

# Factor graph-based ground truth trajectory estimation by fusing robotic total station and inertial measurements

Manuel Mittelstedt<sup>1,\*</sup>, Felix Esser<sup>1</sup>, Gereon Tombrink<sup>1</sup>, Lasse Klingbeil<sup>1</sup> and Heiner Kuhlmann<sup>1</sup>

**Abstract**—The application of mobile mapping systems (MMS) has increased continuously in the last decades in fields like infrastructure or ecosystem monitoring. Equipped with multiple laser scanners and cameras, these systems can generate high-resolution 3D point clouds of the environment in a short time. In this process, the accuracy of the trajectory of the system is of central importance as it directly affects the accuracy of the resulting point cloud. However, since the trajectory estimation depends on sensor observations that are often affected by unknown systematic errors, the actual accuracy of the trajectory remains mainly unknown. To uncover the gap in the trajectory accuracy assessment, we present a method to create ground truth trajectories for mobile mapping systems by integrating millimeter-accurate total station measurements. We mount an Inertial Measurement Unit (IMU) and two 360-degree prisms on a mobile platform, track them with two Robotic Total Stations (RTS) during motion, and fuse these prism measurements with the IMU readings using a factor graph-based trajectory estimation approach. To evaluate the quality of this ground truth trajectory, we record repeated measurements on a closed-loop rail track close to Bonn, Germany. The results show that the generated ground truth trajectory estimated with RTS and IMU data achieves a precision of around 1 mm in position and  $0.05^\circ$  in orientation. To show the potential of the method, we detect systematic deviations of an example MSS that uses Real-Time Kinematic Global Navigation Satellite System (RTK-GNSS) and IMU data for trajectory estimation. The results show that even under good GNSS conditions, the ground truth trajectory from our proposed approach has significantly better precision and less systematic errors than the trajectory based on RTK-GNSS and IMU data.

**Index Terms**—Localization, Sensor Fusion

## I. INTRODUCTION

MOBILE MAPPING SYSTEMS like field robots or drones equipped with multiple cameras or laser scanners can create high-resolution and high-precision 3D point clouds of the environment in a short time [1]. Therefore, the application of such systems has rapidly increased during the past decades in fields like building and ecosystem monitoring or the creation of large-scale city models. To generate 3D point clouds, Simultaneous Localization and Mapping (SLAM) techniques are frequently applied that use sensor data from cameras or laser scanners to estimate both the trajectory

of the system and the environmental map simultaneously [2]. Direct georeferencing techniques are also commonly used, which first estimate the trajectory using for example Global Navigation Satellite System (GNSS) and Inertial Measurements Unit (IMU) data and use this trajectory to register the recorded images and laser scans to a global reference frame. Since both methods depend on sensor data that are affected by random noise and systematic errors, the accuracy of the trajectory is often unknown.

To evaluate the accuracy of the trajectory, direct methods by comparing it with a ground truth trajectory [3]–[5] or indirect methods by using a resulting point cloud to draw conclusions on the trajectory [6], [7] are available. The direct approaches have the advantage that the results are neither environment-dependent nor affected by potential additional systematic deviations of cameras or laser scanners, such as calibration errors. Therefore, many researchers focus on the generation of ground truth trajectories to evaluate the trajectories of mobile mapping systems. Existing methods are usually limited either by their suitability only for small-scale areas or by their accuracy. Motion-capturing systems with cameras enable a high-accuracy ground truth trajectory, but can only be used in laboratories and small-scale areas [4]. An alternative for large-scale outdoor areas is to use Real-time kinematic GNSS in combination with a high-grade IMU [8]. However, this is, in some cases, the solution we want to evaluate, and the accuracy is still limited by the available accuracy of the GNSS positions in the particular test environment. Robotic Total Stations (RTS) provide position data with higher accuracy and, in particular, less systematic errors than RTK-GNSS. By mounting a 360° prism on the platform and tracking it continuously during motion with the RTS, a position measurement with an accuracy in the order of millimeters can be achieved [3], [9]–[11]. A simultaneous use of three RTSs can be used to estimate both ground truth position and orientation, as demonstrated by [5]. However, this method relies heavily on prior knowledge of the prism positions on the platform and does not provide redundant observations of the orientation, which may compromise its robustness regarding deviations and outliers.

In this paper we present a factor graph-based method to fuse the position measurements of two RTSs with the data of an IMU to estimate a ground truth trajectory. While the RTS measurements deliver millimeter-precise position measurements in motion, the IMU obtains relative changes along the trajectory with high frequency. The integration of IMU measurements over a longer period leads to drift effects and systematic deviations, which cannot be corrected for the heading angle without additional measurements or restrictions. To avoid these drift effects in the heading angle, we include the position

Manuscript received: March 31, 2025; Revised: May 19, 2025; Accepted: July 15, 2025

This paper was recommended for publication by Editor Javier Civera upon evaluation of the Associate Editor and Reviewers' comments.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2070–390732324 - PhenoRob.

<sup>1</sup>University of Bonn, Institute of Geodesy and Geoinformation (Nussallee 17, 53115 Bonn, Germany)

\*Corresponding author, mittelstedt@igg.uni-bonn.de

Digital Object Identifier (DOI): see top of this page.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.**

measurements of both RTSs using a novel RTS baseline factor in our factor graph for trajectory estimation. Since roll and pitch values are obtained with sufficient accuracy from the IMU measurements, two RTSs are sufficient for the method presented. Compared to [5], this reduces the effort and increases flexibility, as only two instead of three RTSs are required and therefore an additional line of sight constraint is eliminated. An additional advantage of our method is that the IMU measurements are available at a very high frequency, which means that even fine movements can be taken into account. We evaluate our approach using repeated closed-loop rail track measurements to analyze the precision of the ground truth trajectory and the noise model of the RTS-related factors. To show the potential of the ground truth, we additionally uncover systematic trajectory deviations of a state-of-the-art system that uses centimeter accurate RTK-GNSS and IMU data.

In summary, the main contribution of this paper is a method that estimates the ground truth trajectory by fusing two RTS prism positions with IMU data using a factor graph-based approach. The resulting ground truth trajectory is characterized by millimeter precision without the presence of systematic deviations. Since the RTS data is geo-referenced with high precision, the ground truth allows the evaluation of state-of-the-art systems that use GNSS and IMU data for trajectory estimation.

The paper is structured as follows: First, in Section II, the methodology to estimate the ground truth trajectory using a factor graph-based approach is presented. The system and the sensor setup we used in our work for experiments are described in Section III. In Section IV, the results of an evaluation concerning the estimated ground truth trajectory are presented. In addition, we evaluate a GNSS/IMU trajectory and demonstrate how the ground truth trajectory enables the detection of systematic deviations in Section IV-C.

## II. TRAJECTORY ESTIMATION METHODOLOGY

The aim is to estimate a high-accurate ground truth trajectory by fusing total station and IMU measurements. The total station measurements provide an absolute position observation. Relative changes in position and orientation are derived from the IMU measurements. Since the IMU observation of horizontal orientation (yaw angle) is subject to drift effects, absolute information of the yaw angle is integrated using total station measurements as a baseline.

The trajectory estimation considering different sensors like IMU and absolute measuring sensors like total station is an essential part of any MMS measurement processing. One common approach to calculate the trajectory using several sensor observations is a Kalman filter [10], [12], [13]. In addition, new approaches such as factor graph-based solutions provide the ability to optimize a trajectory efficiently and globally based on sensor observations [14], [15]. The advantage of factor graph-based approaches is that they can be flexibly extended by new sensor observations and directly estimate the trajectory without forward and backward filtering like common Kalman filter smoothers. For the ground truth trajectory estimation, we use a factor graph-based approach.

The trajectory is defined by a sequence of navigation states  $\mathbf{x}_i$  consisting of position  $\mathbf{p}_i = [p_x, p_y, p_z]_i$ , velocity  $\mathbf{v}_i = [v_x, v_y, v_z]_i$ , and orientation  $\mathbf{R}_i$  at each time increment  $i$ . The orientation is represented by the rotation matrix  $\mathbf{R}_i$ , which can be defined by the three Euler angles roll  $\alpha_i$ , pitch  $\beta_i$  and yaw  $\gamma_i$ .

$$\mathbf{x}_i = [\mathbf{p}_i, \mathbf{v}_i, \mathbf{R}_i]^T. \quad (1)$$

The total set of navigation states over the whole trajectory to be estimated is given by the set of  $I$  navigation states.

$$\mathcal{X}_I = \{\mathbf{x}_i \mid i = 1, \dots, I\} \quad (2)$$

Besides the trajectory navigation states, time-dependent IMU bias parameters  $\mathbf{b}_j$  containing acceleration bias  $\mathbf{b}_j^a = [b_x^a, b_y^a, b_z^a]_j$  and gyroscope bias  $\mathbf{b}_j^\omega = [b_x^\omega, b_y^\omega, b_z^\omega]_j$  need to be estimated as well. All bias parameters that need to be estimated over the entire measurement time can be summarized as  $\mathcal{B}_J$

$$\mathcal{B}_J = \{\mathbf{b}_j \mid j = 1, \dots, J\} \quad (3)$$

In total, the parameters to be estimated are

$$\mathcal{V}_K = \{\mathcal{X}_I, \mathcal{B}_J\} \quad (4)$$

The navigation states of the ground truth trajectory are estimated using a factor graph. In principle, graphs consist of nodes and edges that represent relations between the nodes. A special type of graph is the so-called factor graph. A factor graph is a two-sided graph with two types of nodes, variable nodes on the one hand and factor nodes on the other. Variable nodes represent unknown random variables and factor nodes represent impact factors on the random variables obtained from measurements or previous knowledge.

The trajectory estimation using multiple sensor measurements can also be represented as a factor graph. The parameters  $\mathcal{V}_K$  to be estimated are introduced as variable nodes. The measurements of the sensors are integrated as factor nodes. A factor node can also be understood as a function between a measurement and variable nodes. By linearizing the factor graph, optimization based on least squares can be applied. A visualization example of a factor graph is shown in Figure 1.

The navigation states and IMU bias parameters are represented as variables and the measurements as factor nodes. New bias parameters are estimated when an absolute measurement is included, as they are observable through absolute comparison. The squared factor nodes of Figure 1 highlight the non-linear error functions depending on unknown parameters and sensor measurements.

Equation 5 describes the corresponding least squares optimization problem of the factor graph as shown in Figure 1. Each function  $f()$  represents the corresponding error function of the respective factor. The error functions define the functional relationship between the parameters to be estimated and the observations. To perform a least squares adjustment, a covariance matrix  $\Sigma$  is required for each measurement included as a factor.

The IMU acceleration and gyroscope measurements can be used through a pre-integration to calculate the relative movements between two navigation states  $i$  and  $j$ . This is

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

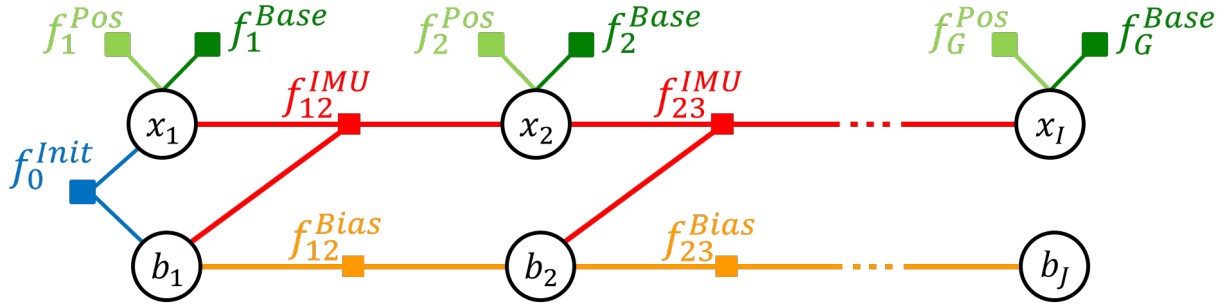


Fig. 1. Example for the visualization of a factor graph, consisting of variable nodes and factor nodes (initial-, IMU-, IMU-bias-, position and baseline-factors).

$$\mathcal{V}_K^* = \arg \min_{\mathcal{V}_K} \left\{ \sum_{i=0, j=i+1}^K \|f_{ij}^{IMU}\|_{\Sigma_{ij}}^2 + \sum_{q=0}^Q \|f_q^{Bias}\|_{\Sigma_{(q-1)q}}^2 + \sum_{g=1}^G \|f_g^{Pos}\|_{\Sigma_{pp}}^2 + \sum_{g=1}^G \|f_g^{Base}\|_{\Sigma_{ll}}^2 \right\}. \quad (5)$$

represented in the factor graph as an IMU factor  $f_{ij}^{IMU}$  to predict the next pose with accuracy given by the covariance matrix  $\Sigma_{ij}$ . For pre-integration, the approach of [16], [17] and [18] is used. Since the IMU observations are influenced by a time-dependent bias, the current bias  $b_q$  has to be taken into account in pre-integration. The bias factor  $f_q^{Bias}$  contains the prediction, how the bias changes over the measurement period, described by a random walk process.

The absolute observations of the total stations are integrated using two kinds of factors. A position measurement is integrated via the  $f_g^{Pos}$  factor. The error function (Eq. 6) is calculated from the position measurement  $\mathbf{l}_i^{Pos}$ , the lever arm  $\mathbf{t}$ , i.e. the referred point of the position measurement on the origin of the platform coordinate system and the position  $\mathbf{p}_i$  and orientation  $\mathbf{R}_i$  of the navigation state  $\mathbf{x}_i$  at time  $i$ .

$$f^{Pos}(\mathbf{p}_i, \mathbf{R}_i, \mathbf{l}_i^{Pos}) = \mathbf{l}_i^{Pos} - (\mathbf{p}_i + \mathbf{R}_i \cdot \mathbf{t}) \quad (6)$$

The matrix  $\mathbf{R}$  describes the rotation matrix that specifies the orientation of the platform at navigation state  $\mathbf{x}_i$ . The covariance matrix  $\Sigma_{pp}$  gives the accuracy of the position measurement.

Since the integration of a position factor does not provide any direct information about the orientation, a baseline vector is derived by the difference of the two position measurements. The orientation of the baseline vector in space can be traced back to the orientation of the MMS if the extrinsic calibration is known. The extrinsic calibration i.e. the lever arms  $\mathbf{t}_1, \mathbf{t}_2$  describe the two reference points, referring to the origin of the platform, of the position observations. A baseline vector is integrated into the factor graph as the factor  $f_g^{Base}$ . The error function 7 specifies the functional relationship between the observation  $\mathbf{l}_i^{Base}$  and the rotation  $\mathbf{R}_i$  at time  $i$ . So the error function is the difference between the predicted baseline vector and the measured one.

$$f^{Base}(\mathbf{R}_i, \mathbf{l}_i^{Base}) = \mathbf{l}_i^{Base} - \mathbf{R}_i \cdot (\mathbf{t}_2 - \mathbf{t}_1) \quad (7)$$

The accuracy of the baseline measurement  $\mathbf{l}_i^{Base}$  is specified by the covariance matrix  $\Sigma_{ll}$ . The covariance matrix of the baseline measurement depends on the accuracy of the two position measurements used to compute the baseline.

For the optimization of  $\mathcal{V}_K$  we use the iSAM2 algorithm [19] that sequentially builds up and optimizes the factor graph. To estimate the global optimum, approximate values for the parameters  $\mathcal{V}_K$  are required for linearization. Since these are often only insufficiently available over the entire measurement time, the iSAM2 algorithm offers the possibility to optimize sequentially and update the approximate values over time. If parameters in the past are no longer influenced by new measurements, the iSAM2 algorithm no longer optimizes these parameters. This ensures that the factor graph is optimized efficiently. Further information can be found in [19].

### III. EXPERIMENTS

The first part of this Section introduces the system and the sensors used. The second part describes the measurements conducted as part of our work and the measurement setup.

#### A. Sensor System

For ground truth trajectory estimation, we use a mobile sensor system consisting of two 360-degree Leica GRZ122 prisms and an IMU, as shown in Figure 2. The IMU is an Ellipse-2D from SBG-Systems located in the aluminum box.

The prisms, as targets for the total stations, are needed to create the ground truth trajectory. The prisms must be mounted at different heights to ensure a permanent line of sight between the prism and the total station.

Since the lever arms between the IMU center and the prisms  $\mathbf{t}_1, \mathbf{t}_2$  must be known for the error function of the position factor of the factor graph, these vectors must be estimated with high precision. We determine the lever arm as part of a calibration in the laboratory. According to Equation 7, the baseline factor is particularly determined by the difference vector between the two lever arms. We determine this vector in the lab by using two Leica round prisms and multiple measurements of a total station, which allow sub-millimeter accurate position measurements. The distance between the prisms, also known as the baseline length, is 0.92 meters.

Total stations are high-precision measuring sensors that calculate the position coordinate of a targeted point by measuring

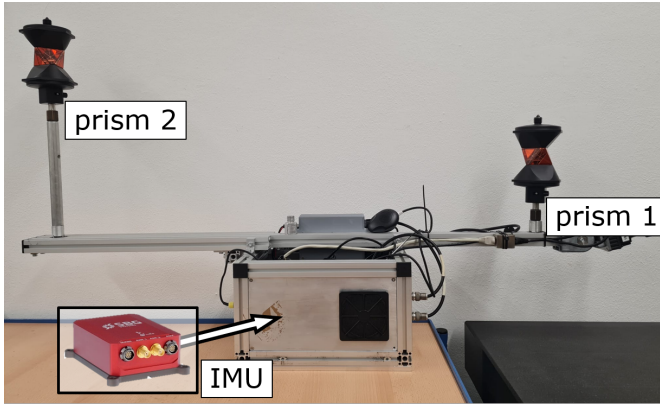


Fig. 2. System for ground truth trajectory estimation, consisting of an IMU and two Leica GRZ122 prisms

two angles and a distance. Robotic total stations additionally provide the feature of automatic prism tracking. As a part of the sensors used for the estimation of the ground truth trajectory, two RTSs are set up externally and statically in the measurement environment. The two RTSs must be set up in a way that ensures a permanent line of sight to two prisms on the platform. If the ground truth trajectory requires to be geo-referenced, for example to enable evaluation of systems that use GNSS for direct geo-referencing, the RTS observations must be available in a geo-referenced coordinate system. This can be realized, for example, by using existing geo-referenced control points in the measuring environment.

Both RTSs, Leica TS60 and Leica MS60 were synchronized by connecting each to a GPS-time synchronized logging device (Raspberry Pi Single Board Computer). This logging device timestamps incoming GeoCOM messages containing the measurements of the RTS in GPS time and writes them to a file. Due to the very similar architecture of both the RTS and the logging devices used, low synchronization latencies can be achieved. Empirical investigations showed a time offset of  $-0.2$  ms which does not differ significantly from zero given the observation accuracy. Note that the RTS measurements are not triggered or performed at the same point in time. By synchronization, we mean that the time systems of both RTSs run synchronously. Measured values for both RTSs for a specific point in time can then be obtained via interpolation.

### B. Measurement Setup

The measurements are realized on the test site of the University of Bonn at the Klein-Altendorf campus. An approximately 140-meter-long closed loop rail track, as highlighted in Figure 3 c, enables repeated measurements of the ground truth trajectory. The rail track contains variations of all trajectory parameters, thus also for the orientation in roll, pitch, and yaw. Conducting measurements on the rail track allows investigating the repeatability of the ground truth trajectory [3].

The sensor system we used for the measurements is shown in Figure 3 b. In comparison to Figure 2, two Leica AS10 GNSS antennas are additionally installed on the platform.

They are used to compare the method presented using observations of two RTSs and an IMU to a conventional system using GNSS and IMU. To move the entire sensor system along the rail track, it is mounted on a test trolley, attached to the rails with three wheels. We moved the trolley manually along the rail track. The velocity varies between 0.2 and 0.8 meters per second, after a zero velocity period at the beginning of the measurement to initialize the IMU.

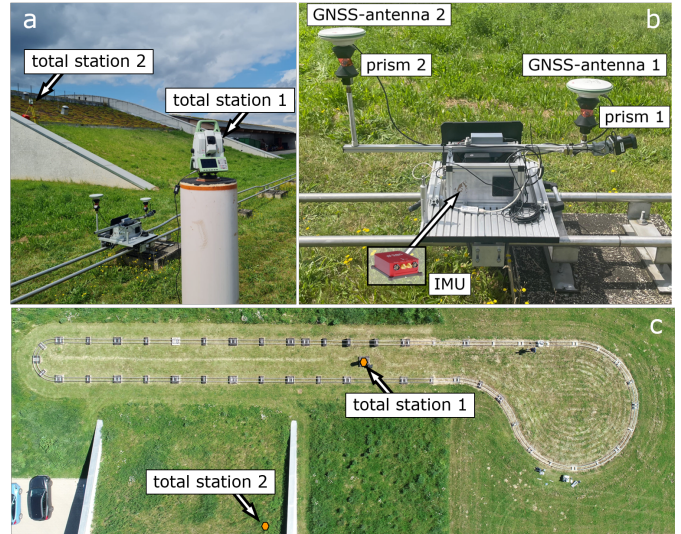


Fig. 3. a: Measurement set up with two total stations; b: Platform with prisms, IMU, and GNSS antennas; c: Rail track

To allow prism tracking along the whole rail track, we place the first total station on a pillar in the middle of the test track, see Figure 3 a. The second total station is positioned on a hill 20 meters from the rail track. A fixed geo-referenced control point network consisting of several pillars is located in the area around the rail track. A network adjustment carried out in advance with static GNSS and total station measurements provides the control points with an accuracy in the sub-millimeter range. We use these network points to transform the local total station position measurements into the global WGS84 reference frame. This ensures the same spatial reference in which for example GNSS measurements were obtained, so trajectories of GNSS/IMU systems can be compared with the ground truth trajectory.

A dataset consisting of 11 laps was measured simultaneously for all sensors. We used the measurements on the rail track to investigate the repeatability based on multiple measurements. Simultaneous measurements of GNSS, RTS and IMU allow a direct comparison between the ground truth and a trajectory estimated using GNSS and IMU. Since RTK corrected GNSS measurements are used, centimeter-level accuracy can be assumed for the GNSS positions.

## IV. EVALUATION AND RESULTS

In this Section, we present the results and findings of the measurements described in Section III. First, in Section IV-A, we describe how the noise model for the factor graph is selected and how this can be confirmed based on the

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.**

measurements. The repeatability of the ground truth trajectory based on multiple measurements of the laps on the rail track is analyzed and compared with the trajectory from GNSS and IMU in Section IV-B. Finally, in Section IV-C, using the example of the GNSS/IMU trajectory, we demonstrate how the high-accurate ground truth trajectory can be used to detect systematic deviations.

In general, accuracy, as defined by ISO 5725-1, is a combination of random and systematic deviations. Random deviations are termed as precision and can be covered by repeated measurements. Systematic deviations are termed trueness and can be determined using a ground truth value.

#### A. Factor graph noise model

To integrate the observations correctly into a factor graph, a specification of the noise model is required. First, prior knowledge, such as manufacturer specifications of sensor accuracy, can be used to define the noise model. Once measurements have been conducted, we use the results to validate the noise model.

The 3D accuracy of the position observation  $I^{Pos}$  measured with a total station depends on several influencing factors and the measurement setup. Besides the measurement noise specified by the manufacturer, potential total station position and system-related deviations have to be considered. The total station position deviations can be calculated based on the network adjustment carried out previously and are in the sub-millimeter range. Two main system-related deviations must be taken into account. One of these is small vibrations of the prisms on the platform, which can amount to around one millimeter. The second are deviations of the prisms depending on the incident angle of the target. A closer examination of this deviation can be found in [20]. The resulting deviation of the measured prism position is up to  $\pm 1.5$  mm. Overall, a noise model of 3 millimeters is assumed for 3D-position accuracy. The noise model of the baseline observation  $I^{Base}$  is derived based on the accuracy of the total station position measurements as well.

To validate the noise model, the raw total station measurements are examined empirically. One way to do so is to check the baseline length between the prisms over the measurement period. According to [21], the baseline length is also known as the inter-prism distance. Since the measurements were carried out on a closed loop rail track, repeated measurements are available over several laps. As constant systematic deviations have the same effect in each lap, the laps are analyzed concerning similarities in the variation of the baseline length. Figure 4 visualizes the baseline lengths as an example for 5 laps of the measured dataset. Besides the measurement noise, in each lap the same deviations of around  $\pm 1.5$  millimeters compared to the baseline length determined in the laboratory can be recognized. According to [20], this corresponds to the order of magnitude of the deviation of the prisms depending on the incident angle of the target. Given that this provides an explanation for the existing systematics, which are smaller than the assumed noise, it can be approved that there are no other significant systematic deviations. As a result, the selected

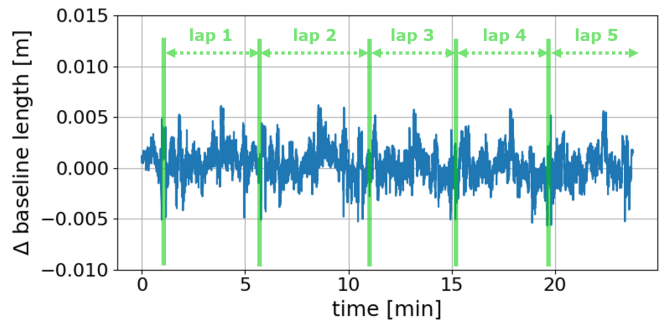


Fig. 4. Difference between the baseline calibrated in the laboratory and estimated baseline length out of measurements along five laps on the rail track. The five laps are divided by the green lines.

noise model fits the acquired measurement data and we can assume that all deviations have been taken into account.

#### B. Trajectory repeatability

Since our dataset contains repeated measurements on a closed-loop rail track, the repeatability of trajectories is computed, which we use to measure the trajectory precision. This is done by estimating a mean trajectory based on the repeated laps and computing the errors to this mean afterward. We first do so for the ground truth trajectory and, to highlight the precision gain of the ground truth, then additionally for the trajectory estimated from GNSS and IMU data. The GNSS/IMU trajectory is estimated using a factor graph-based approach as described in our previous work [22].

To compute the mean trajectories, all poses of the trajectory are first arranged based on the arc length of the rail track. The mean trajectory is determined as a continuous function using a least-squares adjustment. The errors  $\delta$  to the mean trajectory are then calculated over the whole trajectory for each position and orientation parameter. Therefore, we use the popular trajectory evaluation software Trajectory<sup>1</sup> as proposed in [3].

Since the trajectory was sorted spatially along the rail track before calculating the errors, analyzing the along-track errors is senseless. The analysis of the cross-track error  $\delta_c$  provides information about the repeatability of the horizontal component and the vertical error  $\delta_v$  about the height component of the trajectory. To compute the mean values for the orientation parameters, we use the chordal L2-mean, as described in [23]. Since the orientation parameter roll pitch and yaw errors cannot be calculated directly in the Euclidean space, the orientation parameters are converted to the quaternion space. After calculating the error in quaternion space, the error is converted back to roll pitch and yaw errors  $\delta_\alpha$ ,  $\delta_\beta$  and  $\delta_\gamma$ .

To obtain a single scalar that values the repeatability for each position and rotation parameter separately, the standard deviation  $\sigma$  is computed over all  $N$  trajectory poses with

$$\sigma = \frac{1}{N} \sqrt{\delta^T \delta} \quad (8)$$

<sup>1</sup><https://github.com/gereon-t/trajectory>

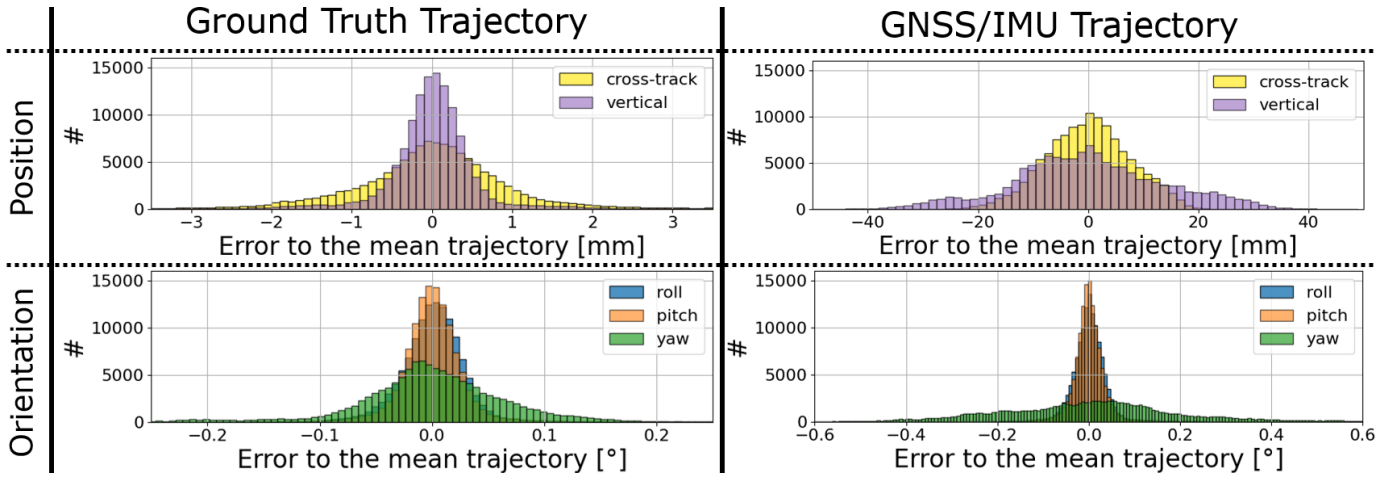


Fig. 5. Distribution of the trajectory errors for the ground truth trajectory and trajectory estimated with GNSS and IMU data divided into cross-track, cross-vertical, roll, pitch, and yaw. Due to visualization reasons, the interval width of the histograms and the scaling of the x-axis differ.

Both the error distributions and their standard deviations are used to evaluate the ground truth trajectory and trajectory estimated with GNSS and IMU data.

Figure 5 shows the resulting error histograms for cross-track, vertical, roll, pitch, and yaw. The cross-track and vertical position errors of the ground truth trajectory are distributed between  $\pm 2.5$  mm. The roll and pitch errors extend between  $\pm 0.1^\circ$ , and the yaw angle errors between  $\pm 0.2^\circ$ . The errors of each component follow a normal distribution.

The corresponding standard deviations of the ground truth for the error histograms are summarized in Table I. The millimeter range standard deviation of 1.4 mm and 0.8 mm for cross-track and vertical positions show the high repeatability of the prism measurements in the trajectory estimation. The standard deviation of the rotations is estimated with  $0.03^\circ$  for roll and pitch and  $0.07^\circ$  for the yaw rotation.

The error distributions of the GNSS/IMU trajectories differ in the magnitude of the errors. The cross-track and vertical position errors are distributed between  $\pm 40$  mm and thus differ by an order of magnitude from the errors of the ground truth trajectory. While the errors of the roll and pitch orientation components are in the same order of magnitude as for the ground truth trajectory, the errors for the yaw component are significantly larger and are distributed between  $\pm 0.5^\circ$ . The errors of the GNSS/IMU trajectory cannot generally be described by a normal distribution for each component. Thus, for example, several local maxima are recognizable in the histogram of the vertical errors.

The empirical standard deviation for each parameter of the GNSS/IMU trajectory is higher than the standard deviation of the ground truth trajectory. The position's standard deviation of 7.6 mm and 13.8 mm is significantly higher than the estimated ground truth trajectory. The empirical standard deviation of the yaw orientation is  $0.19^\circ$ , which is more than twice as high as the standard deviation of the ground truth yaw orientation.

The results demonstrate that the assumptions made about the noise model correspond to the real measurements and that a highly accurate ground truth trajectory with high repeatability

TABLE I  
STANDARD DEVIATION OF THE TRAJECTORY ERROR DIVIDED INTO HORIZONTAL CROSS-TRACK, VERTICAL, ROLL, PITCH, AND YAW FOR THE GROUND TRUTH TRAJECTORY AND GNSS/IMU TRAJECTORY.

Parameter	RTS and IMU	GNSS/IMU
cross-track [mm]	1.4	7.6
vertical [mm]	0.8	13.8
roll [ $^\circ$ ]	0.03	0.04
pitch [ $^\circ$ ]	0.03	0.04
yaw [ $^\circ$ ]	0.07	0.19

is estimated. Even though the GNSS conditions in the measurement environment are good, the ground truth trajectory is proven to outperform GNSS/IMU trajectories in terms of precision.

### C. Uncover systematic deviations using ground truth

As described in Sections IV-A and IV-B, the results prove that the ground truth trajectory is not affected by any significant systematic deviations and has high precision. In the following Section, we show how the millimeter-precise ground truth trajectory can be used to detect systematic deviations of the GNSS/IMU trajectories and evaluate the trueness. Note that the combination of GNSS and IMU data is just an example, but any kind of trajectory estimate, such as SLAM systems, could be evaluated here as well.

To determine systematic deviations of the GNSS/IMU trajectory, it can be compared directly with the ground truth trajectory. Since both trajectories are in the same coordinate system and were measured simultaneously, the matching of poses to be compared is realized by time in this paper.

The time-varying systematic deviations of the GNSS/IMU trajectory are investigated by analyzing the errors to the ground truth trajectory. Figure 6 shows the difference between the height components of the GNSS/IMU solution and the ground truth trajectory. This time series clearly shows a temporal variation of up to 4 cm over the measurement period of approximately 40 minutes. It can be assumed that the system-

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

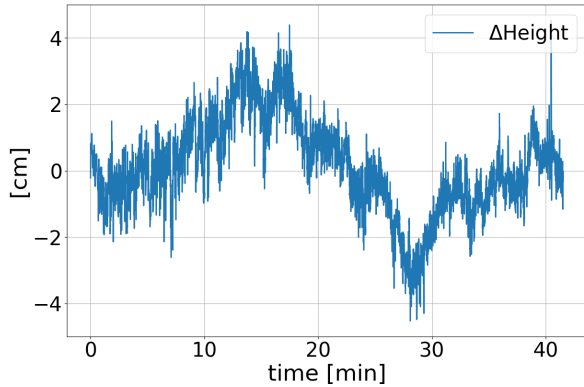


Fig. 6. Height differences over time between GNSS/IMU trajectory and the ground truth trajectory

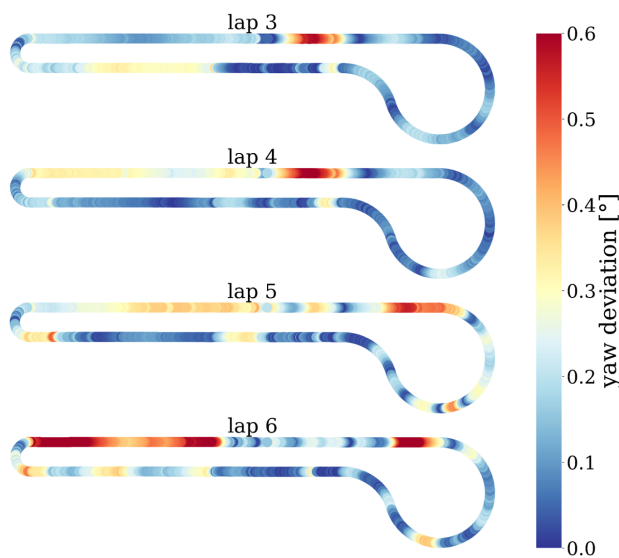


Fig. 7. Yaw orientation deviations between ground truth and GNSS/IMU trajectory plot over the rail track for the example of four sequential laps (The plot is generated with Trajectory [3])

atic deviations are caused by varying systematic deviations of GNSS over time.

The time-varying systematic deviations of the GNSS/IMU trajectory can also be identified by the residuals of the orientation parameters to the ground truth trajectory. Figure 7 shows the residuals of the yaw orientation for four sequential laps. First, it can be noted that there is no section of the rail track where the same systematic occurs in every lap. This proves that the deviations change over the measurement time. Particularly in laps 3 and 6, the sections where the largest deviations occur are different. However, a continuous evolution of the deviations can be observed over the four sequential laps. As an example, the deviation in the upper right part of the rail track increases, while the deviations in the upper middle part decrease from lap to lap. This demonstrates the fact that the systematic deviations change over time and indicate deviations resulting from the GNSS measurements.

The experiments conducted prove that the ground truth

trajectory estimated from RTS and IMU is highly accurate and can be used for evaluation purposes. Systematic deviations of a GNSS/IMU system were detected. Since the rail track is not required for this analysis of systematic deviations, the experiment is also transferable to other environments. As described in Section III-A, a prerequisite for analyzing a test trajectory using GNSS observations is that control points are available in a global coordinate system to geo-reference both RTSs. This enables a more comprehensive analysis of the impact of GNSS on trajectory estimation, for example, in challenging environments such as urban canyons or forests.

## V. CONCLUSION

In this paper, we present a method for determining a ground truth trajectory using two total stations and an IMU with a factor graph-based approach.

Conducting tests on a rail track, we have proven that the generated trajectory achieves precision in the millimeter and hundredth of a degree range. The accuracy of the ground truth trajectory exceeds the accuracy of a state-of-the-art configuration of GNSS/IMU system, even under very good GNSS conditions. The potential of the estimated ground truth trajectory for evaluation purposes is highlighted using the example of a GNSS/IMU trajectory.

The ground truth trajectory offers the opportunity to advance research. For example, the accuracy of the trajectory of MMSs can be investigated by comparing it with the ground truth trajectory. Therefore, a conclusion about precision and systematic deviations can be made. Algorithms for trajectory estimation can also be evaluated with the ground truth trajectory. Another example is calibration measurements for sensors like cameras and laser scanners, which can be improved by a highly accurate trajectory.

Future research may investigate testing the ground truth trajectory for different applications. To improve the ground truth trajectory, future research should attempt to overcome the prism deviations. For example, automatic rotating round prisms would be conceivable.

## REFERENCES

- [1] Erik Heinz, Christian Eling, Markus Wieland, Lasse Klingbeil, and Heiner Kuhlmann, "Development, calibration and evaluation of a portable and direct georeferenced laser scanning system for kinematic 3d mapping," *Journal of Applied Geodesy*, vol. 2015, Bd. 9., no. 4, pp. 227–243, Jan. 2016.
- [2] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5135–5142.
- [3] G. Tombrink, A. Dreier, L. Klingbeil, and H. Kuhlmann, "Trajectory evaluation using repeated rail-bound measurements," *Journal of Applied Geodesy*, vol. 17, 02 2023.
- [4] S. Dudzik, "Application of the motion capture system to estimate the accuracy of a wheeled mobile robot localization," *Energies*, vol. 13, no. 23, 2020.
- [5] M. Vaidis, M. Shahraji, E. Daum, W. Dubois, P. Giguère, and F. Pomerleau, "Rts-gt: Robotic total stations ground truthing dataset," 09 2023.
- [6] M. Wagner, B. Jost, L. Klingbeil, and H. Kuhlmann, "Evaluation of mobile mapping point clouds in the context of height difference estimation," *Journal of Applied Geodesy*, 03 2025.
- [7] L. Wiesmann, E. Marks, S. Gupta, T. Guadagnino, J. Behley, and C. Stachniss, "Efficient lidar bundle adjustment for multi-scan alignment utilizing continuous-time trajectories," 2024.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.**

- [8] C. Eling, L. Klingbeil, and H. Kuhlmann, "Real-time single-frequency gps/mems-imu attitude determination of lightweight uavs," *Sensors*, vol. 15, no. 10, pp. 26 212–26 235, 2015.
- [9] T. Thalmann and H. Neuner, "Temporal calibration and synchronization of robotic total stations for kinematic multi-sensor-systems," *Journal of Applied Geodesy*, vol. 15, no. 1, pp. 13–30, 2021.
- [10] F. Keller and H. Sternberg, "Multi-sensor platform for indoor mobile mapping: System calibration and using a total station for indoor applications," *Remote Sensing*, vol. 5, no. 11, pp. 5805–5824, 2013.
- [11] S. Vogel and F. Hake, "Development of gps time-based reference trajectories for quality assessment of multi-sensor systems," *Journal of Applied Geodesy*, vol. 18, pp. 597–612, 05 2024.
- [12] F. Caron, E. Duflos, D. Pomorski, and P. Vanheeghe, "Gps/imu data fusion using multisensor kalman filtering: introduction of contextual aspects," *Information fusion*, vol. 7, no. 2, pp. 221–230, 2006.
- [13] B. Schaffrin and H. Iz, "Towards total kalman filtering for mobile mapping," in *5th Int. Symp. on Mobile Mapping Technol*, 2008.
- [14] W. Wen and L.-T. Hsu, "Towards robust gnss positioning and real-time kinematic using factor graph optimization," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5884–5890.
- [15] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," *Robotics and Autonomous Systems*, vol. 61, pp. 721–738, 08 2013.
- [16] T. Lupton and S. Sukkariéh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 61–76, 2012.
- [17] L. Carlone, Z. Kira, C. Beall, V. Indelman, and F. Dellaert, "Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors," *Proceedings - IEEE International Conference on Robotics and Automation*, 06 2014.
- [18] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "Imu preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation," *Robotics: Science and Systems (RSS) conference*, 01 2015.
- [19] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "isam2: Incremental smoothing and mapping using the bayes tree," *International Journal of Robotic Research - IJRR*, vol. 31, pp. 216–235, 05 2012.
- [20] S. Lackner and W. Lienhart, "Impact of prism type and prism orientation on the accuracy of automated total station measurements," in *Proc. Joint International Symposium on Deformation Monitoring (JISDM)*, Mar. 2016, p. 8p, joint International Symposium on Deformation Monitoring (JISDM) ; Conference date: 30-03-2016 Through 01-04-2016.
- [21] M. Vaidis, W. Dubois, E. Daum, D. LaRocque, and F. Pomerleau, "Uncertainty analysis for accurate ground truth trajectories with robotic total stations," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 5312–5319.
- [22] F. Esser, R. A. Rosu, A. Cornelißen, L. Klingbeil, H. Kuhlmann, and S. Behnke, "Field robot for high-throughput and high-resolution 3d plant phenotyping: towards efficient and sustainable crop production," *IEEE Robotics & Automation Magazine*, vol. 30, no. 4, pp. 20–29, 2023.
- [23] R. Hartley, J. Trunpf, Y. Dai, and H. li, "Rotation averaging," *International Journal of Computer Vision*, vol. 103, 07 2013.