustness and higher accuracy than existing initialization baselines as our experiments on multi-sensor datasets demonstrate. Our initialization framework yields 60% improvement in scale and gravity direction estimation over existing baselines, exhibiting also enhanced robustness in challenging scenarios.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-143-2026 | © Author(s) 2026. CC BY 4.0 License.

2. Related Work

2.1 Visual-Inertial Odometry Initialization

VIO plays a pivotal role in modern navigation and mapping systems due to its low cost and ability to provide high-frequency motion estimation (Mur-Artal and Tardós, 2017b; Qin et al., 2018; Campos et al., 2021). Accordingly, VIO has been extensively studied with many methods addressing both online state estimation and initialization (Mur-Artal and Tardós, 2017a; Demmel et al., 2021; Zhang et al., 2025). The main challenge in VIO initialization lies in accurately estimating the initial system states from noisy visual and inertial measurements. Early work, e.g., Mourikis and Roumeliotis (2007) employed static initialization that assumed zero initial velocity and used IMU data to estimate the gravity and IMU biases. This simplified assumption failed in dynamic scenarios where the platform moved during initialization. Later frameworks based their initialization on loosely- or tightly-coupled approaches. Martinelli (2014) proposed a tightly-coupled closed-form SfM that jointly estimated velocity, gravity, and scale using both visual and inertial data. Nonetheless, as Kaiser et al. (2016) and Domínguez-Conti et al. (2018) demonstrated, ignoring the gyroscope bias dropped the accuracy when using low-cost IMUs. Qin et al. (2018) estimated the initial states by introducing a loosely-coupled initialization that first computed the camera motion using SfM and then aligned it with IMU rotations. Later, Campos et al. (2021) considered the IMU measurement uncertainties and used the maximum a posteriori (MAP) estimation to robustify the initialization. To improve efficiency and accuracy, He et al. (2023) proposed a rotation-translation-decoupled framework. The authors estimated first the gyroscope bias through a rotation-only optimization using camera observations, and then solved for the initial velocity and gravity vectors using linear translation constraints. This decoupled formulation substantially improved both computational efficiency and robustness, but still suffered from scale drift due to the lack of global position constraints. Merrill et al. (2025) used scale-less single-image depth to reduce the number of feature parameters to the scale and bias of the depth map, which accelerated the initialization phase.

2.2 GNSS and VIO Integration

Due to the value in combining the complementary strengths of the different sensors, GNSS/VIO integration has been receiving increased attention in recent years (Gu et al., 2022; Cao et al., 2022; Cremona et al., 2024; Gu et al., 2025). Current initialization methods tend to focus on providing VIO with initial pose from GNSS or GNSS/inertial navigation system (INS) fusion, or aligning the VIO trajectory with the global frame to obtain the initial velocity and attitude. Yu et al. (2019) proposed a global positioning system (GPS)-aided visual-inertial system, in which the IMU was loosely coupled with the visual measurements for VIO initialization, and then the VIO trajectory was aligned with GPS measurements to obtain the initial velocity and attitude. Lee et al. (2020) aligned the VIO frame with the GNSS global frame where only the yaw angle was being considered. To make GNSS measurements more supportive for VIO initialization, Jin et al. (2021) proposed a fast initialization method that aligned the GNSS/INS trajectory with that of the VIO via nonlinear optimization. Similarly, Niu et al. (2022) used GNSS to initialize the IMU biases, then the INS pose was used to aid VIO initialization. However, visual measurements were not fully exploited during the initialization stage. Cao et al. (2022) presented a multi-stage coarse-to-fine GNSS/VIO initialization method that first estimated coarse GNSS positions, then calibrated the yaw offset, and finally refined all states via nonlinear optimization. Their VIO initialization was loosely coupled, relying on the SfM to provide camera motion. Recently, Zhang et al. (2024) and Hu et al. (2025) used GNSS measurements to initialize the IMU biases, then used the IMU pose as the initial states of VIO. As the review demonstrates, multi-stage nonlinear optimization is often involved in these frameworks, which can be computationally expensive. In addition, these models can also be overly complex for practical applications, limiting their usability and generality.

3. Methodology

Given GNSS, IMU and camera measurements, our initialization process consists of two main steps: *i*) IMU gyroscope bias estimation using camera observations, and *ii*) scale, initial velocity, and gravity vector estimation through GNSS, inertial, and visual measurements integration. Here, the gyroscope bias estimation is solved first through a rotation-only optimization using camera observations, while the accelerometer bias, which is small in effect (He et al., 2023) is ignored. Then, the scale, initial velocity, and gravity vectors are estimated by the three following steps: *i*) estimation of the local initial velocity and gravity vectors using visual translation constraints, sans GNSS measurements; *ii*) alignment of the initial velocity and gravity directions with ones derived from GNSS measurements, thereby computing the rotation matrix from the global frame to the IMU frame at the first keyframe to align the local frame with the global one; and *iii*) incorporation of both the GNSS-derived and visual constraints into a linear initialization framework to obtain globally consistent and drift-free scale, initial velocity, and gravity vectors. Our overall initialization framework is illustrated in Fig. (1).

We use first a tightly coupled formulation to jointly exploit both visual and inertial data. The IMU integration follows a standard approach, which can be expressed as (Forster et al., 2016):

$$\begin{aligned}
\mathbf{p}_{b_1 b_k} &= \mathbf{p}_{b_1 b_i} + \mathbf{v}_{b_1}\Delta t_{ik} - \frac{1}{2}\mathbf{g}_{b_1}\Delta t_{ik}^2 + \mathbf{R}_{b_1 b_i}\boldsymbol{\alpha}_{b_k}^{b_i} \\
\mathbf{v}_{b_1 b_k} &= \mathbf{v}_{b_1 b_i} - \mathbf{g}^{b_1}\Delta t_{ik} + \mathbf{R}_{b_1 b_i}\boldsymbol{\beta}_{b_k}^{b_i} \\
\mathbf{R}_{b_1 b_k} &= \mathbf{R}_{b_1 b_i}\boldsymbol{\gamma}_{b_k}^{b_i}
\end{aligned} \tag{1}$$

where $\mathbf{p}_{b_1 b_k}$, $\mathbf{v}_{b_1 b_k}$, and $\mathbf{R}_{b_1 b_k}$ are the respective position, velocity vector, and rotation matrix from IMU frame b_1 to b_k; $\mathbf{g}_{b_1}$ and $\mathbf{v}_{b_1}$ are the *unknown* initial gravity and velocity vectors in the IMU frame b_1 to be estimated; Δt_{ik} is the time interval between IMU frames b_i and b_k; and $\boldsymbol{\alpha}_{b_k}^{b_i}$, $\boldsymbol{\beta}_{b_k}^{b_i}$, and $\boldsymbol{\gamma}_{b_k}^{b_i}$ are the IMU preintegration terms between IMU frames b_i and b_k, such that:

$$\begin{aligned}
\boldsymbol{\alpha}_{b_k}^{b_i} &= \sum_{j=i}^{k-1}\left(\left(\sum_{f=i}^{j-1}\mathbf{R}_{b_i b_f}\mathbf{a}_f^{imu}\Delta t\right)\Delta t + \frac{1}{2}\mathbf{R}_{b_i b_j}\mathbf{a}_j^{imu}\Delta t^2\right) \\
\boldsymbol{\beta}_{b_k}^{b_1} &= \sum_{j=i}^{k-1}\mathbf{R}_{b_i b_j}\mathbf{a}_j^{imu}\Delta t \\
\boldsymbol{\gamma}_{b_k}^{b_1} &= \prod_{j=i}^{k-1}\exp\left(\boldsymbol{\omega}_j^{imu}\Delta t\right)
\end{aligned} \tag{2}$$

where $\mathbf{a}_j^{imu}$ and $\boldsymbol{\omega}_j^{imu}$ are the specific force and angular velocity measurements from the IMU at time step j; and Δt is the IMU sampling interval.

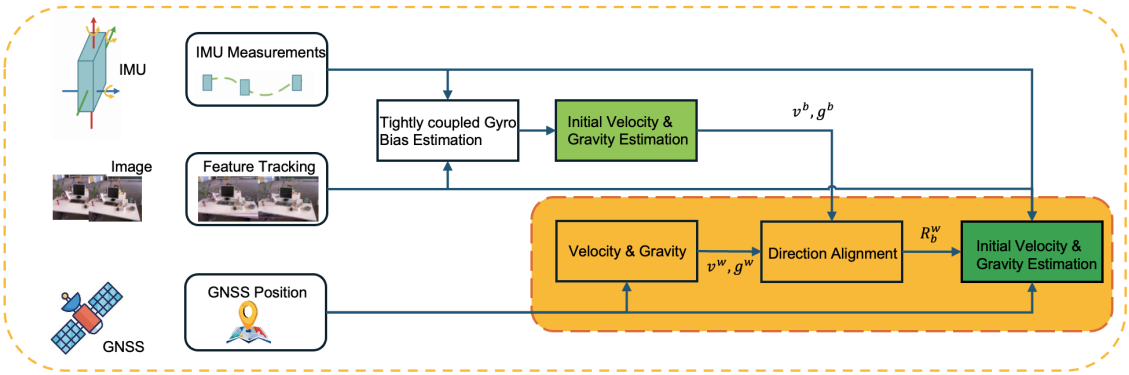

Figure 1. The framework of GNSS-constrained visual-inertial initialization. There are two initial velocity and gravity vector estimation steps: one in local without GNSS constraints and one with GNSS constraints. In the orange box is our proposed GNSS-constrained translation estimation contribution.

Visual measurements provide relative translation constraints between keyframes up to an unknown scale-factor but can be used to constrain the initial velocity and gravity vectors. To estimate them we follow He et al. (2023) and make use of the linear global translation (LiGT, Cai et al., 2021) form. The relative translations between three camera keyframes c_l, c_m and c_r with respect to c_1 can be expressed as:

$$\mathbf{B}\mathbf{p}_{c_1c_l} + \mathbf{C}\mathbf{p}_{c_1c_m} + \mathbf{D}\mathbf{p}_{c_1c_r} = \mathbf{0}, 1 \leq m \leq N, m \neq r \quad (3)$$

where,

$$\begin{aligned} \mathbf{B} &= [\mathbf{X}_m]_\times \mathbf{R}_{c_mc_r} \mathbf{X}_r \boldsymbol{a}_{rl}^T \mathbf{R}_{c_l} \\ \mathbf{C} &= \theta_{rl}^2 [\mathbf{X_m}]_\times \mathbf{R}_m \\ \mathbf{D} &= -(\mathbf{B} + \mathbf{C}) \\ \boldsymbol{a}_{rl}^T &= \left([\mathbf{R}_{c_rc_l}\mathbf{X}_l]_\times \mathbf{X}_l\right)^T [\mathbf{X}_l]_\times \\ \theta_{rl} &= \|[\mathbf{X}_l]_\times \mathbf{R}_{c_rc_l} \mathbf{X}_r\| \end{aligned} \quad (4)$$

and where $\mathbf{p}_{c_1c_l}$, $\mathbf{p}_{c_1c_m}$, and $\mathbf{p}_{c_1c_r}$ are the camera positions at keyframes c_l, c_m, and c_r, respectively; $\mathbf{R}_{c_mc_r}$ is the relative rotation from camera frame c_m to c_r; $\mathbf{X}_l$, $\mathbf{X}_m$, and $\mathbf{X}_r$ are the normalized image coordinates of the matched feature point in camera frames c_l, c_m, and c_r; and $[\cdot]_\times$ denotes the skew-symmetric matrix.

To establish a tightly coupled form, the IMU integration equations and visual translation constraints (Eqs. 1 & 3) are combined to yield a linear system for estimating the initial velocity and gravity vectors. The spatial relation between the camera and IMU frames is defined by a known extrinsic calibration, allowing us to express the camera positions in terms of the IMU frame:

$$\begin{aligned} \mathbf{R}_{c_ic_j} &= \mathbf{R}_{bc}^T \mathbf{R}_{b_ib_j} \mathbf{R}_{bc} \\ \mathbf{p}_{c_ic_j} &= \mathbf{R}_{bc}^T \left(\mathbf{p}_{b_ib_j} + \mathbf{R}_{b_ib_j}\mathbf{p}_{bc} - \mathbf{p}_{bc}\right) \end{aligned} \quad (5)$$

where $\mathbf{R}_{bc}$ and $\mathbf{p}_{bc}$ are the rotation and translation from the IMU frame to the camera frame, known from extrinsic calibration. $\mathbf{R}_{b_ib_j}$ and $\mathbf{p}_{b_ib_j}$ are the rotation and translation from the IMU frame b_i to b_j. By substituting the IMU integration equations (Eq. 1) into the visual translation constraints (Eq. 3) and using Eq. (5) to express positions and rotations in the body frame, we obtain:

$$\mathbf{B}\mathbf{R}_{bc}^T \mathbf{p}'_{b_1b_l} + \mathbf{C}\mathbf{R}_{bc}^T \mathbf{p}'_{b_1b_m} + \mathbf{D}\mathbf{R}_{bc}^T \mathbf{p}'_{b_1b_r} = \mathbf{0} \quad (6)$$

where

$$\mathbf{p}'_{b_1b_k} = \mathbf{p}_{b_1b_k} + \mathbf{R}_{b_1b_k}\mathbf{p}_{bc} - \mathbf{p}_{bc}, \ k = l, m, r \quad (7)$$

The initial velocity, $\mathbf{v}_{b_1}$, and gravity vector, $\mathbf{g}_{b_1}$, are the only unknowns in Eq. (6), as all other terms can be computed from the IMU preintegration and visual measurements. By stacking all such constraints from multiple feature tracks across the sliding window of keyframes, we can formulate a linear system of equations in terms of the unknown $\mathbf{v}_{b_1}$ and $\mathbf{g}_{b_1}$:

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_{b_1} \\ \mathbf{g}_{b_1} \end{bmatrix} = \mathbf{L}_1 \quad (8)$$

where

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{B}\mathbf{R}_{bc}^T \Delta t_{1l} + \mathbf{C}\mathbf{R}_{bc}^T \Delta t_{1m} + \mathbf{D}\mathbf{R}_{bc}^T \Delta t_{1r} \\ \mathbf{A}_2 &= -\frac{1}{2}\left(\mathbf{B}\mathbf{R}_{bc}^T \Delta t_{1l}^2 + \mathbf{C}\mathbf{R}_{bc}^T \Delta t_{1m}^2 + \mathbf{D}\mathbf{R}_{bc}^T \Delta t_{1r}^2\right) \\ \mathbf{L} &= -\left(\mathbf{B}\mathbf{R}_{bc}^T \mathbf{s}_{1l} + \mathbf{C}\mathbf{R}_{bc}^T \mathbf{s}_{1m} + \mathbf{D}\mathbf{R}_{bc}^T \mathbf{s}_{1r}\right) \\ \mathbf{s}_{1k} &= \boldsymbol{\alpha}_{b_k}^{b_1} + \mathbf{R}_{b_1b_k}\mathbf{p}_{bc} - \mathbf{p}_{bc} \end{aligned} \quad (9)$$

Eq. (8) forms a tightly coupled visual-inertial initialization framework. Then, turning to our GNSS-constrained initialization (orange box in Fig. 1), we incorporate the GNSS constraints into the translation estimation step by expressing the GNSS positional offset between two IMU frames as:

$$\mathbf{p}_{b_1b_k} = \mathbf{R}_{w_1b_1}^T \left(\mathbf{p}_{w_k}^G - \mathbf{p}_{w_1}^G\right), \ k = l, m, r \quad (10)$$

where $\mathbf{p}_{b_1b_k}$ is the relative position derived from IMU; $\mathbf{p}_{w_k}^G$ and $\mathbf{p}_{w_1}^G$ are the GNSS positions in the global frame w_k and w_1, respectively; and $\mathbf{R}_{w_1b_1}$ is the rotation from the global frame to the IMU frame at b_1. In typical platforms, the GNSS antenna is mounted close to the IMU, resulting in a short lever-arm. The displacement induced by rotational motion is therefore small and generally within the noise level of GNSS measurements. Consequently, the lever-arm effect has a negligible impact on the initialization process and is ignored for simplicity (Cahyadi et al., 2024). To formulate the GNSS-derived translation constraints, we need first to compute the rotation $\mathbf{R}_{w_1b_1}$ that aligns the global and IMU frames at b_1. This can be achieved using the gravity and initial velocity estimated from the visual-inertial initialization. The rotation can be computed by aligning

the velocity and gravity directions between the global and IMU frames at b_1:

$$\mathbf{R}_{w_1 b_1} \begin{bmatrix} \mathbf{v}_{b_1} & \mathbf{g}_{b_1} & \mathbf{v}_{b_1} \times \mathbf{g}_{b_1} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_{w_1} & \mathbf{g}_{w_1} & \mathbf{v}_{w_1} \times \mathbf{g}_{w_1} \end{bmatrix} \tag{11}$$

where $\mathbf{v}_{w_1}$ can be approximated by the GNSS-derived velocity at IMU frame b_1, and $\mathbf{g}_{w_1}$ is the known gravity direction in the global frame. Substituting the preintegration model (Eq. 1) into Eq. (10), the GNSS-derived translation constraints can be expressed as:

$$\mathbf{v}_{b_1} \Delta t_{ik} - \frac{1}{2} \mathbf{g}_{b_1} \Delta t_{ik}^2 = \mathbf{R}_{w_1 b_1}^T \left(\mathbf{p}_{w_k}^G - \mathbf{p}_{w_1}^G \right) - \mathbf{p}_{b_1} - \boldsymbol{\alpha}_{b_k}^{b_1} \tag{12}$$

By stacking all such constraints from multiple keyframes, we can formulate another linear system for the initial velocity and gravity vector as:

$$\begin{bmatrix} \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix} \begin{bmatrix} \mathbf{v}_{b_1} \\ \mathbf{g}_{b_1} \end{bmatrix} = \mathbf{L}_2 \tag{13}$$

where

$$\begin{aligned} \mathbf{A}_3 &= \Delta \mathbf{t}_{1k} \\ \mathbf{A}_4 &= -\frac{1}{2} \Delta \mathbf{t}_{1k}^2 \\ \mathbf{L}_2 &= \mathbf{R}_{w_1 b_1}^T \left(\mathbf{p}_{w_k}^G - \mathbf{p}_{w_1}^G \right) - (\boldsymbol{\alpha}_{b_k}^{b_1} + \mathbf{p}_{b_1}) \end{aligned} \tag{14}$$

By combining the GNSS, inertial and visual constraints, we arrive at our the final linear system form:

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix} \begin{bmatrix} \mathbf{v}_{b_1} \\ \mathbf{g}_{b_1} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}2 \end{bmatrix}. \tag{15}$$

This augmented system ensures that the translation estimation is both drift-free and metrically consistent, leveraging the absolute position information provided by GNSS.

4. Experiments

We use the GVINS datasets (Cao et al., 2022) as our benchmark for evaluating our GNSS-constrained visual-inertial initialization model. These datasets provide synchronized GNSS, IMU, and stereo camera data collected in different outdoor environments, including urban and suburban areas. Two of them are used here, the *Sports Field* and the *Complex Environment*, where typical scenes of both are presented in Fig. (2).[1] The *Sports Field* was acquired while moving along an athletic track for five laps. It features open sky with strong GNSS signals and moderate motion dynamics. Nevertheless, significant parts of each image are featureless due to the sky and sports field layout, making visual feature extraction and tracking challenging. The second dataset was collected in a complex urban environment with tall buildings and trees. GNSS signals are intermittent due to occlusions, while illumination is uneven due to shadows, which can degrade visual feature tracking. Therefore, it presents a more challenging initialization scenario due to the presence of GNSS outages and visual degradation.

Initialization methods We use the GVINS (Cao et al., 2022) and DRT-t (He et al., 2023) algorithms as baselines for comparison. As our method extends the DRT-t framework by integrating GNSS constraints into the translation estimation, it is a natural choice for testing the benefits these measurements bring. GVINS is a state-of-the-art GNSS/VIO integration method, whose initialization is the baseline against which we test our formulation. The GNSS settings follow those in Cao et al. (2022) for a fair comparison. We divide the two datasets into multiple segments and perform initialization on each segment independently. We use only the left camera as our focus is on monocular VIO initialization and for fair comparison with DRT-t. All algorithms adopted the same image processing operations including corner feature detection and tracking algorithm. We evaluate the initialization accuracy using three metrics: scale, gravity direction, and velocity error. The improvement is calculated as the percentage reduction in mean and RMS compared to DRT-t.

Figure 2. Typical scenarios in the *Sports Field* (a) and *Complex Environment* (b) datasets.

Initialization results for the *Sports Field* dataset are listed in Table (1). Our proposed method shows significant improvements in accuracy compared with DRT-t, alluding to effectiveness of integrating GNSS constraints into the translation estimation step. Though GVINS includes GNSS constraints in its multi-stage optimization, it does not fully utilize them during the initialization of the scale, initial velocity, and gravity states, while our model integrates these constraints directly into the translation estimation step. As our integration is direct, we achieve up to 80% improvement in accuracy compared to it. The initialization error box plots (Fig. 3) show how we outperform both baselines with fewer outliers. The results in all three metrics demonstrate more compact error distributions, highlighting our improvement in accuracy and robustness. Though DRT-t shows a smaller first quartile and median gravity errors, ours shows fewer outliers, and therefore higher reliability of the estimated parameters. We also achieve smaller quartile ranges and median errors and fewer outliers compared to GVINS.

The results of the more challenging *Complex Environment* dataset are listed in Table (2). Here again, utilizing the available GNSS constraints during translation estimation yields signific-

[1] Notably, a third one exists, but being collected at dusk and night, it is unsuitable for our evaluation.

Table 1. Comparison of initialization errors on the *Sports Field* dataset.

Method	**Scale Error (-)**				**Gravity Error (deg)**				**Velocity Error (m/s)**			
	Mean	Improve	RMS	Improve	Mean	Improve	RMS	Improve	Mean	Improve	RMS	Improve
DRT-t	0.65		0.67		1.14		2.84		0.47		0.55	
GVINS	0.35	46.2%	0.41	38.8%	1.04	8.8%	1.21	57.4%	0.25	46.8%	0.34	38.2%
Our method	**0.06**	90.8%	**0.06**	91.0%	**0.77**	32.5%	**0.85**	70.1%	**0.16**	66.0%	**0.19**	65.5%

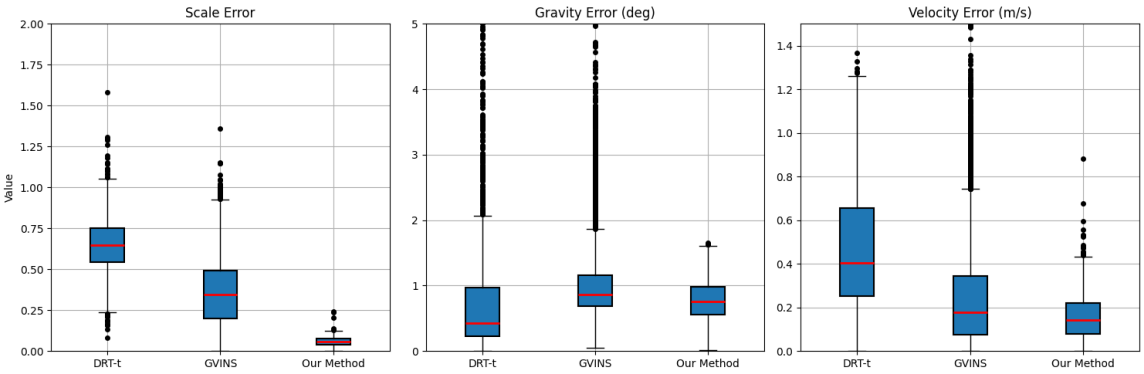

Figure 3. Scale, velocity, and gravity error box plots on the *Sports Field* dataset.

ant improvements compared to DRT-t. Concurrently, as our method fully exploits the GNSS constraints during translation estimation, it also shows improvement in scale and gravity estimation compared to GVINS. The initialization errors box plots (Fig. 4) show how we outperform both baselines in scale, gravity, and velocity estimation accuracy with more compact error distributions, and smaller quartile ranges and median errors in scale and gravity. DRT-t and GVINS show similar performance in gravity estimation, whereas ours shows a more compact error distribution with fewer outliers, leading to a significant improvement in gravity estimation accuracy. Overall, the error distribution is wider than that of the *Sports Field* dataset, which is due to intermittent GNSS signals and more challenging visual conditions which degrade the performance of all methods.

In both test cases, our proposed GNSS-constrained visual-inertial initialization method consistently outperforms the baselines in terms of scale, gravity, and velocity estimation accuracy. It demonstrates the effectiveness of integrating GNSS constraints directly, through a tightly coupled formulation, into the translation estimation step, providing a robust and metrically consistent initialization for GNSS/VIO integrated systems. These accurate scale, gravity and velocity initializations are expected to contribute to lower trajectory errors in mapping and navigation.

5. Conclusion

This paper presented a GNSS-constrained visual-inertial initialization method that integrates GNSS position measurements into the rotation-translation-decoupled framework. We jointly estimated the initial velocity and gravity vector in a linear, closed-form manner, leveraging absolute position constraints from GNSS measurements to achieve drift-free and metrically consistent initialization. Our proposed initialization solution is verified on different outdoor datasets in open-sky and shaded environments. The results show that our method consistently outperforms existing VIO initialization baselines, demonstrating significant improvements in scale, gravity, and velocity estimation accuracy. These improvements lead to a more reliable initial state for subsequent optimization or filtering, and thus to improved trajectory accuracy, faster estimator convergence, and more reliable operation in real-world applications e.g., autonomous navigation and visual mapping.

6. Appendix

We show the detailed derivation of Eq. (8) in this section. First, substitute Eqs. (1) & (7) into Eq. (6):

$$
\begin{aligned}
&\mathbf{B}\mathbf{R}_{bc}^{T}\left(\mathbf{p}_{b_1b_l}+\mathbf{R}_{b_1b_l}\mathbf{p}_{bc}-\mathbf{p}_{bc}\right)+\mathbf{C}\mathbf{R}_{bc}^{T}(\mathbf{p}_{b_1b_m}+\mathbf{R}_{b_1b_m}\mathbf{p}_{bc}\\
&-\mathbf{p}_{bc})+\mathbf{D}\mathbf{R}_{bc}^{T}\left(\mathbf{p}_{b_1b_r}+\mathbf{R}_{b_1b_r}\mathbf{p}_{bc}-\mathbf{p}_{bc}\right)\\
=&\mathbf{B}\mathbf{R}_{bc}^{T}\left(\mathbf{v}_{b_1}\Delta t_{1l}-\frac{1}{2}\mathbf{g}_{b_1}\Delta t_{1l}^{2}+\boldsymbol{\alpha}_{b_l}^{b_1}+\mathbf{R}_{b_1b_l}\mathbf{p}_{bc}-\mathbf{p}_{bc}\right)+\\
&\mathbf{C}\mathbf{R}_{bc}^{T}\left(\mathbf{v}_{b_1}\Delta t_{1m}-\frac{1}{2}\mathbf{g}_{b_1}\Delta t_{1m}^{2}+\boldsymbol{\alpha}_{b_m}^{b_1}+\mathbf{R}_{b_1b_m}\mathbf{p}_{bc}-\mathbf{p}_{bc}\right)+\\
&\mathbf{D}\mathbf{R}_{bc}^{T}\left(\mathbf{v}_{b_1}\Delta t_{1r}-\frac{1}{2}\mathbf{g}_{b_1}\Delta t_{1r}^{2}+\boldsymbol{\alpha}_{b_r}^{b_1}+\mathbf{R}_{b_1b_r}\mathbf{p}_{bc}-\mathbf{p}_{bc}\right)
\end{aligned}
\tag{16}
$$

Denoting $\boldsymbol{s}_{1k}=\boldsymbol{\alpha}_{b_k}^{b_1}+\mathbf{R}_{b_1b_k}\mathbf{p}_{bc}-\mathbf{p}_{bc}$, we can rearrange the above equation as:

$$
\begin{aligned}
&\mathbf{B}\mathbf{R}_{bc}^{T}\left(\mathbf{v}_{b_1}\Delta t_{1l}-\frac{1}{2}\mathbf{g}_{b_1}\Delta t_{1l}^{2}+\boldsymbol{s}_{1l}\right)+\\
&\mathbf{C}\mathbf{R}_{bc}^{T}\left(\mathbf{v}_{b_1}\Delta t_{1m}-\frac{1}{2}\mathbf{g}_{b_1}\Delta t_{1m}^{2}+\boldsymbol{s}_{1m}\right)+\\
&\mathbf{D}\mathbf{R}_{bc}^{T}\left(\mathbf{v}_{b_1}\Delta t_{1r}-\frac{1}{2}\mathbf{g}_{b_1}\Delta t_{1r}^{2}+\boldsymbol{s}_{1r}\right)\\
=&\left(\mathbf{B}\mathbf{R}_{bc}^{T}\Delta t_{1l}+\mathbf{C}\mathbf{R}_{bc}^{T}\Delta t_{1m}+\mathbf{D}\mathbf{R}_{bc}^{T}\Delta t_{1r}\right)\mathbf{v}_{b_1}-\\
&\left(\frac{1}{2}\mathbf{B}\mathbf{R}_{bc}^{T}\Delta t_{1l}^{2}+\frac{1}{2}\mathbf{C}\mathbf{R}_{bc}^{T}\Delta t_{1m}^{2}+\frac{1}{2}\mathbf{D}\mathbf{R}_{bc}^{T}\Delta t_{1r}^{2}\right)\mathbf{g}_{b_1}=\\
&-\left(\mathbf{B}\mathbf{R}_{bc}^{T}\mathbf{s}_{1l}+\mathbf{C}\mathbf{R}_{bc}^{T}\mathbf{s}_{1m}+\mathbf{D}\mathbf{R}_{bc}^{T}\mathbf{s}_{1r}\right)=\mathbf{0}
\end{aligned}
\tag{17}
$$

Writing Eq. (17) in the matrix form, we obtain the linear system in Eq. (8).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-143-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 2. Comparison of initialization errors on the *Complex Environment* dataset.

Method	Scale Error (-)				Gravity Error (deg)				Velocity Error (m/s)			
	Mean	Improve	RMS	Improve	Mean	Improve	RMS	Improve	Mean	Improve	RMS	Improve
DRT-t	4.4		6.4		25.18		30.70		0.60		0.96	
GVINS	0.42	90.5%	0.48	92.5%	21.4	15.0%	26.08	15.0%	0.59	1.7%	0.92	4.2%
Our method	**0.12**	97.3%	**0.47**	92.7%	**4.77**	81.1%	**6.22**	79.7%	**0.52**	13.3%	**0.70**	27.1%

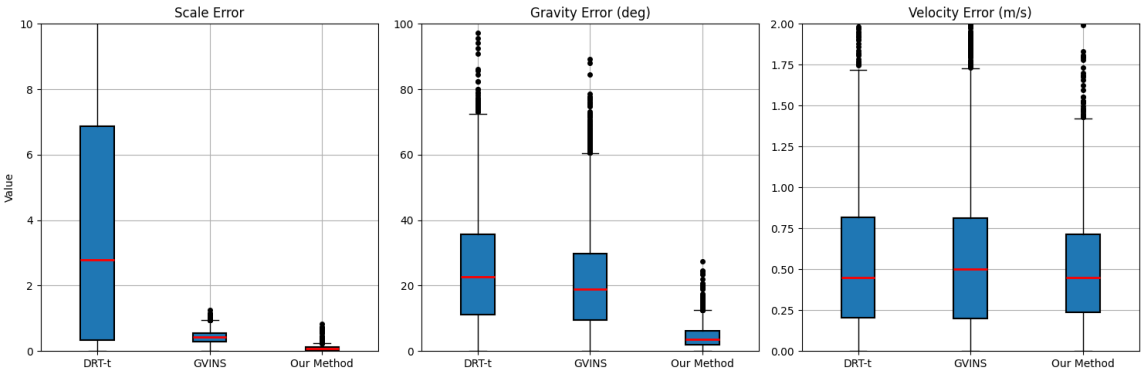

Figure 4. Scale, velocity, and gravity error box plots on the *Complex Environment* dataset.

References

Cahyadi, M. N., Asfihani, T., Suhandri, H. F., Erfianti, R., 2024. Unscented Kalman filter for a low-cost GNSS/IMU-based mobile mapping application under demanding conditions. *Geodesy and Geodynamics*, 15(2), 166–176.

Cai, Q., Zhang, L., Wu, Y., Yu, W., Hu, D., 2021. A pose-only solution to visual reconstruction and navigation. *IEEE TPAMI*, 45(1), 73–86.

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., Tardós, J. D., 2021. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE transactions on robotics*, 37(6), 1874–1890.

Cao, S., Lu, X., Shen, S., 2022. GVINS: Tightly coupled GNSS–visual–inertial fusion for smooth and consistent state estimation. *IEEE Transactions on Robotics*, 38(4), 2004–2021.

Cheng, J., Zhang, L., Chen, Q., Hu, X., Cai, J., 2023. Map aided visual-inertial fusion localization method for autonomous driving vehicles. *Measurement*, 221, 113432.

Cremona, J., Civera, J., Kofman, E., Pire, T., 2024. GNSS-stereo-inertial SLAM for arable farming. *Journal of field robotics*, 41(7), 2215–2225.

Demmel, N., Schubert, D., Sommer, C., Cremers, D., Usenko, V., 2021. Square root marginalization for sliding-window bundle adjustment. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13260–13268.

Domínguez-Conti, J., Yin, J., Alami, Y., Civera, J., 2018. Visual-inertial slam initialization: A general linear formulation and a gravity-observing non-linear optimization. *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, 37–45.

Dong-Si, T.-C., Mourikis, A. I., 2012. Estimator initialization in vision-aided inertial navigation with unknown camera-imu calibration. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 1064–1071.

Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D., 2016. On-manifold preintegration for real-time visual–inertial odometry. *IEEE Transactions on Robotics*, 33(1), 1–21.

Gu, S., Dai, C., Mao, F., Fang, W., 2022. Integration of multi-GNSS PPP-RTK/INS/vision with a cascading kalman filter for vehicle navigation in urban areas. *Remote Sensing*, 14(17), 4337.

Gu, S., Zhou, S., Song, W., Li, R., 2025. Feature augmented PPP-RTK/INS/Vision integration based on combinatorial optimization for urban vehicle navigation. *Measurement Science and Technology*, 36(5), 055101.

He, Y., Xu, B., Ouyang, Z., Li, H., 2023. A rotation-translation-decoupled solution for robust and efficient visual-inertial initialization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 739–748.

Hu, S., Liu, G., Lyu, M., Wang, R., Zhao, W., Zhang, B., 2025. Visual-Inertial-GNSS Fusion Positioning for Vehicles With Deep Learning-Based Feature Extraction and Outlier Detection. *IEEE Internet of Things Journal*.

Huai, Z., Huang, G., 2022. Robocentric visual–inertial odometry. *The International Journal of Robotics Research*, 41(7), 667–689.

Jin, R., Liu, J., Zhang, H., Niu, X., 2021. Fast and accurate initialization for monocular vision/INS/GNSS integrated system on land vehicle. *IEEE Sensors Journal*, 21(22), 26074–26085.

Kaiser, J., Martinelli, A., Fontana, F., Scaramuzza, D., 2016. Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation. *IEEE Robotics and Automation Letters*, 2(1), 18–25.

Lee, W., Eckenhoff, K., Geneva, P., Huang, G., 2020. Intermittent gps-aided vio: Online initialization and calibration. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 5724–5731.

Li, J., Pan, X., Huang, G., Zhang, Z., Wang, N., Bao, H., Zhang, G., 2024. RD-VIO: Robust visual-inertial odometry for mobile augmented reality in dynamic environments. *IEEE transactions on visualization and computer graphics*, 30(10), 6941–6955.

Martinelli, A., 2014. Closed-form solution of visual-inertial structure from motion. *International journal of computer vision*, 106(2), 138–152.

Merrill, N., Geneva, P., Katragadda, S., Chen, C., Huang, G., 2025. Fast and robust learned single-view depth-aided monocular visual-inertial initialization. *The International Journal of Robotics Research*, 44(10-11), 1619–1647.

Mourikis, A. I., Roumeliotis, S. I., 2007. A multi-state constraint kalman filter for vision-aided inertial navigation. *Proceedings 2007 IEEE international conference on robotics and automation*, IEEE, 3565–3572.

Mur-Artal, R., Tardós, J. D., 2017a. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE transactions on robotics*, 33(5), 1255–1262.

Mur-Artal, R., Tardós, J. D., 2017b. Visual-inertial monocular SLAM with map reuse. *IEEE Robotics and Automation Letters*, 2(2), 796–803.

Niu, X., Tang, H., Zhang, T., Fan, J., Liu, J., 2022. IC-GVINS: A robust, real-time, INS-centric GNSS-visual-inertial navigation system. *IEEE robotics and automation letters*, 8(1), 216–223.

Qin, T., Li, P., Shen, S., 2018. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE transactions on robotics*, 34(4), 1004–1020.

Yu, Y., Gao, W., Liu, C., Shen, S., Liu, M., 2019. A gps-aided omnidirectional visual-inertial state estimator in ubiquitous environments. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 7750–7755.

Zhang, X., Qu, A., Zhang, Y., Wei, J., Dai, Z., Wang, Y., Sun, J., 2024. Gnss/ins/visual integrated navigation algorithm and performance analysis based on factor graph optimization. *Proceedings of the 2024 9th International Conference on Cyber Security and Information Engineering*, 515–520.

Zhang, Y., Tang, F., Xu, Z., Wu, Y., Ma, P., 2025. PGD-VIO: A Plane-Aided RGB-D Inertial Odometry with Graph-Based Drift Suppression. *IEEE Robotics and Automation Letters*.