UWB-VIO Fusion for Accurate and Robust Relative Localization of Round Robotic Teams

S. Zheng, Z. Li, Y. Liu, H. Zhang, P. Zheng, X. Liang, Y. Li, X. Bu, and X. Zou

***Abstract*—The relative pose estimation is one of the most fundamental components for multi-robot systems, while it still remains an open and challenging research topic in infrastructure-free environment. In this letter, we target improving the accuracy and robustness of relative pose estimation for ground robotic teams, and propose to fuse range and odometry measurements to estimate the relative pose using sliding window optimization. In the system, multiple UWB tags for ranging are equipped on each robot, and visual inertial odometry is applied for estimating the ego-motion pose for each robot. Aiming for simple and effective relative pose initialization, the triangulation uncertainty for multi-tag robots is analyzed, and an initialization method is designed. To cope with the complex environments such as continuous NLOS condition, a NLOS detection and range measurements filtering method is presented. We have conducted series of experiments to demonstrate the performance of the proposed approach.**

***Index Terms*—Multi-robot systems, relative localization, sensor fusion.**

I. Introduction

TEAMS of Unmanned Ground Vehicles (UGVs) have gained increasing popularity in numerous scenarios, such as collaborative surveillance, search and rescue [1]. These applications involving a team of robots performing coordinated tasks require the robots to have a shared understanding of their location. Compared with the single robot localization, which is well-addressed through approaches such as visual-inertial odometry (VIO) [2], [3], the problem of multi-robot co-localization is much more complex, requiring not only the ego-motion poses of each robot, but also the relative poses between robots.

To date, most of existing researches of multi-robot relative localization have been adopting external devices, for instance, motion tracking systems [4], range-based systems with anchors [5], GPS systems [6], [7]. However, such schemes are not friendly to open sites or indoor scenes where is difficult to carry out infrastructure transformation, which not only increases the difficulty and cost of use, but also restricts the application potential of multi-robot system. A popular approach to overcome these limitations is to use vision sensors and visual-detection-based or visual-feature-based [8], [9] methods to estimate the relative pose, but the camera ‘s field of view, ambient illumination, the association of the data from the detection results and other factors all limit the scenarios in which vision-based method can be applied.

Beyond visual sensors, portable range sensors using Ultra-Wideband (UWB) technology is gaining increasing popularity in relative localization research [10] due to its omnidirectional sensing range, environment independence, and data association friendliness. The general approach is to equip each robot with a UWB tag and use the time-of-flight principle to measure the distance between pairs of robots [11], [12]. These single-range based localization approaches usually require persistent relative motion between the robots to initialize the relative pose [13], which reduce the flexibility and safety of robotic formations. The use of multi-tag robot alleviates the unobservability of relative pose [14], [15], but due to the limitation of the short baseline between tags, the relative pose may still be unobservable when the robots are far apart. On the other hand, multi-robot relative localization also faces challenges in Non-Line-of-Sight (NLOS) conditions, especially the continuous NLOS condition common to ground robots, which has not been well handled by traditional methods [16], [17]. We believe implementing effective relative pose initialization and NLOS processing would improve the robustness and accuracy of multi-robot systems in practical applications.

In this spirit, this letter proposes a novel UWB range and VIO fusion relative localization system that integrates the proposed relative pose initialization and NLOS processing methods for accurate and robust relative pose estimation. The main contributions of this letter are summarized as follows:

Manuscript received 15 July 2022; accepted 12 September 2022. Date of publication 21 September 2022; date of current version 30 September 2022. This letter was recommended for publication by Associate Editor D. Schulz and Editor S. Behnke upon evaluation of the reviewers' comments. This work was supported by the Key Research Program of Frontier Science, CAS, under Grant ZDBS-LY-JSC028. *(Corresponding authors: Z. Li; X. Zou.)*

S. Zheng, Y. Liu, H. Zhang, and P. Zheng are with the State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhengshuaikang18@mails.ucas.ac.cn; liuyunfei17@mails.ucas.ac.cn; zhanghaifeng19@mails.ucas.ac.cn; zhengpengcheng20@mails.ucas.ac.cn).

Z. Li is with the State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China (e-mail: ztli@mail.ie.ac.cn).

X. Liang, Y. Li, and X. Bu are with the State Key Laboratory of Microwave Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China (e-mail: xdliang_aircas@163.com; yllee@mail.ie.ac.cn; buxx@aircas.ac.cn).

X. Zou is with the State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China, also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the QiLu Aerospace Information Research Institute, Chinese Academy of Sciences, Jinan 250100, China (e-mail: zouxd@aircas.ac.cn).

Digital Object Identifier 10.1109/LRA.2022.3208354

1) A relative localization system based on UWB range and VIO fusion and using sliding window optimization for accurate and robust relative pose estimation of ground robotic teams.
2) A relative pose initialization method based on triangulation uncertainty to ensure that the relative pose can still be effectively initialized even when the robots are far apart.
3) A NLOS detection and range measurements filtering method based on consistency check to improve the accuracy of the relative pose estimation in NLOS conditions.
4) A series of field experiments to demonstrate the performance of the proposed methods.

Our approach presents a possible solution for accurate and robust relative pose estimation independent of infrastructure, which will be conducive to the practical applications of ground robotic teams in complex environments.

II. Related Work

A. Range-Based Relative Localization

Inter-robot range measurement is becoming increasingly accessible due to the wide application of technologies such as UWB which has the appeal of low cost, low power and high accuracy. In range-based relative localization approaches that do not require anchor, usually each robot carries a UWB tag for inter-robot range measuring. In [11], [12], the range measurements from the UWB tags are combined with the displacement measurements from optical flow method to obtain a relative pose estimation. In [18], a sliding window filter is used to fuse range measurements and 9-axis inertial measurement units (IMUs) to estimate the relative position between two robots. However, as outlined in [13], these single-range based localization approaches usually require persistent relative motion between the robots, and the relative pose can become unobservable in some cases, e.g., parallel motion.

Recently, the idea of using multi-tag robot has been presented for person tracking [19], autonomous docking [20], and so on. In [14], two two-tag robots coupled with an altimeter and optical flow velocity measurements are used for relative localization, and the results are validated in an experiment with limited motion. In [15], a sliding window optimizer fusing multi-range and onboard odometry measurements is used to improve the accuracy of relative pose estimation between robots. In [21], authors analyze the instantaneous observability of the relative pose between two-tag robots with attitude measurements based on the rigidity theory, but the magnetometer for attitude measurement is not always available due to the magnetic interference.

These approaches use multi-tag robot to alleviate the relative pose unobservability, but ignore an important issue, that is, the triangulation uncertainty increases with the distance between robots due to range measurement noise. This results in that they cannot guarantee to initialize the relative pose correctly when the robots are far apart. Therefore, we analyze the triangulation uncertainty for multi-tag robots and propose an initialization method to ensure that the relative pose can still be effectively initialized even when the robots are far apart.

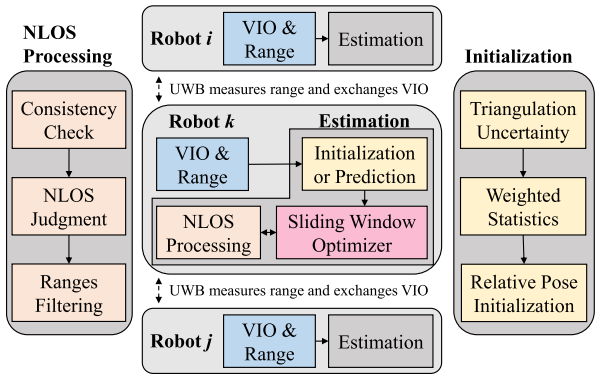


Fig. 1. The system structure of UWB and VIO fusion relative localization of multiple robots. For each robot, use UWB to measure the range to neighbor robots and collect their VIO for relative pose estimation. The system also integrates the proposed initialization and NLOS processing methods.

B. NLOS Processing

In practical application, UWB range may be trapped in NLOS conditions, which may greatly reduce the accuracy of localization, or even lead to localization failure. Therefore, the NLOS detection and error mitigation are very important and they have received attention in the past already [22], [23]. The channel statistics-based approaches [24], [25] are effective for reducing multipath interference but have little direct effect on the ranging errors due to NLOS reception [26]. Hence, we focus on range estimation-based methods, which do not require channel parameters and are more general in most cases.

Hot et al. [16] use the one-step prediction of Extended Kalman Filter (EKF) to discard individual measurements which fall outside of their confidence interval, but do not take consistency between ranges into account. Hepp et al. [19] use Mahalanobis distance to check whether all ranges at a time can be attributed to the supposed distribution and all ranges must be valid at once to be kept. This method considers the consistency between ranges, but will cause the system to fail due to long periods of no available ranges in some cases. Dwek et al. [17] use conditional probabilities to check the consistency between ranges, and iteratively reject the strongest outlier until all ranges agree each other. However, this method is implemented on the assumption that there are enough ranges free of NLOS contamination, which only works in the presence of multiple scattered anchors. In a relative localization system, it is very likely that all ranges between mobile robots is contaminated with NLOS, rendering this approach ineffective.

Therefore, we propose to first use Mahalanobis distance to check the consistency between measurements in the sliding window, which can improve the performance of NLOS detection with more measurements; if NLOS is detected, we perform rigorous ranges filtering to select consistent ranges and improve the accuracy and robustness of the relative localization in NLOS conditions.

III. Method

Fig. 1 illustrates the complete structure of the proposed relative localization system. The data from stereo camera and IMU

on each robot is incorporated for VIO measurements using ORB-SLAM3 [2]. Multiple UWB tags are installed on each robot for inter-robot range measuring. The range and VIO measurements are fused through sliding window optimization to estimate the 2D relative pose estimation. Meanwhile, the proposed relative pose initialization and NLOS processing methods are performed to improve the accuracy and robustness of the system.

A. Problem Formulation

Let's denote the 2D pose of robot $i(i \in [1:N])$ in the world frame at time t as transformation matrix ${}^{w}\mathbf{T}_i^t$. The transformation consists of a position vector ${}^{w}\mathbf{p}_i^t(x,y)$ and a rotation matrix ${}^{w}\mathbf{R}_i^t(\phi)$. Measurements are denoted by $\tilde{(\cdot)}$ and estimated values are indicated with a hat (i.e., $\hat{(\cdot)}$). ${}^{b_i}(\cdot)$ and ${}^{l_i}(\cdot)$ denote the values in the robot body frame and VIO local frame respectively.

For a robotic team that contains N robots, the relative pose estimation problem can be represent as: For an arbitrary robot k, estimate ${}^{b_k}\hat{\mathbf{T}}_i^t$ for each neighbor robot at time t. Together with the ego-motion pose ${}^{l_k}\tilde{\mathbf{T}}_k^t$ (as a measurement) obtained by VIO, the pose of robot i in robot k's local frame at time t can be easily retrieved: ${}^{l_k}\hat{\mathbf{T}}_i^t = {}^{l_k}\tilde{\mathbf{T}}_k^t\,{}^{b_k}\hat{\mathbf{T}}_i^t$.

B. Measurement Modeling

1) UWB Range: The position of UWB tag g on robot i in world frame is expressed as

$$ {}^{w}\mathbf{p}_{i_g}^t = {}^{w}\mathbf{T}_i^t \cdot {}^{b_i}\mathbf{p}_{i_g}, \tag{1} $$

where ${}^{b_i}\mathbf{p}_{i_g}$ is the calibrated extrinsic transformation between the tag frame and the body frame.

Denote $\tilde{r}_{i_g,j_h}^t$ as the range measurement between tag $g(g \in [1:G])$ on robot i and tag $h(h \in [1:H])$ on robot j. Assuming range measurement noise follows an approximately normal distribution $\mathrm{n}_r \sim \mathcal{N}(0,\sigma_r)$, the range measurement can be modeled as

$$ \begin{aligned} \tilde{r}_{i_g,j_h}^t &= ||{}^{w}\mathbf{p}_{i_g}^t - {}^{w}\mathbf{p}_{j_h}^t||_2 + \mathrm{n}_r \\ &= ||{}^{w}\mathbf{T}_i^t \cdot {}^{b_i}\mathbf{p}_{i_g} - {}^{w}\mathbf{T}_j^t \cdot {}^{b_j}\mathbf{p}_{j_h}|| + \mathrm{n}_r. \end{aligned} \tag{2} $$

There are multiple range measurements between every two robots, and all range measurements at the same time between two robots is abbreviated as $\mathbf{z}_{r_{i,j}}^t = \{\tilde{r}_{i_g,j_h}^t\}_{g,h=1}^{G,H}$.

2) Visual Inertial Odometry: In our system, odometry output of VIO is used to provide real-time local pose estimation. In terms of statistics, the first-order Markov process dominates the long-term drift of VIO. [27]. In our approach, we make use of a simplified model in order to reduce the complexity of the optimization scheme and use relative pose extracted from odometry in the pose solver to cope with the difference of the local frames, which can be modeled as

$$ \mathbf{z}_{\delta\mathbf{T}_i}^t = \left({}^{l_i}\tilde{\mathbf{T}}_i^{t-1}\right)^{-1}\left({}^{l_i}\tilde{\mathbf{T}}_i^t\right) = \left({}^{w}\mathbf{T}_i^{t-1}\right)^{-1}\left({}^{w}\mathbf{T}_i^t\right) + \mathbf{n}_o, \tag{3} $$

where the noise of the relative odometry pose is assumed as Gaussian $\mathbf{n}_o \sim \mathcal{N}(0,\mathbf{\Sigma}_o)$ and $\mathbf{\Sigma}_o$ can be obtained through the statistical results of VIO drift. Moreover, with the help of accelerometers and gyroscopes, the z-axis of the local frame of all robots can be aligned with the gravity direction.

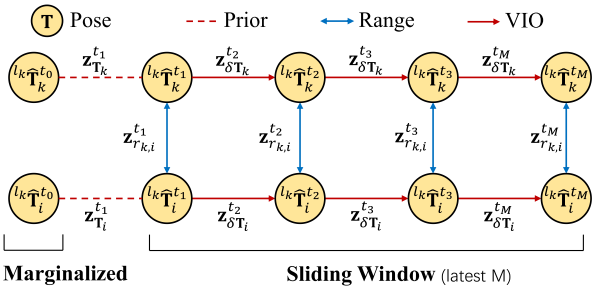


Fig. 2. An illustration of the sliding window relative pose estimation. The sliding window uses all range and odometry measurements at multiple consecutive times to jointly estimate the poses of all neighbor robots relative to robot k. The poses before the window are marginalized and provide the prior measurements.

C. Optimization-Based Relative Pose Estimation

For robot k, exchange data with neighbor robots at a certain frequency to measure the ranges to neighbors and collect the odometry of each neighbor simultaneously. A sequence of measurements in chronological order forms a sliding window for optimization. An illustration of a sliding window is shown in Fig. 2. Note that due to the nature of temporal information coupling [28], instead of extracting partial-time ranges, we use all-time ranges for estimation.

Once the window size reaches $M(M$ time steps aligned with ranges), we solve the relative pose estimation problem through sliding window optimization. The full state vector $\mathcal{X}_k$ for optimization is defined as

$$ \mathcal{X}_k = \left\{ \underbrace{\overbrace{{}^{l_k}\hat{\mathbf{T}}_1^{t_1}, {}^{l_k}\hat{\mathbf{T}}_1^{t_2} \ldots {}^{l_k}\hat{\mathbf{T}}_1^{t_M}}^{Melements}, {}^{l_k}\hat{\mathbf{T}}_2^{t_1}, \ldots {}^{l_k}\hat{\mathbf{T}}_N^{t_M}}_{M \times Nelements} \right\}. \tag{4} $$

By minimizing the Mahalanobis norm of the residuals of the measurements, a maximum a posterior estimation can be obtained. The optimization is expressed as

$$ \begin{aligned} f(\mathcal{X}_k) := \min_{\mathcal{X}_k} \Bigg\{ & \sum_{i=1:N,t=t_1} E_p(\mathbf{z}_{\mathbf{T}_i}^t, \mathcal{X}_k) \\ & + \sum_{(k,i,t)\in\mathcal{R}} E_r(\mathbf{z}_{r_{k,i}}^t, \mathcal{X}_k) \\ & + \sum_{(i,t)\in\mathcal{O}} E_o(\mathbf{z}_{\delta\mathbf{T}_i}^t, \mathcal{X}_k) \Bigg\}, \end{aligned} $$

$$ E_p(\mathbf{z}_{\mathbf{T}_i}^t, \mathcal{X}_k) = \left\| \left((\mathbf{z}_{\mathbf{T}_i}^t)^{-1} ({}^{l_k}\hat{\mathbf{T}}_i^t) \right) \right\|_{\mathbf{\Sigma}_p^{-1}}^2 $$

$$ \begin{aligned} E_r(\mathbf{z}_{r_{k,i}}^t, \mathcal{X}_k) = \sum_{g,h=1}^{G,H} \rho \Bigg\| & \tilde{r}_{k_g,i_h}^t - \Big\| {}^{l_k}\hat{\mathbf{T}}_k^t \cdot {}^{b_k}\mathbf{p}_{k_g} \\ & - {}^{l_k}\hat{\mathbf{T}}_i^t \cdot {}^{b_i}\mathbf{p}_{i_h} \Big\| \Bigg\|_{\sigma_r^{-1}}^2, \end{aligned} $$

$$ E_o(\mathbf{z}_{\delta\mathbf{T}_i}^t, \mathcal{X}_k) = \left\| (\mathbf{z}_{\delta\mathbf{T}_i}^t)^{-1} \left(({}^{l_k}\hat{\mathbf{T}}_i^{t-1})^{-1} ({}^{l_k}\hat{\mathbf{T}}_i^t) \right) \right\|_{\mathbf{\Sigma}_o^{-1}}^2, \tag{5} $$

where $\mathcal{R}$ is the set of all range measurements, O is the set of all odometry measurements. The notation $\|\cdot\|^2_{\mathbf{W}}$ denotes the squared Mahalanobis norm with the information matrix $\mathbf{W}$. $E_p(\mathbf{z}^t_{\mathbf{T}_i}, \mathcal{X}_k)$ is the residual of prior item, $\mathbf{z}^t_{\mathbf{T}_i}$ is the result from previous window and $\sum_p$ is the covariance carried over from historical measurements [28]. $E_r(\mathbf{z}^t_{r_{k,i}}, \mathcal{X}_k)$ is the residual of the range measurements. $E_o(\mathbf{z}^t_{\delta\mathbf{T}_i}, \mathcal{X}_k)$ is the residual of odometry measurements, ensuring the local consistency of the robot i's poses. Since some anomalous measurements may be generated in range measurements, we adopt Huber norm $\rho(\cdot)$ [29] to reduce the effect of possible outlier.

The Ceres-solver [30] is adopted to solve the optimization problem, while the Trust Region with the Levenberg-Marquardt method is chosen. After optimization, the oldest poses in the window will be marginalized and provide the prior measurements. Meanwhile, the latest odometry measurement can be used to predict the robot pose as the initial value for the optimizer:

$$ {}_{l_k}\breve{\mathbf{T}}^t_i = {}_{l_k}\hat{\mathbf{T}}^{t-1}_i \left({}_{l_i}\tilde{\mathbf{T}}^{t-1}_i \right)^{-1} {}_{l_i}\tilde{\mathbf{T}}^t_i. \tag{6} $$

D. NLOS Detection and Ranges Filtering

In addition to using the robust Huber norm as (5) to reduce the impact of outliers, we also design a NLOS detection method based on consistency check, as well as a range measurements filtering method based on normal distribution to improve the accuracy of the system.

As we know, the range measurement error will deviate from the normal distribution in NLOS conditions, and hypothesis testing is usually used to detect NLOS. The general method [16] is to directly reject the range that are inconsistent with the prediction of the odometry as

$$ \left| \tilde{r}^t_{k_g,i_h} - \left\| {}_{l_k}\tilde{\mathbf{T}}^t_k \cdot {}^{b_k}\mathbf{p}_{k_g} - {}_{l_k}\breve{\mathbf{T}}^t_i \cdot {}^{b_i}\mathbf{p}_{i_h} \right\|_{\text{abs}} \right| > 2\sigma_r. \tag{7} $$

In fact, similar to the Huber norm, this method can only reject strong outliers without considering the consistency between measurements. Therefore, for more robust detection of NLOS, we perform chi-square test based on the Mahalanobis distance of the residuals of all measurements in the sliding window. NLOS is detected according to

$$ \chi^2_n = \sum_{(k,i,t)\in\mathcal{R}} E_r(\mathbf{z}^t_{r_{k,i}}, \mathcal{X}_k) + \sum_{(i,t)\in\mathcal{O}} E_o(\mathbf{z}^t_{\delta\mathbf{T}_i}, \mathcal{X}_k) \geq \chi^2_{thresh}, \tag{8} $$

where χ^2_{thresh} is equal to the 0.9 probability quantile of the χ^2 distribution of n degrees of freedom and can be proper adjusted in practice. The NLOS detection is performed after the optimization of problem (5). If NLOS is not detected, the optimization result is used as estimation output, otherwise the range measurements are filtered to re-optimize the problem.

For a set of consistent range measurements, the error should be close to a normal distribution, that is, the mean of the error is close to zero, and the error is concentrated around the mean. Therefore, the residuals set $\{\mathrm{e}^{+,1}_r, \cdots \mathrm{e}^{+,A}_r, \mathrm{e}^{-,1}_r, \cdots \mathrm{e}^{-,B}_r\}$ of the selected consistent range measurements should be satisfied with

$$ \sum_{1\leq a\leq A} \mathrm{e}^{+,a}_r \leq res\_sum < \sum_{1\leq a\leq A+1} \mathrm{e}^{+,a}_r, $$

$$ \sum_{1\leq b\leq B} -\mathrm{e}^{-,b}_r \leq res\_sum < \sum_{1\leq b\leq B+1} -\mathrm{e}^{-,b}_r, $$

$$ res\_sum = \min\left(\sum_{1\leq a\leq A'} \mathrm{e}^{+,a}_r, \sum_{1\leq b\leq B'} -\mathrm{e}^{-,b}_r) \right), $$

$$ -2\sigma_r \leq \mathrm{e}^{-,B'}_r \leq \cdots \mathrm{e}^{-,1}_r < 0 \leq \mathrm{e}^{+,1}_r \leq \cdots \mathrm{e}^{+,A'}_r \leq 2\sigma_r, \tag{9} $$

where $\{\mathrm{e}^{+,1}_r, \cdots \mathrm{e}^{+,A'}_r\}$ and $\{\mathrm{e}^{-,1}_r, \cdots \mathrm{e}^{-,B'}_r\}$ are two sets of residuals of range measurements divided according to positive and negative, and each has been arranged in order. This method of selecting consistent measurements as much as possible rather than rejecting all measurements helps to reduce the influence of odometry drift and priori error, which is necessary in continuous NLOS conditions. Meanwhile, one-time ranges filtering can be realized with the aid of the odometry, instead of iterative selection like other methods [17].

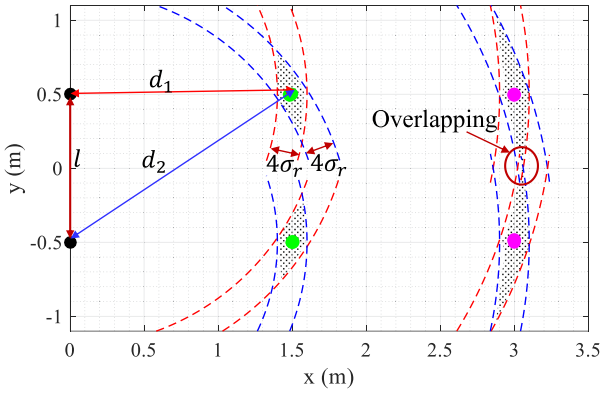


Fig. 3. The analytical results of possible positions of every tag using circle geometry, as the shaded areas bounded by the dashed circle arcs. Solid circles of the same color represent true positions of the tags on the same robot, and we consider initializing the poses of the nearer robot (green circles) and the farther robot (magenta circles) relative to the reference robot (black circles). l is the baseline between tags on one robot. d_1 and d_2 are the reference distances for one tag, and the confidence interval for range measurements is $\pm 2\sigma_r$. The possible position areas of the tags on the farther robot partially overlap, indicating that the relative pose initialization may fail.

E. Triangulation Uncertainty and Relative Pose Initialization

The use of multi-tag robot simplifies the relative pose initialization and enables single-shot relative pose estimation when the robots are close together. But, due to the triangulation uncertainty caused by range measurements noise, the relative pose estimation using single-shot measurements may still be invalid when the robots are far apart. Therefore, by analyzing the triangulation uncertainty, we propose a relative pose initialization method to extend the coverage of the multi-tag robotic teams.

Fig. 3 shows how a multi-tag robot uses circular geometry to determine the position of tags on other robots at different distances. It can be seen that the possible position areas

of the tags are changed with inter-robot distances, and the position areas of the two tags on the farther robot partially overlap, which may cause the relative orientation to be invalid. The description of the possible position area is already characterized as Horizontal Dilution of Precision (HDOP). The computed HDOP [31] for the 2D case with the baseline l between the UWB tags and the distance d^t between the robots is

$$HDOP(d^t) = \sqrt{2}\frac{\sqrt{d^t \cdot d^t + 4l \cdot l}}{2l}. \tag{10}$$

The trilateration uncertainty caused by the noise of range is mainly distributed in the tangential direction of the connection between the robots and the uncertainty in the tangential direction can be assumed to be normally distributed with a standard deviation of $\sigma^t = HDOP(d^t) \cdot \sigma_r$.

From the odometry trajectory, the relative position between tags on a robot at different times can be obtained, which can be seen as virtual baseline between tags, denoted as l_v^t. A simple way to ensure proper relative pose estimation is to make l_v^t long enough so that the tags' enclosed areas do not overlap:

$$l_v^{t_m} = \left\| {}^{l_i}\tilde{\mathbf{T}}_i^{t_m} \cdot {}^{b_i}\mathbf{p}_{i_1} - {}^{l_i}\tilde{\mathbf{T}}_i^{t_1} \cdot {}^{b_i}\mathbf{p}_{i_2} \right\| \geq 4\left(\sigma^{t_m} + \sigma^{t_1}\right). \tag{11}$$

This method must wait for the robot to move a sufficient distance. From another point of view, the invalid orientation estimation is a probabilistic event, which is determined by the proportion of overlap areas as

$$p^{t_m} = \frac{1}{2}\mathrm{erf}\left(\frac{-l_v^{t_m}}{2\sqrt{2}\sigma^{t_m}}\right). \tag{12}$$

Where $\mathrm{erf}(\cdot)$ is the error function representing the indefinite integral of the Gaussian function. The statistical methods can be used to find the correct results as much as possible from multiple estimations (samples). The calculation of the sample size for a proportion can be found in [32]. But this method does not take into account the motion of the robot.

In order to include the contribution of both estimation times and robot motion to speed up initialization, we propose a weighted statistics-based method. Assuming that each range measurement is independent, and the required estimation times n needs to meet:

$$\begin{aligned} \sum {}^{1}\!/_{s^{t_m}} &\geq 1, m = 1, 2, \cdots n, \\ s^{t_m} &= Z_{\alpha/2}^2 \frac{p^{t_m}(1-p^{t_m})}{e^2} + c, \\ \alpha &= 0.001, e = 0.01, c = 20. \end{aligned} \tag{13}$$

Where s^{t_m} is the expected sample size at t_m and represents the weight, α is the significance level, $Z_{\alpha/2}$ is the value of normal distribution for $\alpha/2$, e is the accepted error, c is a constant represents the minimum sample size. From these estimations, we can count the proper relative pose using histogram. In each estimation, the relative pose is obtained from the positions of the tags, which can be solved by

$$\begin{aligned} \min_{{}^{l_k}\hat{\mathbf{p}}_{i_1}^{t_m}, {}^{l_k}\hat{\mathbf{p}}_{i_2}^{t_1}} \Bigg\{ & \sum_{g\in[1:G]} \left\| \tilde{r}_{k_g,i_1}^{t_m} - \left\| {}^{l_k}\hat{\mathbf{T}}_k^{t_m} \cdot {}^{b_k}\mathbf{p}_{k_g} - {}^{l_k}\hat{\mathbf{p}}_{i_1}^{t_m} \right\| \right\|_{\sigma_r^{-1}}^2 \\ & + \sum_{g\in[1:G]} \left\| \tilde{r}_{k_g,i_2}^{t_1} - \left\| {}^{l_k}\hat{\mathbf{T}}_k^{t_1} \cdot {}^{b_k}\mathbf{p}_{k_g} - {}^{l_k}\hat{\mathbf{p}}_{i_2}^{t_1} \right\| \right\|_{\sigma_r^{-1}}^2 \\ & + \left\| \left\| {}^{l_i}\tilde{\mathbf{T}}_i^{t_m} \cdot {}^{b_i}\mathbf{p}_{i_1} - {}^{l_i}\tilde{\mathbf{T}}_i^{t_1} \cdot {}^{b_i}\mathbf{p}_{i_2} \right\| - \left\| {}^{l_k}\hat{\mathbf{p}}_{i_1}^{t_m} - {}^{l_k}\hat{\mathbf{p}}_{i_2}^{t_1} \right\| \right\|_{\sigma_r^{-1}}^2 \Bigg\}. \end{aligned} \tag{14}$$

Where the third item is the virtual baseline length constraint to improve the accuracy of the relative orientation. Usually, we use the robot with 4 tags as the reference robot, which can avoid the flip ambiguity of the relative pose, and the enclosed area formed by the 4 distances is almost the same as that by 2 distances when the robots are far apart. After this, the sliding window optimizer can be implemented to obtain accurate relative pose estimation.

Fig. 4. The experiment field and one robot platform in the system, which is equipped with a RealSense D435i stereo camera, 4 Noploop UWB tags and a Nvidia Jetson AGX embedded computer. The dual antenna GNSS-RTK is used to provide location truth. The only difference in the hardware of the other robots is that there are only 2 UWB tags.

IV. Experiment

To fully demonstrate the proposed relative localization system in this letter, the customized multi-robot platforms is designed and implemented. As shown in Fig. 4, every robot is equipped with multiple Noploop UWB tags, an Intel RealSense D435i stereo camera integrated with IMU module and a Nvidia Jetson AGX embedded computer. The raw data from the IMU at 200 Hz and images from the stereo camera at 30 Hz are fused together for the VIO estimation of individual robot using ORB-SLAM3. The baseline between UWB tags on every robot is 0.68 m, the maximum output frequency is 50 Hz and the nominal range standard deviation is 0.05 m. The UWB tags

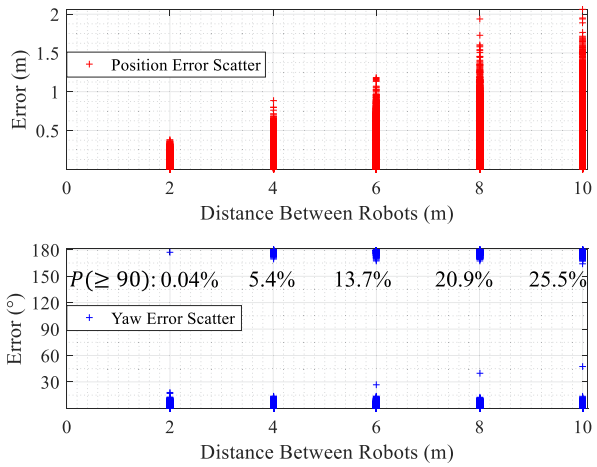


Fig. 5. Relative pose estimation error distribution at different distances between 2 robots using only single-shot range measurements, which is obtained by 10000 Monte Carlo simulations, respectively. The bottom figure also shows the proportion of relative orientation estimation failures (greater than 90°). The baseline between tags is 0.68 m, and the range standard deviation σ_r is 0.05 m.

complete the broadcast of data with synchronized time stamps while measuring the inter-robot distances. For the VIO drift, we use $\mathbf{\Sigma}_{o;x,y} = 0.005$ m, $\mathbf{\Sigma}_{o;\phi} = 0.05\,\text{deg}$. This results in a drift of 4% of the trajectory length, which is a typical order for state-of-the-art VIO.

In order to validate the robustness and accuracy of our system, the experiments are divided into three parts: 1): Initialization tests at different inter-robot distances as well as simulations to show relative pose estimation error distribution using only single-shot range measurements; 2): NLOS condition test to verify the effect of our NLOS processing algorithm; 3): Three-robot test to show the accuracy of our relative localization system, which also shows the results of different methods through ablation studies. The ground truth is achieved from dual antenna GNSS-RTK. The relative pose estimation algorithms are running on the individual robot platforms and the entire system is built under the ROS framework. The evaluation indicators of relative position error and orientation error (two-dimensional) are both Root Mean Square Error (RMSE).

A. *Initialization Test*

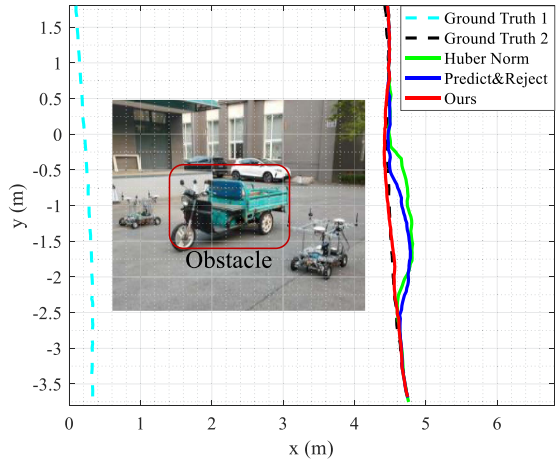
In this experiment, we first show the relative pose estimation error distribution using only single-shot range measurements through simulations. Second, we test the relative pose initialization ability of the system at different inter-robot distances through field experiment.

In the simulation, we analyze the relative pose estimation error distribution using only single-shot range measurements. From Fig. 5 we can see that, consistent with the uncertainty analysis in Section III-E, as the distance between robots increases, the relative position error becomes more dispersed, that is, the average error becomes larger. Meanwhile, the relative orientation error presents a state of bipolar distribution, and the proportion of the error greater than 90° increases with the distance between

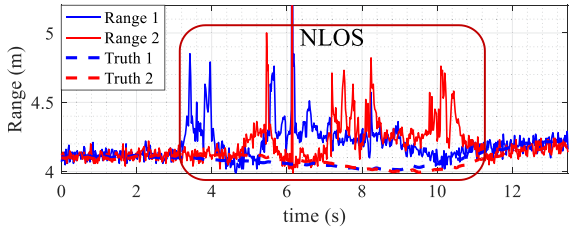
TABLE I
INITIALIZATION ACCURACY IN FIELD TEST

Inter-Robot Distance (m)	Estimations Number	Moving Distance (m)	Position Error (m)	Yaw Error (°)
2	21	0.08	0.099	2.314
4	41	0.28	0.155	2.185
6	78	0.41	0.228	3.036
8	99	0.59	0.269	2.745
10	96	0.94	0.233	2.492

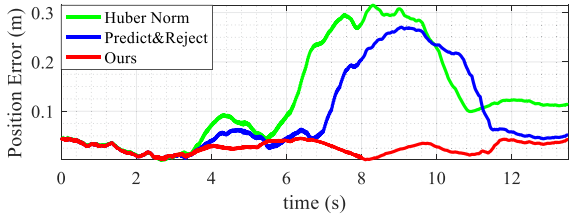
The inter-robot distances are approximate values measured by hand. The reference robot is robot 1 and the moving distances is obtained from the odometry of robot 2.



(a) The trajectories obtained by using different NLOS processing methods (Huber norm in (5), use VIO prediction to reject outliers like (7), and our proposed method). Trajectory of robot 1 is obtained by RTK to visualize the relative pose estimation results of robots 2.



(b) Range measurements between robots (only two are displayed) in the continuous NLOS condition. The truth is calculated from RTK.



(c) Position error of robot 2 using different NLOS processing methods.

Fig. 6. Results of the NLOS condition test. (a) The trajectories of the two robots. (b) Range measurements between the robots. (c) Position error of robot 2.

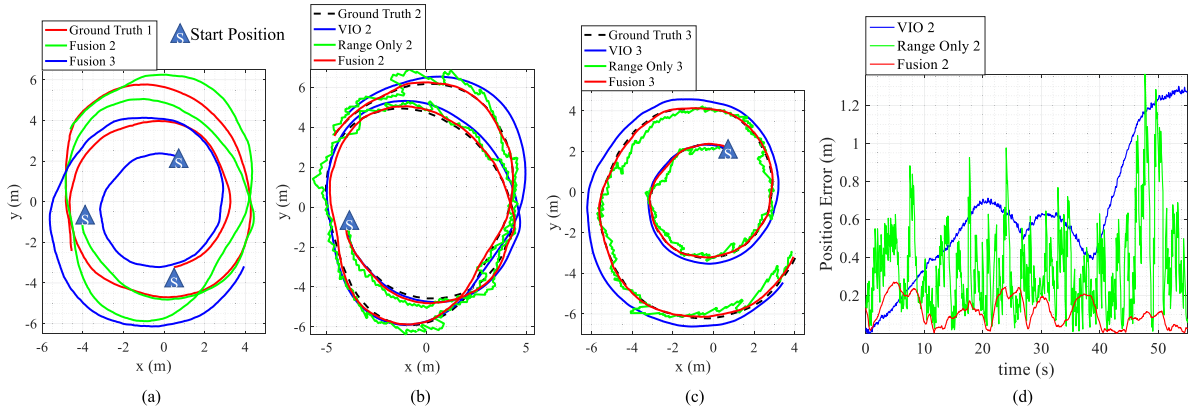


Fig. 7. The results of three-robot test. (a) The trajectories of all robots. (b)–(c) The detailed comparison of the trajectories obtained from different methods (VIO: The result obtained by aligning the VIO with RTK in the first 2 seconds; Range Only: The result obtained by using only range measurements; Fusion: The result obtained by fusion of range and VIO). Trajectory of robot 1 is directly obtained by RTK to visualize the relative pose estimation results of robots 2 and 3. (d) The position error curve of robot 2 obtained from different methods.

robots, which means that the sliding window optimizer is more likely to fail.

In the field experiment, we test the effect of the initialization method by making two robots move in parallel at different inter-robot distances, and Table I shows the relative pose initialization results. It can be seen that as the inter-robot distance increases, more robot motion and more estimations are required to meet the initialization condition. At the same time, the proposed weighted statistics-based method can consider the contribution of both estimation times and robot motion to speed up initialization. This test proves the initialization ability of the system, which is important to ensures the correct execution of the optimizer and expands the coverage of the robotic teams.

B. NLOS Condition Test

In this experiment, as shown in Fig. 6(a), the two robots are separated by obstacle during near-parallel movement to verify the effect of our proposed NLOS processing algorithm. As shown in Fig. 6(b), when being blocked by obstacle, range measurements between robots have partial loss and many outliers. As shown in Fig. 6(c), we compared the positioning accuracy of different algorithms, including 1) Huber Norm: only use Huber norm in optimizer like (5); 2) Predict & Reject: reject outliers according to the prediction of VIO like (7); and 3) our proposed NLOS processing algorithm. Moreover, the trajectory of robot 1 is directly obtained from ground truth to visualize the results of relative pose estimation.

From the results we can see that, compared with using only Huber norm, the effect of the method using the prediction results to reject outliers is slightly improved, but it is still not satisfactory. While our method of range measurements filtering is hardly affected by NLOS, which improves the robustness and accuracy of relative pose estimation of robotic teams in complex environment.

C. Three-Robot Test

In this section, we use 3 robots to discuss the performance of the UWB range and VIO fused relative localization system.

TABLE II
VIO DRIFT ERROR IN THREE-ROBOT TEST

	Trajectory Length (m)	VIO (m | °)
Robot 1	50.81	0.706 | 5.531
Robot 2	58.18	0.616 | 4.014
Robot 3	44.71	0.697 | 5.621

RELATIVE LOCALIZATION ACCURACY IN THREE-ROBOT TEST

	Fusion (m | °)	Range Only (m | °)
Robot 2 to 1	**0.113 | 3.234**	0.366 | 5.357
Robot 3 to 1	**0.086 | 2.489**	0.321 | 5.297
Robot 3 to 2	**0.144 | 3.750**	0.594 | 6.750

The data in the table are position error (m) and yaw error (°), separated by vertical symbol.

In the experiment, the three robots circle a field of size over 10 m $\times$ 12 m in the counterclockwise direction. Fig. 7 shows the trajectories obtained by the ablation studies and the position error curve of robot 2, and in order to visualize the results of relative pose estimation, the trajectory of robot 1 is directly obtained from RTK. It can be seen that the VIO drifts more and more with time, and the position obtained only by range is very jittery, while the fusion of range and VIO can obtain stable and drift-free localization results.

The relative pose estimation results obtained by different methods are shown in Table II. It can be seen that our UWB and VIO fusion relative position error can reach within 0.15m, which reaches the state-of-the-art in range-based relative localization approaches and can meet the needs of many multi-robot applications.

In addition, the algorithm runs on the CPU of ARMv8.2@2.26GHz (Jetson AGX), the computer system is Ubuntu 18.04, and the project is based on ROS and C++. The bandwidth required to exchange floating-point odometry data with a maximum rate of 30Hz is less than 500 Bytes per second, which is very easy for UWB technology. All range and odometry measurements are used for estimation and the oldest measurements in the sliding window are marginalized

after optimization. The size of sliding window is 20 and the time consumption of sliding window optimization is less than 10 milliseconds, which means it can accurately estimate the relative pose in real time.

V. Conclusion

In this letter, a novel UWB range and VIO fusion relative pose estimation approach using sliding window optimization is introduced, which provides a possible solution for relative localization of ground robotic teams in infrastructure-free environments. The proposed relative pose initialization method extends the coverage of the multi-robot teams, the NLOS processing method helps the system to run in extreme NLOS condition, both of which improve the robustness and accuracy of the system and will be conducive to the practical applications of multi-robot systems in complex environments. In the series of field experiments, the effect of the initialization and NLOS processing methods is demonstrated and the accuracy within 0.15m proves the performance of the system. In addition, there are still some areas to be improved. One direction of future work is to improve our approach and apply it to multi-robot systems in 3D space. Another point is to integrate collaborative mapping and other functions to extend the application scenarios of the system.

References

[1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2568–2599, Thirdquarter 2019.

[2] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021, doi: 10.1109/TRO.2021.3075644.

[3] T. Qin, P. Li, and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018, doi: 10.1109/TRO.2018.2853729.

[4] J. A. Preiss, W. Honig, G. S. Sukhatme, and N. Ayanian, "Crazyswarm: A large nano-quadcopter swarm," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 3299–3304.

[5] A. Ledergerber, M. Hamer, and R. D'Andrea, "A robot self-localization system using one-way ultra-wideband communication," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 3131–3137.

[6] A. Jaimes, S. Kota, and J. Gomez, "An approach to surveillance an area using swarm of fixed wing and quad-rotor unmanned aerial vehicles UAV(s)," in *Proc. IEEE Int. Conf. Syst. Syst. Eng.*, 2008, pp. 1–6.

[7] S. Moon, Y. Choi, D. Kim, M. Seung, and H. Gong, "Outdoor swarm flight system based on RTK-GPS," *J. KIISE*, vol. 43, no. 12, pp. 1315–1324, 2016.

[8] P.-Y. Lajoie, B. Ramtoula, Y. Chang, L. Carlone, and G. Beltrame, "DOOR-SLAM: Distributed, online, and outlier resilient SLAM for robotic teams," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1656–1663, Apr. 2020, doi: 10.1109/LRA.2020.2967681.

[9] H. Xu et al., "Omni-swarm: A decentralized omnidirectional visual-inertial-UWB state estimation system for aerial swarm," *IEEE Trans. Robot.*, to be published, doi: 10.1109/TRO.2022.3182503.

[10] W. Shule, C. M. Almansa, J. P. Queralta, Z. Zou, and T. Westerlund, "UWB-based localization for multi-UAV systems and collaborative heterogeneous multi-robot systems," *Procedia Comput. Sci.*, vol. 175, pp. 357–364, 2020.

[11] K. Guo, X. Li, and L. Xie, "Ultra-wideband and odometry-based cooperative relative localization with application to multi-UAV formation control," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2590–2603, Jun. 2020.

[12] T.-M. Nguyen, Z. Qiu, T. H. Nguyen, M. Cao, and L. Xie, "Distance-based cooperative relative localization for leader-following control of MAVs," *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 3641–3648, Oct. 2019, doi: 10.1109/LRA.2019.2926671.

[13] P. Batista, C. Silvestre, and P. Oliveira, "Single range aided navigation and source localization: Observability and filter design," *Syst. Control Lett.*, vol. 60, no. 8, pp. 665–673, 2011, doi: 10.1016/j.sysconle.2011.05.004.

[14] T.-M. Nguyen, A. H. Zaini, C. Wang, K. Guo, and L. Xie, "Robust target-relative localization with ultra-wideband ranging and communication," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 2312–2319.

[15] Z. Cao et al., "Relative localization of mobile robots with multiple ultra-wideband ranging measurements," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 5857–5863, doi: 10.1109/IROS51168.2021.9636017.

[16] J. D. Hol, F. Dijkstra, H. Luinge, and T. B. Schöny, "Tightly coupled UWB/IMU pose estimation," in *Proc. IEEE Int. Conf. Ultra-Wideband*, 2009, pp. 688–692.

[17] N. Dwek et al., "Improving the accuracy and robustness of ultra-wideband localization through sensor fusion and outlier detection," *IEEE Robot. Automat. Lett.*, vol. 5, no. 1, pp. 32–39, Jan. 2020, doi: 10.1109/LRA.2019.2943821.

[18] C. C. Cossette, M. Shalaby, D. Saussié, J. R. Forbes, and J. le Ny, "Relative position estimation between two UWB devices with IMUs," *IEEE Robot. Automat. Lett.*, vol. 6, no. 3, pp. 4313–4320, Jul. 2021.

[19] B. Hepp, T. Nägeli, and O. Hilliges, "Omni-directional person tracking on a flying robot using occlusion-robust ultra-wideband signals," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2016, pp. 189–194, doi: 10.1109/IROS.2016.7759054.

[20] T. M. Nguyen, T. H. Nguyen, M. Cao, Z. Qiu, and L. Xie, "Integrated UWB-vision approach for autonomous docking of UAVs in GPS-denied environments," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 9603–9609, doi: 10.1109/ICRA.2019.8793851.

[21] M. Shalaby, C. C. Cossette, J. R. Forbes, and J. le Ny, "Relative position estimation in multi-agent systems using attitude-coupled range measurements," *IEEE Robot. Automat. Lett.*, vol. 6, no. 3, pp. 4955–4961, Jul. 2021, doi: 10.1109/LRA.2021.3067253.

[22] J. Khodjaev, Y. Park, and A. S. Malik, "Survey of NLOS identification and error mitigation problems in UWB-based positioning algorithms for dense environments," *Ann. Des. Telecommun./Ann. Telecommun.*, vol. 65, no. 5/6, pp. 301–311, 2010, doi: 10.1007/s12243-009-0124-z.

[23] Z. Zeng, S. Liu, and L. Wang, "UWB/IMU integration approach with NLOS identification and mitigation," in *Proc. IEEE 52nd Annu. Conf. Inf. Sci. Syst.*, 2018, pp. 1–6.

[24] S. Maranò, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1026–1035, Sep. 2010, doi: 10.1109/JSAC.2010.100907.

[25] D. Neirynck, M. O'Duinn, and C. McElroy, "Characterization of the NLOS performance of an IEEE 802.15.4a receiver," in *Proc. 22nd IET Ir. Signals Syst. Conf.*, 2015.

[26] P. D. Groves and Z. Jiang, "Height aiding, C/N_0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas," *J. Navigation*, vol. 66, no. 5, pp. 653–669, 2013, doi: 10.1017/S0373463313000350.

[27] R. Jiang, R. Klette, and S. Wang, "Statistical modeling of long-range drift in visual odometry," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 214–224.

[28] Y. Shen, S. Mazuelas, and M. Z. Win, "Network navigation: Theory and interpretation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 9, pp. 1823–1834, Oct. 2012, doi: 10.1109/JSAC.2012.121028.

[29] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964, doi: 10.1214/aoms/1177703732.

[30] S. Agarwal et al., "Ceres solver," 2010. [Online]. Available: http://ceres-solver.org

[31] R. B. Langley et al., "Dilution of precision," *GPS world*, vol. 10, no. 5, pp. 52–59, 1999.

[32] G. D. Israel, *Determining Sample Size 1 The Level of Precision*. Gainesville, FL, USA: Univ. Florida, 1992.