

GPS-aided Visual Wheel Odometry

Junlin Song¹, Pedro J. Sanchez-Cuevas², Antoine Richard¹ and Miguel Olivares-Mendez¹

Abstract—This paper introduces a novel GPS-aided visual-wheel odometry (GPS-VWO) for ground robots. The state estimation algorithm tightly fuses visual, wheeled encoder and GPS measurements in the way of Multi-State Constraint Kalman Filter (MSCKF). To avoid accumulating calibration errors over time, the proposed algorithm calculates the extrinsic rotation parameter between the GPS global coordinate frame and the VWO reference frame online as part of the estimation process. The convergence of this extrinsic parameter is guaranteed by the observability analysis and verified by using real-world visual and wheel encoder measurements as well as simulated GPS measurements. Moreover, a novel theoretical finding is presented that the variance of unobservable state could converge to zero for specific Kalman filter system. We evaluate the proposed system extensively in large-scale urban driving scenarios. The results demonstrate that better accuracy than GPS is achieved through the fusion of GPS and VWO. The comparison between extrinsic parameter calibration and non-calibration shows significant improvement in localization accuracy thanks to the online calibration.

Index Terms—Sensor Fusion, State Estimation, Kalman Filter

I. INTRODUCTION

The localization algorithm of mobile ground robots is an essential component in critical industries like industrial automation, autonomous driving and even planetary exploration. In the last two decades, the research community has been looking into high-precision and robust localization algorithms to enable the exploitation of autonomous vehicles to improve the efficiency and productivity of those industries. Typical sensor setups use camera, inertial measurement unit (IMU) and wheels' position encoder (also referred to as odometer in this paper) to solve this localization problem. It is well known that the fusion of camera and IMU measurements can provide accurate motion tracking. This 6-degrees of freedom (DoF) localization technique is called visual-inertial odometry (VIO). Previous studies have shown that the scale information of VIO system is not observable when the robot undergoes constant acceleration motion [1][2]. Some approaches additionally fuse wheel encoder measurements overcome this scale issue, as wheel-encoders can provide valid scale information [2][3].

We are interested in the fusion of camera and wheel encoder called visual-wheel odometry (VWO) to deal with situation where the IMU is unavailable or too noisy to be used, or the system undergoes degenerated motion profile.

Thus the system reliability is increased. Compared with the IMU, the wheel-encoders can directly obtain the velocity, thus avoiding the double integration of the acceleration from the IMU. Moreover, its measurements are not affected by biases, unlike the IMU. The localization result obtained by any previously mentioned odometry is doomed to drift, as these sensors only generate relative motion constraints. A straightforward strategy to prevent this drifting is to integrate global position measurements such as GPS.

To correlate GPS measurements with the position output of the VWO system, it is necessary to calculate or estimate the spatial transformation of the two reference frames, referred to as extrinsic parameters in the rest of the paper. In this study, we propose a tightly-coupled filter-based GPS-VWO system with online spatial calibration to fully exploit the information of those sensors. The observability analysis of the extrinsic parameters linked to the spatial calibration shows that it is reasonable to add this calibration as an observable variable to the state vector. The online correction of the spatial calibration improves the localization accuracy and avoids dragging large uncertainties over time. The contributions of this work are summarized as:

- To the best of our knowledge, this is the first time that a tightly-coupled filter-based GPS-VWO system is proposed to optimally fuse camera, wheel encoder and GPS measurements. The extrinsic parameter between GPS frame and VWO frame is calibrated online.
- Nonlinear observability theory is employed to analyze the observability of extrinsic parameter between the GPS global coordinate frame and the VWO reference frame. We validate the observability conclusion through simulation experiments.
- Observability analysis of extrinsic parameter reveals a novel theoretical finding that the variance of unobservable state could converge to zero. We provide a detailed proof for a simplified Kalman filter system.
- The proposed GPS-VWO system is evaluated on large-scale urban driving datasets. The drift of VWO is significantly reduced by GPS. Meanwhile, the system is resilient to GPS interruption and challenge GPS noise (range from 6m to 20m).
- Moreover, real experiments show that VIO easily suffers from constant velocity motion, while our VWO not. Compared to the state-of-the-art GPS-VIO [4], GPS-VWO achieves up to 66% accuracy improvements.

II. RELATED WORK

High-precision and drift-free localization is essential for the autonomy of ground robots. In recent years, tightly-

This research was supported by the European Union's Horizon 2020 project SESAME (grant agreement No 101017258).

¹ Space Robotics (SpaceR) Research Group, Int. Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg.

²Advanced Centre for Aerospace Technologies (CATEC) Seville, Spain.

coupled VIO approaches fusing measurements from camera and IMU have shown great success [5]. Practically, VIO algorithms can be classified into two branches, one is based on nonlinear optimization such as VINS-Mono [6] and ORB-SLAM3 [7], and the other is based on Kalman Filter such as the commonly used algorithm named Multi-State Constraint Kalman Filter (MSCKF) [8], [9]. The MSCKF-based VIO algorithm is appealing for platforms with limited computing resources, since it consumes much less computational resources, but yet obtains comparable accuracy with optimization-based competitors [10].

However, in some specific motion profiles, VIO algorithms can suffer from unobservability [1], that is the impossibility to observe specific elements of the state. A relevant case leading to unobservable scale is constant acceleration motion, which is very common for ground robot [2]. A detailed motion degradation analysis of VIO is also presented in [11]. Therefore, it is necessary to consider additional sensor modality to resolve this issue. Wheel-odometers can provide velocity and scale information. Hence, it makes sense to combine camera, IMU and wheel-encoders measurements in a tightly-coupled fashion [2], [3], [12]. A recent work [13] adapts the above strategy through the optional usage of IMU measurements. In [13], the motion manifold of ground robots is modeled in detail to improve the localization accuracy.

Compared with the three sensors mentioned above, GPS is able to provide drift-free absolute position information directly. In order to achieve high-quality global localization, it is natural to integrate GPS measurements with those sensor suits or their subsets in large-scale environments. As eluded to earlier, the focus of this paper is the tightly-coupled fusion of GPS measurements in a VWO system, which is not studied in the literature. First, compared to loosely-coupled algorithms [14], tightly-coupled algorithms provides higher localization accuracy, because tightly-coupled algorithms fully exploited the inner correlation of all measurements [15]. Second, in order to realize multi-sensor fusion, we need to calculate extrinsic parameter to model GPS measurements [4]. Third, it is critical to prove that the extrinsic parameter is observable before including this parameter to state vector [4]. If this parameter is observable, performing online estimation can reduce the initialization error, which helps to improve the accuracy of the estimator output. If not, it will deteriorate the estimation. Lie derivative is a readily-available tool for the observability analysis of nonlinear system [16]. This tool was previously used to analyze the observability of extrinsic parameter between camera and IMU in VIO systems [17], [18]. Here, we employ Lie derivative to analyze the observability of extrinsic parameter between the GPS reference frame and the VWO reference frame, and drawn non-trivial conclusion, which is different from [4].

III. PROBLEM FORMULATION

A. Notation

The coordinate frames used in this paper are illustrated in the Fig. 1a. $\{E\}$ is the East-North-Up (ENU) coordinate frame. The GPS measurements are described in this frame.

Usually, the first received GPS measurement is selected as the origin of $\{E\}$. $\{V\}$ denotes the reference frame of the VWO system. $\{O\}$ and $\{C\}$ denote the wheel-odometer frame and camera frame respectively. G represents the position of the receiving antenna of GPS. The studied vehicle follows the differential drive kinematics. Fig. 1b shows the relationship between the velocity of two differential (left and right) wheels. v_l and v_r are the velocity of left and right wheel respectively. The ICR represents the instantaneous center of rotation. The origin of $\{O\}$ is located at the center of the robot's wheels. b is the distance between the wheels.

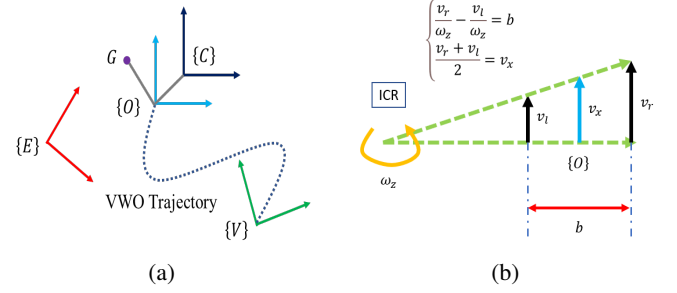


Fig. 1: (a) Coordinate frames. (b) The relationship between the velocity of two differential (left and right) wheels.

We use ${}^V(\bullet)$ to represent a physical quantity in the frame $\{V\}$. The position of a point O in the frame $\{V\}$ is expressed as Vp_O . The velocity of a point O in the frame $\{V\}$ is expressed as Vv_O . A Unit quaternion is employed to represent the rotation of the rigid body [19]. Oq represents the orientation of the frame $\{O\}$ with respect to the frame $\{V\}$, and its corresponding rotation matrix is O_R . $\{ {}^O_R \quad {}^Cp_O \}$ is the extrinsic parameter set between camera and wheel encoder. Op_G is the extrinsic parameter from the GPS sensor to frame $\{O\}$. These extrinsic parameters can be obtained from CAD model or calibrated before system running. This paper assumes that they are known. $[\bullet]_\times$ is denoted as the skew symmetric matrix corresponding to a three-dimensional vector. The transpose of a matrix is $[\bullet]^T$. e_i is a 3×1 vector, with the i th element to be 1 and other elements to be 0.

B. MSCKF based VWO

In this section, we first briefly review the pipeline of standard MSCKF [8]. The classical state estimation algorithm based on MSCKF usually defines a state vector containing the current state and multiple historical state clones:

$$\begin{aligned} x_O &= ({}^Oq^T \quad {}^Vp_O^T \quad {}^Vp_f^T)^T \\ x_{c_i} &= ({}^Oq_i^T \quad {}^Vp_{O_i}^T)^T \\ x &= (x_O^T \quad x_{c_1}^T \quad \cdots \quad x_{c_N}^T)^T \end{aligned} \quad (1)$$

Where x_O includes the position and orientation of the wheel-odometer frame $\{O\}$ at the current time and the visual feature position, Vp_f , in the reference frame of the VWO. To be concise, only one visual feature point is described here, namely f . This simplified case has no impact on the subsequent theoretical analysis. x_{c_i} is obtained by cloning the first two physical quantities of x_O at different times.

Then the state vector of the VWO system can be obtained by augmenting x_O with N historical clones of x_{c_i} . N is the sliding window size, a fixed parameter.

1) Wheel Odometer Propagation:

We propagate the state and covariance based on the kinematic model of a differential wheels vehicle taken from [20], [21]. The wheeled encoder equipped by the ground-robot provides local linear and angular velocity measurement (see Fig. 1b):

$$\begin{aligned} v_l &= \frac{\Delta m_l + n_l}{M_l \Delta t} \pi d_l \\ v_r &= \frac{\Delta m_r + n_r}{M_r \Delta t} \pi d_r \\ v_x &= \frac{v_l + v_r}{2} \\ \omega_z &= \frac{v_r - v_l}{b} \end{aligned} \quad (2)$$

Where Δm_l is the reading difference between two adjacent measurements of left wheel's encoder. Time difference between these two adjacent timestamps is denoted as Δt . The meaning of Δm_r is the same as Δm_l . n_l and n_r are zero mean Gaussian noise corresponding to Δm_l and Δm_r respectively. M_l and M_r represent the resolution of left and right wheels' encoders respectively. d_l and d_r represent the left and right wheels' diameters respectively. The above measurements can be used to compose the linear and angular velocity in frame $\{O\}$:

$${}^O\omega_O = \begin{bmatrix} n_{\omega_x} \\ n_{\omega_y} \\ \omega_z \end{bmatrix} \quad {}^Ov_O = \begin{bmatrix} v_x \\ n_{v_y} \\ n_{v_z} \end{bmatrix} \quad (3)$$

Where $n_{[\bullet]}$ represents the zero mean Gaussian noise of $[\bullet]$. The continuous-time kinematic model of wheel-odometer is given by:

$$\begin{aligned} {}^O\dot{q} &= \frac{1}{2} \Omega ({}^O\omega_O) {}^Oq \\ {}^V\dot{p}_O &= {}^O R^T {}^Ov_O \end{aligned} \quad (4)$$

Meanwhile,

$$\Omega(\omega) = \begin{bmatrix} -[\omega]^\times & \omega \\ \omega^T & 0 \end{bmatrix} \quad (5)$$

2) Visual Measurement Update:

We refer to Open-VINS [10] for the technical details of the visual measurement update. When a new image comes, the latest state is cloned and the state vector is augmented by this clone state. Then, the carefully selected feature points are utilized to update the poses over the sliding window and the position of the feature points, through camera measurement model and triangulation.

C. GPS Measurement Update

The GPS position measurement is expressed in the frame $\{E\}$. The measurement equation is modelled as ¹:

$$\begin{aligned} {}^E p_G &= {}^E p_V + {}^E_V R^V p_G \\ &= {}^E p_V + {}^E_V R ({}^V p_O + {}^O_V R^T {}^O p_G) \end{aligned} \quad (6)$$

¹Measurement equation (6) is used here just for the convenience of the observability analysis. In the implementation, we adopt the interpolation measurement equation as [4].

Where ${}^O p_G$ is the extrinsic parameter from the GPS sensor to frame $\{O\}$. ${}^V p_O$ and ${}^O_V R$ are described in section III-B. $\{ {}^E_V R \quad {}^E p_V \}$ is the extrinsic parameter set between the frame $\{V\}$ and the frame $\{E\}$.

The parameterization of ${}^E_V R$ will be discussed in detail. This paper mainly analyzes two parameterization methods. One is a 3DoF parameterization, while the other is 1DoF. One attitude angle from yaw, pitch, or roll can be selected as “**active DoF**”, while others are denoted as “**fixed DoF**”. For example, if yaw is used as “active DoF”, ${}^E_V R$ can be expressed in the following form:

$$\begin{aligned} {}^E_V R &= R_z(\alpha) R_y(\beta) R_x(\gamma) \\ &= \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} R_y(\beta) R_x(\gamma) \end{aligned} \quad (7)$$

Where α , β and γ are yaw, pitch and roll respectively. However, β and γ are fixed, marked as “fixed DoF”. α is the only variable can be changed during the update, marked as “active DoF”, thus 1DoF.

To analyse the observability of the extrinsic parameter between the frame $\{V\}$ and the frame $\{E\}$, this parameter is included to the state vector.

The new system state vector becomes:

$$x = (x_O^T \quad x_{c_1}^T \quad \cdots \quad x_{c_N}^T \quad {}^E_V q^T \quad {}^E p_V^T)^T \quad (8)$$

The subset of state variables related to the GPS measurement equation is noted as x_s :

$$x_s = [{}^O_V q^T \quad {}^V p_O^T \quad {}^E_V q^T \quad {}^E p_V^T]^T \quad (9)$$

The corresponding error state is expressed as:

$$\tilde{x}_s = [{}^O_V \tilde{\theta}^T \quad {}^V \tilde{p}_O^T \quad {}^E_V \tilde{\theta}^T \quad {}^E \tilde{p}_V^T]^T \quad (10)$$

The measurement Jacobian of GPS is calculated as:

$$\begin{aligned} H &= \frac{\partial {}^E \tilde{p}_G}{\partial \tilde{x}_s} \\ &= \begin{bmatrix} -{}^E_V R^O R^T [{}^O p_G]_\times & {}^E_V R & H_\theta & I_3 \end{bmatrix} \end{aligned} \quad (11)$$

If ${}^E_V \theta$ is a 3DoF parameterization:

$$H_\theta = [{}^E_V R^V p_G]_\times \quad (12)$$

If ${}^E_V \theta$ is a 1DoF parameterization:

$$H_\theta = \begin{bmatrix} -\sin \alpha & -\cos \alpha & 0 \\ \cos \alpha & -\sin \alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} R_y(\beta) R_x(\gamma) {}^V p_G \quad (13)$$

We omit the initialization of GPS-VWO system and the proper handling of time offset among multi-sensors, because these are not the main focus of this work. Interested readers can refer to [4].

IV. OBSERVABILITY ANALYSIS

The observability plays an important role in state estimation. In this section, we use Lie derivatives to perform the observability analysis of a nonlinear system [16]. To the best of our knowledge, this is the first time that a paper studies the observability of the extrinsic parameter of a GPS-VVO system. Subsequently, We will prove that the translation part of the extrinsic parameter is unobservable and the rotation part of the extrinsic parameter is unobservable when it is parameterized with 3DoF, yet observable when it is parameterized with 1DoF.

To simplify the formulation, the cloned states in the state vector are removed. The system state vector now becomes:

$$x = \begin{bmatrix} {}^O q^T & {}^V p_O^T & {}^V p_f^T & {}^E \theta^T & {}^E p_V^T \end{bmatrix}^T \quad (14)$$

The kinematic equation is rewritten into a suitable form to calculate the Lie derivatives.

$$\dot{x} = \begin{bmatrix} {}^O \dot{q} \\ {}^V \dot{p}_O \\ {}^V \dot{p}_f \\ {}^E \dot{\theta} \\ {}^E \dot{p}_V \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{2}\Xi({}^O q) \\ 0_3 \\ 0_3 \\ 0_{D \times 3} \\ 0_3 \end{bmatrix}}_{f_1} {}^O \omega_O + \underbrace{\begin{bmatrix} 0_3 \\ {}^O R^T \\ 0_3 \\ 0_{D \times 3} \\ 0_3 \end{bmatrix}}_{f_2} {}^O v_O \quad (15)$$

The value of D is either 3 or 1 depending on the number of DoF. In addition, here we use the following property of time derivative of quaternion:

$$\dot{q} = \frac{1}{2}\Omega(\omega)q = \frac{1}{2}\Xi(q)\omega \quad (16)$$

The definition of $\Xi(q)$ can be found in [19]. Next, we list the usable measurement equations. The camera measurement equation is:

$$h_1(x) = {}^C p_f = {}^C R_V^O R p + {}^C p_O \quad (17)$$

Where $p = {}^V p_f - {}^V p_O$.

The norm constraint of the unit quaternion is also considered as its own measurement equation:

$$h_2(x) = {}^O q^T {}^O q - 1 = 0 \quad (18)$$

The measurement equation of GPS is:

$$h_3(x) = {}^E p_G = {}^E p_V + {}^E R^V p_O \quad (19)$$

If the original equation (6) is adopted, the same analysis process will generate consistent conclusions. Therefore, assumption ${}^O p_G = 0_{3 \times 1}$ is made here to simplify the description.

1) Zeroth-Order Lie Derivatives:

The zeroth-order Lie derivative of a function is itself.

$$\begin{aligned} \mathcal{L}^0 h_1 &= {}^C p_O + {}^C R_V^O R p \\ \mathcal{L}^0 h_2 &= {}^O q^T {}^O q - 1 \\ \mathcal{L}^0 h_3 &= {}^E p_V + {}^E R^V p_O \end{aligned} \quad (20)$$

The gradients of the zeroth-order Lie derivatives with respect to x are:

$$\begin{aligned} \nabla \mathcal{L}^0 h_1 &= \begin{bmatrix} \Psi & -{}^C R_V^O R & {}^C R_V^O R & 0_{3 \times D} & 0_3 \end{bmatrix} \\ \nabla \mathcal{L}^0 h_2 &= \begin{bmatrix} 2{}^O q^T & 0_{1 \times 3} & 0_{1 \times 3} & 0_{1 \times D} & 0_{1 \times 3} \end{bmatrix} \\ \nabla \mathcal{L}^0 h_3 &= \begin{bmatrix} 0_{3 \times 4} & {}^E R & 0_3 & H_\theta & I_3 \end{bmatrix} \end{aligned} \quad (21)$$

Meanwhile,

$$\Psi = {}^C R \frac{\partial R({}^O q)}{\partial {}^O q} p \quad (22)$$

2) First-Order Lie Derivatives:

The first-order Lie derivative of h_1 with respect to f_1 is computed as:

$$\mathcal{L}_{f_1}^1 h_1 = \nabla \mathcal{L}^0 h_1 \bullet f_1 = \frac{1}{2} \Psi \Xi({}^O q) e_3 \quad (23)$$

The gradient of $\mathcal{L}_{f_1}^1 h_1$ with respect to x is:

$$\nabla \mathcal{L}_{f_1}^1 h_1 = \begin{bmatrix} \Gamma & -\Upsilon & \Upsilon & 0_{3 \times D} & 0_3 \end{bmatrix} \quad (24)$$

Where Γ and Υ do not need to be explicitly computed because they do not affect the observability analysis.

The first-order Lie derivative of h_3 with respect to f_2 is computed as:

$$\mathcal{L}_{f_2}^1 h_3 = \nabla \mathcal{L}^0 h_3 \bullet f_2 = {}^E R_V^O R^T e_1 \quad (25)$$

The gradient of $\mathcal{L}_{f_2}^1 h_3$ with respect to x is:

$$\nabla \mathcal{L}_{f_2}^1 h_3 = \begin{bmatrix} X & 0_3 & 0_3 & Y & 0_3 \end{bmatrix} \quad (26)$$

Where X is not explicitly computed as Γ and Υ . Y is:

$$Y = \frac{\partial {}^E R_V^O R^T e_1}{\partial {}^E \theta} \quad (27)$$

If ${}^E \theta$ is a 3DoF parameterization:

$$Y = [{}^E R_V^O R^T e_1]_\times \quad (28)$$

If ${}^E \theta$ is a 1DoF parameterization:

$$Y = \begin{bmatrix} -\sin \alpha & -\cos \alpha & 0 \\ \cos \alpha & -\sin \alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} R_y(\beta) R_x(\gamma) {}^O R^T e_1 \quad (29)$$

3) observability analysis:

By stacking the gradients of previously calculated Lie derivatives together, the following observability matrix is constructed:

$$\mathcal{O} = \begin{bmatrix} \nabla \mathcal{L}^0 h_1 & \nabla \mathcal{L}^0 h_2 & \nabla \mathcal{L}^0 h_3 & \nabla \mathcal{L}^1_{f_1} h_1 & \nabla \mathcal{L}^1_{f_2} h_3 \end{bmatrix}^T$$

$$= \begin{bmatrix} \Psi & -\mathcal{O} R_V^O R & \mathcal{O} R_V^O R & 0_{3 \times D} & 0_3 \\ 2_V^O q^T & 0_{1 \times 3} & 0_{1 \times 3} & 0_{1 \times D} & 0_{1 \times 3} \\ 0_{3 \times 4} & \frac{E}{V} R & 0_3 & H_\theta & I_3 \\ \Gamma & -\Upsilon & \Upsilon & 0_{3 \times D} & 0_3 \\ X & 0_3 & 0_3 & Y & 0_3 \end{bmatrix} \quad (30)$$

Adding the third column to the second column, \mathcal{O} becomes:

$$\mathcal{O} = \begin{bmatrix} \Psi & 0_3 & \mathcal{O} R_V^O R & 0_{3 \times D} & 0_3 \\ 2_V^O q^T & 0_{1 \times 3} & 0_{1 \times 3} & 0_{1 \times D} & 0_{1 \times 3} \\ 0_{3 \times 4} & \frac{E}{V} R & 0_3 & H_\theta & I_3 \\ \Gamma & 0_3 & \Upsilon & 0_{3 \times D} & 0_3 \\ X & 0_3 & 0_3 & Y & 0_3 \end{bmatrix} \quad (31)$$

$\frac{E}{V} R$ in the second column can be used to eliminate H_θ in the fourth column and I_3 in the fifth column. Thus, \mathcal{O} reduces to:

$$\mathcal{O} = \begin{bmatrix} \Psi & 0_3 & \mathcal{O} R_V^O R & 0_{3 \times D} & 0_3 \\ 2_V^O q^T & 0_{1 \times 3} & 0_{1 \times 3} & 0_{1 \times D} & 0_{1 \times 3} \\ 0_{3 \times 4} & \frac{E}{V} R & 0_3 & 0_{3 \times D} & 0_3 \\ \Gamma & 0_3 & \Upsilon & 0_{3 \times D} & 0_3 \\ X & 0_3 & 0_3 & Y & 0_3 \end{bmatrix} \quad (32)$$

The fifth column corresponds to the translation part of the extrinsic parameter. This column is not full rank, so the translation part is unobservable. Finally, the rotation part of the extrinsic parameter is analyzed. Let us focus on Y in the fourth column. If $\frac{E}{V} \theta$ is a 3DoF parameterization, Y is a skew symmetric matrix, which means the fourth column is not full rank. Therefore, the rotation part is unobservable. If $\frac{E}{V} \theta$ is a 1DoF parameterization, $\|Y\| \neq 0$ is hold in general case. The fourth column is full rank and cannot be eliminated by other columns. So the rotation part is observable. It is worth noting that although yaw is used as "active DoF" here, the above observability conclusion is also applicable to the pitch or roll case.

V. EXPERIMENTS

To compare the performance of the different algorithms, we adapt our main implementation of GPS-VWO system into three variants. When GPS is not used, the algorithm degenerates to "VWO". When fusing GPS, we further distinguish the algorithm according to whether to calibrate the extrinsic parameter online. If the strategy in [4] is adopted, the algorithm is referred as "GPS-VWO-fixed", which means the extrinsic parameter is estimated during the initialization stage, then fixed and marginalized after initialization. If the rotation part of the extrinsic parameter is 1DoF parameterization and calibrated online after initialization, the algorithm is referred as "GPS-VWO".

We verify the observability conclusion derived for GPS-VWO system with simulated Gaussian GPS measurements.

Then, all three implementations are evaluated on the large-scale KAIST dataset which features challenging urban driving scenes [22]. The KAIST dataset is collected from a rear-wheel drive passenger car. This dataset includes the following sensors: GPS, camera, and wheel encoders distributed on the left and right rear wheel. The frequencies of GPS, camera, and wheel encoder are 5Hz, 10Hz, and 100Hz respectively. We calculate the absolute trajectory error (ATE) between GPS trajectory and groundtruth trajectory for all datasets. To better present the performance of different algorithms under relatively large GPS noise, datasets with ATE larger than 6m are selected for real-world experiments.

For all experiments, the wheel-odometer noise parameters are set as: $\sigma_{n_l} = \sigma_{n_r} = 0.01$, $\sigma_{n_{\omega_x}} = \sigma_{n_{\omega_y}} = 0.01$, $\sigma_{n_{v_y}} = 0.1$, $\sigma_{n_{v_z}} = 0.01$. The noise of v_z is smaller than that of v_y as the wheel movement in the z direction is constrained, tightly attached to the road surface.

A. Validation of the Observability Analysis

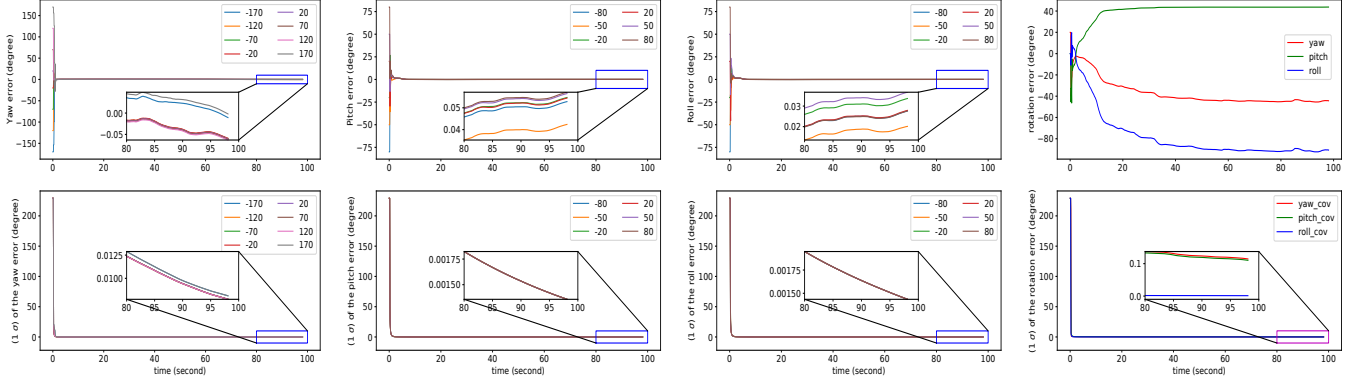
The experiments in this section are here to validate the observability analysis. The KAIST Urban25 dataset is used here. We add zero mean Gaussian noise to the groundtruth position to simulate GPS position measurements. Considering GPS noise in vertical is much larger than that in horizontal, the additive Gaussian noise is defined as:

$$n_{gps} \sim \mathcal{N}(0_{3 \times 1}, \text{diag}(1, 1, 4)) \quad (33)$$

The translation part of the extrinsic parameter between the frame $\{V\}$ and the frame $\{E\}$ is acquired by querying the groundtruth position values at the initialization time. The rotation part of the extrinsic parameter, $\frac{E}{V} R$ can be obtained in the same way. We add some perturbations to the groundtruth $\frac{E}{V} R$ according to different parameterization and set them as the initial estimate of $\frac{E}{V} R$.

First, we test if the 1DoF parameterization of $\frac{E}{V} R$ is observable. The active DoF is yaw in Fig. 2a, pitch in Fig. 2b and roll in Fig. 2c. Fig. 2a shows the convergence of yaw error and the corresponding one standard deviation (1σ) calculated as the square root of the corresponding variance. The range of the initial yaw error is $[-170.0^\circ, 170.0^\circ]$. The estimation of the yaw error consistently converges to near zero with small uncertainty, and the convergence process is robust to the relatively large initial error. Fig. 2b shows the convergence of pitch error and its (1σ) over time. The range of the initial pitch error is $[-80.0^\circ, 80.0^\circ]$. Similar Results of roll parameterization are shown in Fig. 2c. From Fig. 2a to Fig. 2c, We can find that when $\frac{E}{V} R$ is a 1DoF parameterization, the rotational extrinsic parameter is able to perfectly approach to groundtruth value in a short period.

Then, we show that the 3DoF parameterization of $\frac{E}{V} R$ is unobservable. The perturbation angle is set as $[20.0^\circ, 0.0^\circ, 0.0^\circ]$. Fig. 2d shows the rotation error over time. These three components can not converge to zero. The results support the observability conclusion in section IV. Regarding to the convergence of the (1σ) for each component (see the bottom of Fig. 2d), we provide an intuitive explanation in APPENDIX.



(a) Top: Yaw convergence over time with respect to different initial guesses. Bottom: One standard deviation (1σ) of yaw. (b) Top: Pitch convergence over time with respect to different initial guesses. Bottom: One standard deviation (1σ) of pitch. (c) Top: Roll convergence over time with respect to different initial guesses. Bottom: One standard deviation (1σ) of roll. (d) Top: rotation error over time. Bottom: One standard deviation (1σ) of rotation error.

Fig. 2: Calibration results of different parameterization for rotation extrinsic parameter

In order to demonstrate the advantage of GPS-VWO in a intuitive fashion, we draw the distance error over time in Fig. 3. Although the initial yaw error is selected as a relatively large value, the introduction of the GPS significantly reduces the localization error of the VWO. By fusing GPS, the distance error of GPS-VWO is even less than that of the GPS most of the time.

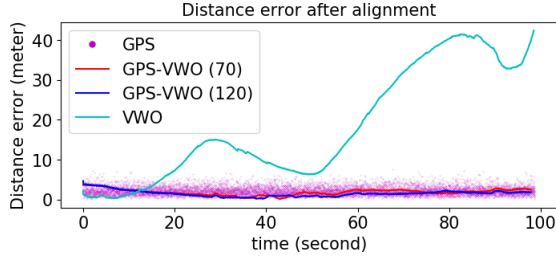


Fig. 3: Comparison of distance error with different methods. The initial yaw errors are 70.0° and 120.0° .

B. Localization Results with real GPS

In this section, real GPS measurements provided by the KAIST dataset [22] are used. We assume that in a ground vehicle, the most important variable necessary to couple the VWO and GPS reference frames is the yaw angle because of the feature that pitch and roll are usually close to zero and prone to be buried by relatively large GPS noise in practice. Moreover, it is effortless to obtain pitch and roll at initialization stage once IMU is utilized, although we do not consider IMU currently.

The initialization of \hat{E}_V^E and E_{p_V} is solved by the SE(3) trajectory alignment of GPS and VWO. Pitch, roll and E_{p_V} are fixed and only yaw is estimated online within the implementation of GPS-VWO. \hat{E}_V^E is the corresponding yaw angle of $\hat{E}_V^E R$. In Fig.4, we plot how $(\hat{E}_V^E \theta - \hat{E}_V^E \theta_0)$ evolves on different datasets, where $\hat{E}_V^E \theta_0$ is the initial value of $\hat{E}_V^E \theta$.

on all datasets show the convergence trend over time, and the deviation from the initial value is less than 3° . The Non-Gaussian behavior of real GPS position measurements is one of the reasons for the unsmooth convergence curves. As an example, the covariance of the GPS position provided by the KAIST Urban39 dataset is shown in the Fig.4. Another reason is that the GPS signal is intermittent, as can be seen in the Fig.6.

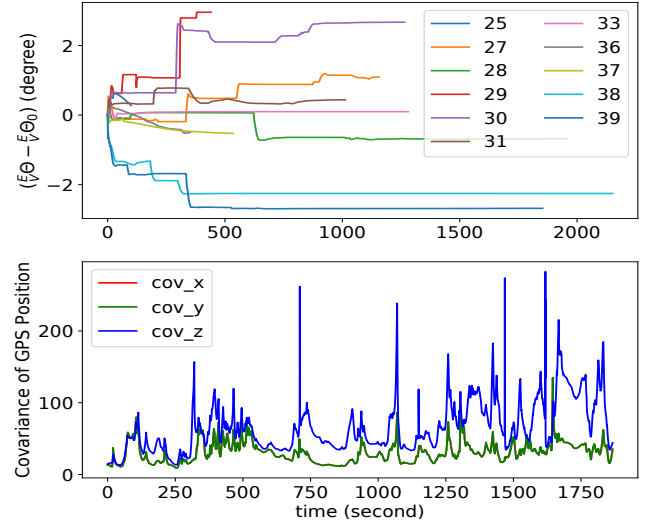


Fig. 4: Top: $(\hat{E}_V^E \theta - \hat{E}_V^E \theta_0)$ convergence over time. Bottom: An example of real world GPS position covariance in three directions. The values in the x and y directions are close.

We evaluate the ATE of GPS, VWO, GPS-VWO-fixed, VIO (Open-VINS [10]), GPS-VIO-fixed and GPS-VWO on all selected datasets. GPS-VIO-fixed is our implementation of [4] since it is not open-sourced. Results are summarized in Tab.I. Half of the sequences cannot be initialized by VIO. This is an expected phenomenon because the data collection

vehicle was driven at approximately constant velocity. VIO initialization process suffers from this specific motion profile [2][3]. The initialization of VWO is much more robust since velocity is directly obtained from the wheels' encoders.

We compare the accuracy of our VWO with Visual-Inertial-Wheel Odometry (VIWO) [12]. [12] only presented the result on one dataset, Urban39. Therefore, this method is not displayed in Tab.I. Their best ATE was 42.748m. Thus we achieve better performance even IMU is not used.

Meanwhile, we notice that the drifts of VWO on Urban36 and Urban37 are relatively larger than others in Tab.I. This is a known issue for wheel-odometer and also reported in the TABLE II of [23]. Although the drifts can be significantly reduced by adjusting the odometer noise parameters, we retain these results for parameter consistency. They can also be seen as challenge cases when fusing with GPS.

It is clear from Tab.I that integrating GPS information can effectively restrict the drift of VWO or VIO, especially for challenge cases (Urban36 and Urban37). Meanwhile, GPS-VWO achieves the best localization accuracy because of the online estimation of the rotational extrinsic parameter, compared with GPS-VWO-fixed and GPS-VIO-fixed.

For a more intuitive presentation, aligned trajectories of the different localization methods on Urban39 dataset are shown in Fig.5. The trajectories of VIO and GPS-VIO-fixed are omitted as they are not the focus of this work, and removing them makes this figure easier to read. In Fig. 6, the distance error after alignment is also presented. The yellow shaded part indicates that there is a GPS signal outage for at least 20s during this period. The longest GPS outage interval exceeds 70s. This shows the robustness of GPS-VWO under intermittent GPS measurements.

TABLE I: ATE(meter) Comparison of Different Methods on the KAIST Dataset

ID	Path len(km)	GPS	A	B	C	D	Ours
25	2.5	7.09	11.35	4.93	×	×	2.09
27	5.4	9.97	22.61	6.84	×	×	6.79
28	11.5	8.66	61.93	7.20	10.78	7.71	3.45
29	3.6	9.63	16.89	8.74	×	×	7.92
30	6.0	10.12	44.55	7.72	×	×	4.57
31	11.4	7.26	49.32	4.58	76.87	6.85	3.78
33	7.6	8.95	31.17	6.40	—	7.77	2.67
36	9.0	20.07	326.22	8.32	×	×	5.88
37	11.8	6.18	156.64	5.13	×	×	3.20
38	11.4	7.09	55.36	4.94	7.53	5.53	3.75
39	11.1	6.43	28.88	5.93	8.73	5.50	4.06

¹ A, B, C, D and Ours represent VWO, GPS-VWO-fixed, VIO [10], GPS-VIO-fixed [4] and GPS-VWO respectively.

² × means VIO initialization is failed. — means trajectory divergence.

VI. CONCLUSION

We propose a novel MSCKF-based GPS-VWO system, which fuses camera, wheel encoder and GPS measurements in a tightly-coupled manner. The observability of the extrinsic parameter between the GPS global coordinate frame and the VWO reference frame is discussed to decide whether

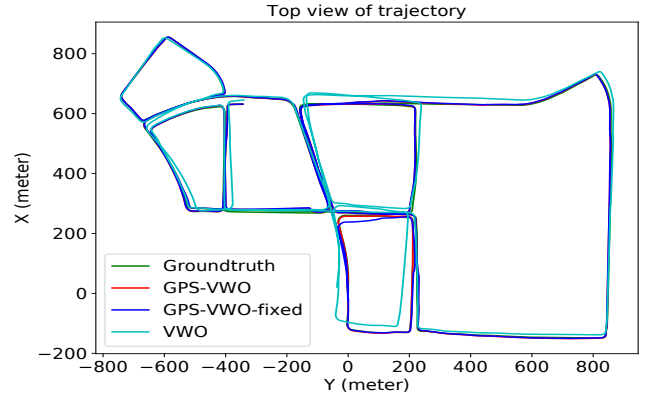


Fig. 5: Comparison of aligned trajectories with different methods.

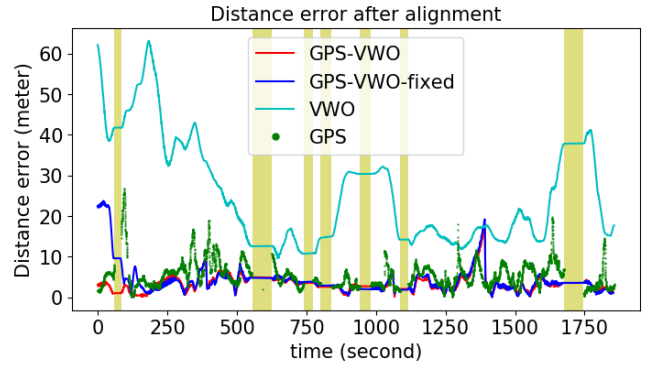


Fig. 6: Distance error over time. The yellow shaded part means that the GPS is unavailable for at least 20s.

to include this parameter in the state vector for online refinement. The observability analysis concludes that the translation part of the extrinsic parameter is unobservable, while the rotation part is observable if it is a 1DoF parameterization. This observation proof is supported by the simulation results. Real-world experimental results on multiple large-scale urban driving datasets demonstrate that the fusion of GPS and VWO provides better accuracy than GPS, and the online calibration of rotational extrinsic parameter further improves the localization accuracy of the estimator.

For future work, we would like to investigate the combination of IMU and implement a GPS-VIWO system.

APPENDIX

The variance of unobservable state converges to zero is a non-trivial phenomenon, therefore we specifically explain here that this phenomenon is theoretically possible to occur.

Here we present a simplified example to explain why the variance of ${}^E_V\theta$ could approach to 0 when ${}^E_V\theta$ is a 3DoF parameterization. Considering the following system model and measurement model:

$${}^E_V\dot{\theta} = 0, {}^E p_G = {}^E_V R^V p_G \quad (34)$$

Assuming ${}^E p_G$ and ${}^V p_G$ are direct measurements with Gaussian noise and the trajectory is a straight line:

$${}^E p_G = \begin{bmatrix} v_x & v_y & 0 \end{bmatrix}^T t \quad (35)$$

Where v_x and v_y represent horizontal velocity. Thus, H_θ becomes (see equation (12)):

$$H_\theta = \begin{bmatrix} 0 & 0 & v_y \\ 0 & 0 & -v_x \\ -v_y & v_x & 0 \end{bmatrix} t \quad (36)$$

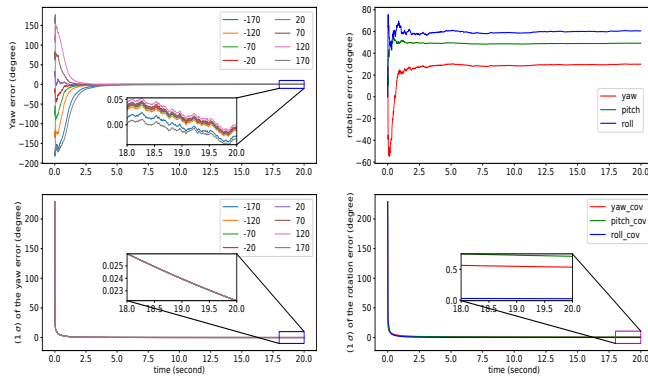
The covariance of ${}^E \theta$, denoted as P , should satisfy Riccati Equation:

$$\begin{aligned} \dot{P} &= -PH_\theta^T H_\theta P \\ H_\theta^T H_\theta &= \begin{bmatrix} v_y^2 & -v_x v_y & 0 \\ -v_x v_y & v_x^2 & 0 \\ 0 & 0 & v_x^2 + v_y^2 \end{bmatrix} t^2 \end{aligned} \quad (37)$$

Measurement noise covariance R is ignored here as it does not affect the analysis. The diagonal elements of $H_\theta^T H_\theta$ are positive, therefore the variance of ${}^E \theta$ could approach to 0.

To validate the above derivation, we conduct simulation experiments as section V-A. v_x and v_y are set as 3 m/s and 4 m/s respectively.

Fig.7 shows the calibration results of rotation extrinsic parameter ${}^E \theta$ when 1DoF parameterization and 3DoF parameterization are adopted for ${}^E \theta$. The state and corresponding variance convergence results are similar as in Fig.2a and Fig.2d. ${}^E \theta$ is observable when it is parameterized with 1DoF, yet unobservable when it is parameterized with 3DoF. Results again support our observability analysis. Moreover, a novel theoretical finding is also demonstrated that the variance of unobservable state could converge to zero for some specific Kalman filter systems (see Fig.2d and Fig.7b).



(a) Top: Yaw convergence over time with respect to different initial guesses. Bottom: One standard deviation (1σ) of yaw. (b) Top: rotation error over time. Bottom: One standard deviation (1σ) of rotation error.

Fig. 7: Calibration results of different parameterization for rotation extrinsic parameter ${}^E \theta$.

REFERENCES

- [1] A. Martinelli, "Vision and imu data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 44–60, 2011.
- [2] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "Vins on wheels," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5155–5162, IEEE, 2017.
- [3] J. Liu, W. Gao, and Z. Hu, "Visual-inertial odometry tightly coupled with wheel encoder adopting robust initialization and online extrinsic calibration," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5391–5397, IEEE, 2019.
- [4] W. Lee, K. Eickenhoff, P. Geneva, and G. Huang, "Intermittent gps-aided vto: Online initialization and calibration," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5724–5731, IEEE, 2020.
- [5] G. Huang, "Visual-inertial navigation: A concise review," in *2019 international conference on robotics and automation (ICRA)*, pp. 9572–9582, IEEE, 2019.
- [6] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [7] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.
- [8] A. I. Mourikis, S. I. Roumeliotis, *et al.*, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *ICRA*, vol. 2, p. 6, 2007.
- [9] M. Li and A. I. Mourikis, "High-precision, consistent ekf-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.
- [10] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4666–4672, IEEE, 2020.
- [11] Y. Yang, P. Geneva, K. Eickenhoff, and G. Huang, "Degenerate motion analysis for aided ins with online spatial and temporal sensor calibration," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2070–2077, 2019.
- [12] W. Lee, K. Eickenhoff, Y. Yang, P. Geneva, and G. Huang, "Visual-inertial-wheel odometry with online calibration," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4559–4566, IEEE, 2020.
- [13] M. Zhang, X. Zuo, Y. Chen, Y. Liu, and M. Li, "Pose estimation for ground robots: On manifold representation, integration, reparameterization, and optimization," *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1081–1099, 2021.
- [14] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," *arXiv preprint arXiv:1901.03642*, 2019.
- [15] G. Cioffi and D. Scaramuzza, "Tightly-coupled fusion of global positional measurements in optimization-based visual-inertial odometry," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5089–5095, IEEE, 2020.
- [16] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on automatic control*, vol. 22, no. 5, pp. 728–740, 1977.
- [17] F. M. Mirzaei and S. I. Roumeliotis, "A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation," *IEEE transactions on robotics*, vol. 24, no. 5, pp. 1143–1156, 2008.
- [18] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.
- [19] N. Trawny and S. I. Roumeliotis, "Indirect kalman filter for 3d attitude estimation," *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep.*, vol. 2, p. 2005, 2005.
- [20] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to autonomous mobile robots*. MIT press, 2011.
- [21] A. Censi, A. Franchi, L. Marchionni, and G. Oriolo, "Simultaneous calibration of odometry and sensor parameters for mobile robots," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 475–492, 2013.
- [22] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 642–657, 2019.
- [23] J. Liu, W. Gao, and Z. Hu, "Bidirectional trajectory computation for odometer-aided visual-inertial slam," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1670–1677, 2021.