Mitigation of odometry drift with a single ranging link in GNSS-limited environments

Young-Hee Lee[1], Chen Zhu[2], Gabriele Giorgi[2], and Christoph Günther[1,2]

[1]*Institute for Communications and Navigation, Technical University of Munich, Germany*
younghee.lee@tum.de
[2]*Institute of Communications and Navigation, German Aerospace Center (DLR), Germany*
{chen.zhu, gabriele.giorgi, christoph.guenther}@dlr.de

BIOGRAPHY

Ms. Young-Hee Lee is an academic researcher at the Institute for Communications and Navigation, Technical University of Munich, Germany. She received her Bachelor's and Master's degrees both in Aerospace Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea.

Mr. Chen Zhu is a researcher at the Institute of Communications and Navigation, German Aerospace Center (DLR), and he received his Ph.D. degree at the Institute for Communications and Navigation, Technical University of Munich, Germany. He received his Bachelor's degree in Automation Engineering from Tsinghua University, Beijing, China, in 2009, and his Master's degree in Communications Engineering from Technical University of Munich, Germany, in 2011. He is interested in the research fields of visual navigation, multi-sensor fusion, robotic swarm navigation, currently focusing on the system integrity.

Dr. Gabriele Giorgi is a researcher at the Institute of Communications and Navigation, German Aerospace Center (DLR). He has Bachelor's degree in Aerospace Engineering, and a Master's degree in Space Engineering from the University of Rome "La Sapienza". He received a Ph.D. degree from the Delft Institute of Earth Observation and Space Systems (DEOS), Delft University of Technology, The Netherlands. His current research interests are satellite navigation, visual navigation, and multi-sensor fusion.

Prof. Christoph Günther received his diploma in 1979 and his Ph.D. in 1984 from the Swiss Federal Institute of Technology in Zurich, Switzerland in theoretical physics. He worked on cryptography, coding, communication, and information theory at Asea Brown Boveri, Ascom and Ericsson, and he became a head of the Institute of Communications and Navigation at the German Aerospace Center (DLR) in 2003. In addition, he became a professor of the Institute for Communications and Navigation at Technical University of Munich in 2004.

ABSTRACT

Vision-based systems can estimate the vehicle's positions and attitude with a low cost and simple implementation, but the performance is very sensitive to environmental conditions. Moreover, estimation errors are accumulated without a bound since visual odometry is a dead-reckoning process. To improve the robustness to environmental conditions, vision-based systems can be augmented with inertial sensors, and the loop closing technique can be applied to reduce the drift. However, only with on-board sensors, vehicle's poses can only be estimated in a local navigation frame, which is randomly defined for each mission. To obtain globally-referred poses, absolute position estimates obtained with GNSS can be fused with on-board measurements (obtained with either vision-only or visual-inertial odometry). However, in many cases (e.g. urban canyons, indoor environments), GNSS-based positioning is unreliable or entirely unavailable due to signal interruptions and blocking, while we can still obtain ranging links from various sources, such as signals of opportunity or low cost radio-based ranging modules. We propose a graph-based data fusion method of the on-board odometry data and ranging measurements to mitigate pose drifts in environments where GNSS-based positioning is unavailable. The proposed algorithm is evaluated both with synthetic and real data.

INTRODUCTION

Vision-based systems can estimate the vehicle's positions and attitude (poses), so they are used on many platforms with relatively low cost and simple on-board hardware, such as UAVs and autonomous cars. However, it is challenging to achieve high quality estimation accuracy and system robustness only with a vision sensor when e.g. the light conditions keep changing or the environment does not have a sufficient number of features (e.g. white wall).

To improve the performance in such environments, vision-based systems can be augmented with inertial sensors (gyroscopes and accelerometers) since they provide additional translational and angular velocities measurements to the system. However, visual-inertial odometry is still a dead-reckoning process, in which the error accumulates without a bound. In

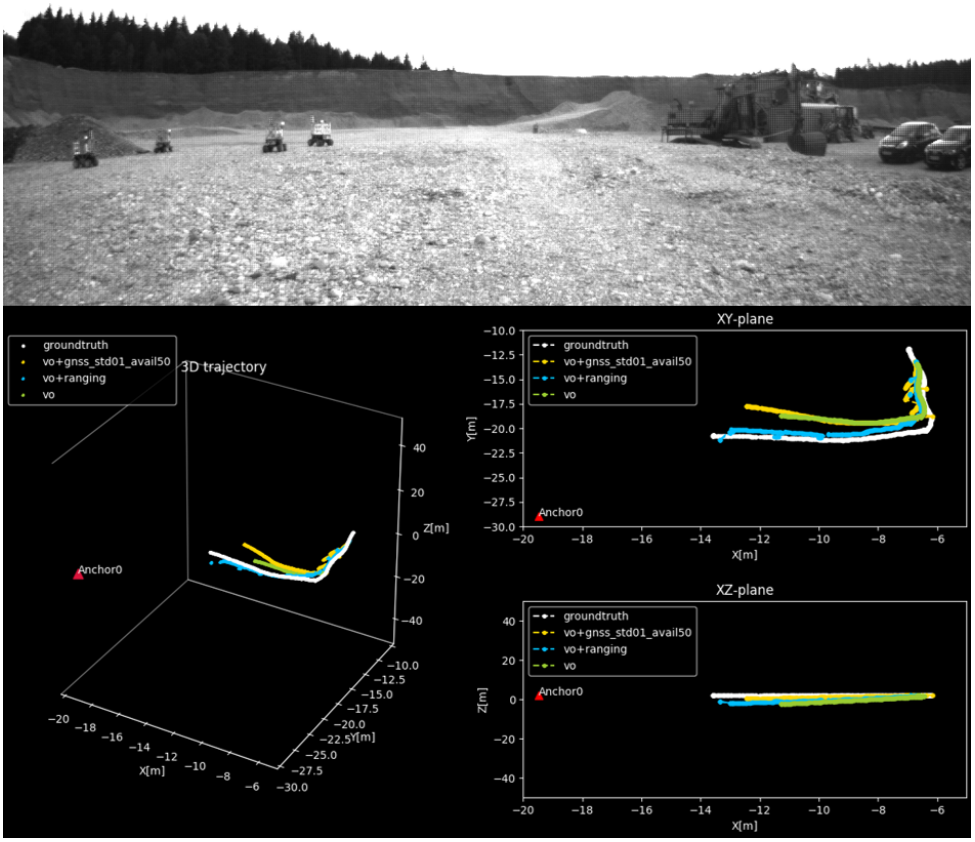


Fig. 1: A potential application of the proposed fusion method. The data are obtained in a featureless rocky environment where accurate GNSS-based positioning (cm-level accuracy) is only available for the first 50% of the entire trajectory, while ranging measurements between the vehicle and the base station (red triangle) are available for the entire time. The green line is the trajectory estimated only with the on-board system (monocular visual odometry), and the skyblue line is the trajectory estimated using the proposed method with ranging measurements. The yellow line is the trajectory estimated with a fusion of odometry and absolute position measurements obtained GNSS-based positioning. This figure qualitatively shows the fusion of odometry with absolute position measurements performs worse than the proposed ranging fusion algorithm if GNSS positioning is only available for a short period of the entire mission, even though it is available with high accuracy.

addition, visual-inertial odometry can only estimate vehicle's poses in a local navigation frame which is initialized differently for every mission.

To mitigate accumulating errors, the loop closing technique can be applied, using the matching features in the current image frame and the map database. However, it continuously requires a large computational load and data storage to detect loop candidates. In addition, mission planning is constrained since a vehicle needs to observe the same sets of features that are already stored in the database for closing a loop. Moreover, without exteroceptive measurements, we can still only estimate the vehicle's poses in a local navigation frame.

To locate a vehicle in a fixed reference frame that is known to users (global frame), vehicle's absolute positions obtained with GNSS can be fused with odometry measurements. Using the additional GNSS observations, accumulating errors in the dead-reckoning estimates can be also eliminated since the GNSS-based estimates are independent on the previous poses. However, there are many cases in which the GNSS-based positioning is unreliable or not available. For example, in urban canyons or indoor environments, we cannot observe a sufficient number of the satellites, and the overall quality of the signal is degraded due to the signal interference or multipath, so GNSS-based positioning becomes untrustworthy or entirely unavailable. In such environments, we can alternatively exploit ranging measurements between a vehicle and one or multiple base stations whose coordinates are known in the global frame. For example, we can use signals of opportunity obtained from radio-frequency networks, such as Wi-Fi signals, mobile networks, or the upcoming 5G cellular network. In addition, low cost ad-hoc ranging devices are available. For instance, ultra-wide band (UWB) modules can provide ranging measurements with a communication range up to few hundred meters at cm-level accuracy.

We propose a graph-based fusion of odometry and ranging measurements to reduce the drift in the trajectory and scale estimates in environments where GNSS-based positioning is unreliable or completely unavailable. Using a public image dataset and synthetically generated ranging measurements, we analyze the system sensitivity to accuracy and data rate of

ranging measurements. Then, we show trajectory estimation results using a real dataset obtained in an outdoor environment with a camera and a UWB module on a rover (see Fig. 1).

RELATED WORKS

To estimate the vehicle's trajectory with respect to a global frame with reduced estimation error, absolute position measurements obtained with GNSS-based positioning can be fused with the on-board measurements (obtained with vision or visual-inertial odometry). In [1], [2], [3], a framework to fuse visual-inertial odometry and GPS positions is proposed, re-formulating the data fusion problem as a frame alignment problem by decoupling locally-referred odometry measurements and globally-referred poses. Although this approach can effectively mitigate the drift of the visual-inertial system, it requires a constantly available and reliable GNSS-based positioning, which is challenging to fulfill in many environments.

In urban areas where not enough number of the satellites is available for positioning, Schreiber *et al.* [4] use pseudorange measurements between a vehicle and satellites, fusing them with stereo visual odometry using a tightly-coupled extended Kalman filter (EKF). However, since pseudoranges are not direct ranging measurements between a receiver (vehicle) and a transmitter (satellite), and include the user clock-offset as an additional unknown, the accuracy of ranging remains at meter- or decimeter-level.

Alternatively, we can obtain ranging measurements from a base station using signals of opportunity, which are wireless signals that are not originally dedicated to navigation purposes. Nikookar *et al.* [5] survey a list of signals of opportunity that can be alternatively used instead of GNSS for ranging and positioning, outlining advantages and disadvantages of using each system for localization purposes. In addition, they give examples of data fusion methods with signals of opportunity and other systems. Tabibiazar *et al.* [6] propose a framework to fuse monocular vision odometry and ranging measurements obtained with signals of opportunity in cellular network systems, by using a particle filter with an adaptive weighting between odometry and ranging measurements.

Instead of signals of opportunity, Perez-Grau *et al.* [7] use radio-based ranging sensors, fusing ranging measurements with on-board stereo visual odometry and inertial data for long-term indoor missions. In the proposed algorithm, the locally-referred poses are estimated by a loosely-coupled EKF with stereo vision and inertial measurements, then distance measurements are added in the update step of the filter. Indoor tests show the reduced positioning errors compared to the stereo vision-only system.

Among radio-based ranging systems, UWB modules are widely used since many low cost products became available. For example, Benini *et al.* [8] propose an EKF-based data fusion of the UWB system with visual-inertial odometry, showing indoor experimental results at cm-level positioning error using quadcopters.

In our previous work [9], a graph-based fusion method was proposed instead of filtering to fuse stereo visual odometry and a single ranging link. The data fusion is formulated as a least-squares estimation minimizing the sum of the squared re-projection and ranging errors. The proposed algorithm is evaluated using the KITTI dataset [10] with synthetic ranging measurements, showing reduced relative pose error (RPE) and absolute trajectory error (ATE) [11], compared to stereo vision-only estimation. In [12], we propose a fusion of monocular visual odometry and ranging measurements also with a graph-based approach, presenting experimental results using a real dataset obtained in a featureless environment (grass field) with a camera and low cost UWB ranging modules. First, we recover the absolute scale in monocular visual odometry measurements, only using ranging links. Then, we reduce accumulating errors in the dead-reckoning process with a graph-based ranging fusion. However, with both [9] and [12], vehicle's poses can only be estimated in a local navigation frame. Furthermore, only post-processing results are presented without an analysis on real-time performance.

Applying a similar tightly-coupled graph-based fusion method, Shi *et al.* [13] propose an exploration and navigation algorithm with a camera and UWB ranging modules. The simulation results with the EuRoC dataset [14] and synthetic ranging measurements show improved estimation accuracy, but they assume that the reference frames for the on-board system and the global frame are the same.

In [15], Wang *et al.* also used a graph-based method to fuse a dense map and IMU with ranging obtained with UWB sensors. They show accurate pose estimation and reconstruction results using real data obtained in indoor environments with a RGBD camera, IMU, and UWB sensors. However, the frame alignment between the on-board and external ranging system is not explained in detail.

In this paper, we propose a graph-based fusion method using on-board visual odometry and ranging data to estimate vehicle's global poses. First, we align the local reference frame with a global frame with a set of globally-referred 3D positions (with e.g. GNSS), which are only available during this initialization process. Afterward, we only use a vision-based system to estimate the vehicle's poses. For correcting the drift inherent to visual odometry, we use additional ranging measurements. To maintain the system complexity and computational load at manageable levels for real-time processing, we use a sliding window, conducting data fusions only with the 10 latest frames.

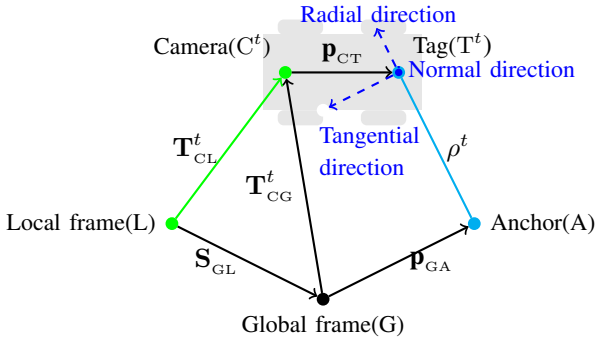


Fig. 2: The system setup with the on-board camera (C^t) and ranging tag module (T^t) at t, and the ranging anchor (A). The local (L) and the global (G) frames are the reference frame of the on-board and external systems, respectively. The relative similarity matrix between the local frame to the global frame is denoted with $\mathbf{S}_{GL}$. The locally-referred camera pose $\mathbf{T}^t_{CL}$ is marked with a green line, and $\mathbf{T}^t_{CG}$ denotes the globally-referred camera pose. The ranging measurement between the tag and anchor ρ^t is marked with a blue line. We assume that the position vector from the camera to the tag $\mathbf{p}_{CT}$ and the anchor position in the global frame $\mathbf{p}_{GA}$ are known. The radial, tangential, and normal directions are additionally defined to analyze the error reduced with ranging measurements.

NOTATIONS AND SYSTEM SETUP

The following is the notation used in this paper:

- ${}_{C}\mathbf{p}^t_{AB}$ A position vector from the origin of the frame A to the origin of the frame B defined with respect to the reference frame C at time t. If the vector is defined with respect to the reference frame A, it is described without a left subscript as $\mathbf{p}^t_{AB}$
- $\mathbf{R}^t_{BA}$ A rotation matrix that converts a vector's reference frame from A to B, i.e. ${}_{B}\mathbf{p}^t_{AB} = \mathbf{R}^t_{BA}\mathbf{p}^t_{AB}$
- $\mathbf{t}^t_{AB}$ A translation vector defined as $\mathbf{t}^t_{AB} = -\mathbf{R}^t_{BA}\mathbf{p}^t_{AB} = \mathbf{p}^t_{BA}$
- $\mathbf{T}^t_{BA}$ A pose matrix from A to B at time t described as a matrix in the Lie group: $\mathbf{T}^t_{BA} = \begin{bmatrix} \mathbf{R}^t_{BA} & \mathbf{t}^t_{AB} \\ \mathbf{0} & 1 \end{bmatrix} \in \mathrm{SE}(3)$
- $\boldsymbol{\delta}^t_{BA}$ A 6×1 vector $[\boldsymbol{\omega}^T, \boldsymbol{u}^T]^T$ belonging to the se(3) group in Lie algebra. Vector $\boldsymbol{\delta}^t_{BA}$ is associated to matrix $\mathbf{T}^t_{BA}$ in SE(3) through relations (1) and (2)
- $\mathbf{S}_{BA}$ A similarity matrix from A to B described as a matrix in the Lie group, including an additional scale factor λ_{BA} from A to B: $\mathbf{S}_{BA} = \begin{bmatrix} \mathbf{R}^t_{BA} & \mathbf{t}^t_{AB} \\ \mathbf{0} & \lambda^{-1}_{BA} \end{bmatrix} \in \mathrm{Sim}(3)$

To describe the camera motion, we use the topological group SE(3). A matrix $\mathbf{T}$ in SE(3) describing a rotation $\mathbf{R}$ and a translation $\mathbf{t}$ can be mapped to a 6×1 vector $[\boldsymbol{\omega}^T, \boldsymbol{u}^T]^T$ in se(3) encoding the rotation and the translation as [16]:

$$\boldsymbol{\omega} = \frac{\theta}{2\sin\theta} \begin{pmatrix} \mathrm{R}_{21} - \mathrm{R}_{12} \\ \mathrm{R}_{02} - \mathrm{R}_{20} \\ \mathrm{R}_{10} - \mathrm{R}_{01} \end{pmatrix}, \tag{1}$$

$$\boldsymbol{u} = \left(\mathbf{I} - \frac{1}{2}\boldsymbol{\omega}_{\mathrm{x}} + \frac{1}{\theta^2}\left(1 - \frac{\sin\theta \cdot \theta}{2(1-\cos\theta)}\right)\boldsymbol{\omega}^2_{\mathrm{x}} \right) \mathbf{t}, \tag{2}$$

where θ is

$$\theta = \arccos\left(\frac{tr(\mathbf{R}) - 1}{2}\right)$$

and $\boldsymbol{\omega}_{\mathrm{x}}$ is the skew-symmetric matrix associated to $\boldsymbol{\omega}$.

In Fig. 2, the system setup is shown with on-board sensors – a monocular camera (C) and a ranging tag module (T) – and a base station as a ranging anchor (A), as well as the two reference frames. The local frame (L) is the reference frame for visual odometry, which is randomly initialized at each mission, coinciding with the initial pose of the camera frame. The global frame (G) is the reference frame for GNSS-based positioning and the anchor position, which is known to the users. In addition, we define the radial direction at each vehicle's position as the unit vector from the anchor to the tag ${}_{G}\mathbf{u}^t_{AT} = \frac{{}_{G}\mathbf{p}^t_{AT}}{||{}_{G}\mathbf{p}^t_{AT}||}$, and the normal direction as the normal vector of the plane defined with the radial unit vector and the position vector from the anchor to the global frame $\frac{{}_{G}\mathbf{u}^t_{AT} \times (-\mathbf{p}_{GA})}{||\mathbf{p}_{GA}||}$. The tangential direction is defined with the right-hand rule.

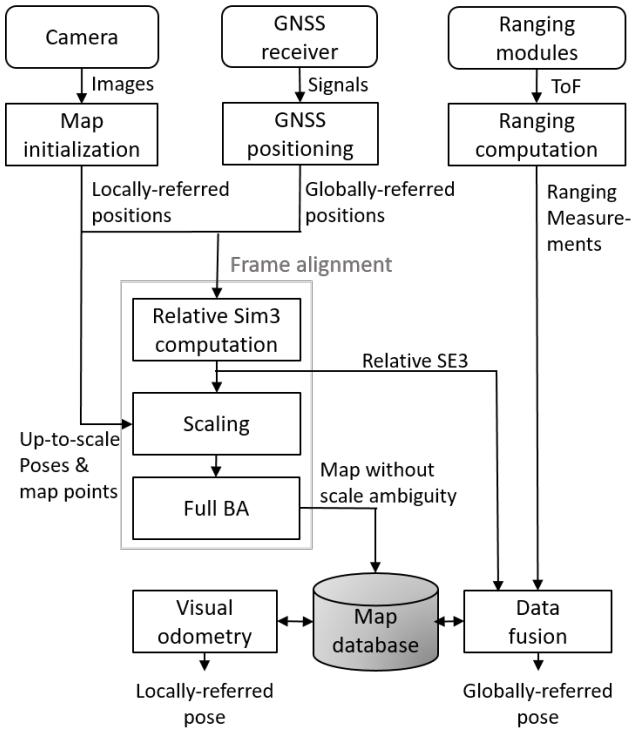


Fig. 3: The flowchart of a localization system with the proposed ranging fusion method. Note that GNSS positioning is only available at the beginning of the process for aligning the local and global frames.

VISUAL-RANGING LOCALIZATION

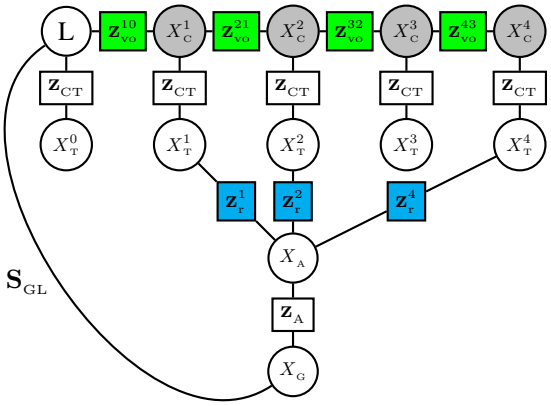
The flowchart of the proposed algorithm is illustrated in Fig. 3. First, vehicle's poses and map points are initialized up to a scale factor in the local frame using monocular visual odometry. The visual odometry part is implemented based on the *Tracking* and *Local Mapping* of ORB-SLAM [17]. With these locally-referred positions obtained with visual odometry and additional absolute globally-referred position measurements obtained with GNSS, we estimate the relative scale and pose between the local and global frames to align the two frames, by applying a deterministic point matching algorithm [18]. Then, we multiply the estimated scale to the up-to-scale locally-referred positions and map points, and execute a full bundle adjustment to minimize the sum of the squared re-projection errors between the adjusted vehicle's poses and map points. After the frame alignment, we use a ranging link between an on-board ranging module and a base station (maximum one link is available at each timestamp) to correct drifts in vehicle pose estimates, assuming that GNSS positioning is no longer available after the frame alignment step. To fuse odometry and ranging measurements, we use a graph-based approach. The updated poses in this fusion process are shared with visual odometry, so the odometry process can also use the corrected estimates.

Frame alignment

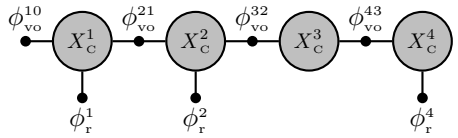
To align the local frame used for visual odometry and the global frame used for GNSS-based positioning and ranging system, we apply a deterministic point matching method proposed in [18]. This method computes the relative similarity matrix between the two frames (7DoF) using the correspondences of the positions in the local frame obtained with visual odometry and the absolute positions in the global frame obtained with GNSS positioning. Theoretically, the relative similarity transformation matrix can be recovered when we have more than three point matches, but we collect at least 10 correspondences, considering the errors in the measurements.

Data fusion of odometry and ranging measurements

1) Bayes network formulation: To mitigate accumulating errors in pose estimates, we conduct a graph-based data fusion of odometry and ranging measurements. Fig. 4a shows the formulated Bayesian network, including nodes of the camera pose X_{C}^{i} and on-board ranging tag modules' position in the camera frame X_{T}^{i} at each timestamp t_i. In this network, the origin of the visual odometry process is denoted with the local frame node (L), and the similarity matrix from the local to the global frame (node X_{G}) is denoted as $\mathbf{S}_{\mathrm{GL}}$. In addition, the ranging anchor position in the global frame is illustrated with a node X_{A}.



(a) Bayes network formulation with the camera poses (node $X_{\mathrm{C}}^{i} = \mathbf{T}_{\mathrm{CL}}^{t_i}$) and on-board ranging tag modules' positions (X_{T}^{i}). The local frame (L=X_{C}^{0}) is the origin of visual odometry, and the similarity matrix from the local to the global frame (node X_{G}) is denoted as $\mathbf{S}_{\mathrm{GL}}$. The ranging anchor position in the global frame (node X_{A}) is denoted as $\mathbf{p}_{\mathrm{GA}}$. The green squares in the edges between the camera poses ($\mathbf{z}_{\mathrm{vo}}^{ji}$) are visual odometry measurements from C^{t_i} to C^{t_j}, and the blue ones ($\mathbf{z}_{\mathrm{r}}^{i}$) are ranging measurements at t_i.



(b) The simplified factor graph with camera poses (X_{C}^{i}) as state factors. ϕ_{vo}^{ji} are factors that are related with odometry measurements, and ϕ_{r}^{i} are the factors related with ranging measurements.

Fig. 4: The Bayes network and simplified factor graph formulations for the proposed data fusion.

2) Factorization: Assuming that the globally-referred anchor position $\mathbf{p}_{\mathrm{GA}}$ and the position vector from the on-board camera to the ranging tag $\mathbf{p}_{\mathrm{CT}}$ are known, we can simplify the network to the graph as shown in Fig. 4b by only retaining the actual unknowns, i.e. camera poses (6DoF) in the local frame:

$$\mathbf{X} = \left[X_{\mathrm{C}}^{1}, X_{\mathrm{C}}^{2}, \ldots, X_{\mathrm{C}}^{N}\right] = \left[\mathbf{T}_{\mathrm{CL}}^{t_1}, \mathbf{T}_{\mathrm{CL}}^{t_2}, \ldots, \mathbf{T}_{\mathrm{CL}}^{t_N}\right].$$

The factors between the camera nodes ϕ_{vo}^{ji} are the likelihood of the states $\mathbf{T}_{\mathrm{CL}}^{t_i}$ and $\mathbf{T}_{\mathrm{CL}}^{t_j}$ given the odometry measurements $\mathbf{z}_{\mathrm{vo}}^{ji} = \mathbf{T}_{\mathrm{C}^{\mathrm{t_j}}\mathrm{C}^{\mathrm{t_i}}}$. Assuming that noise in odometry measurements is Gaussian,

$$\phi_{\mathrm{vo}}^{ji} = l(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}; \mathbf{z}_{\mathrm{vo}}) \propto \exp\{-\frac{1}{2}||\mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}, \mathbf{z}_{\mathrm{vo}}^{ji})||^2\}.$$

To compute the odometry error $\mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}, \mathbf{z}_{\mathrm{vo}}^{ji})$, first, we need to compute the error matrix in SE(3) as

$$\mathbf{E}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}, \mathbf{z}_{\mathrm{vo}}^{ji}) = (\mathbf{T}_{\mathrm{CL}}^{t_j})^{-1}\mathbf{T}_{\mathrm{C}^{\mathrm{t_j}}\mathrm{C}^{\mathrm{t_i}}}\mathbf{T}_{\mathrm{CL}}^{t_i}. \tag{3}$$

Then, we can convert this error matrix to $\mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}, \mathbf{z}_{\mathrm{vo}}^{ji})$ in se(3) of the Lie algebra using (1) and (2).

When ranging is available, camera nodes have additional factors ϕ_{r}^{i}, which is the likelihood of the states $\mathbf{T}_{\mathrm{CL}}^{t_i}$ given the ranging measurement $\mathbf{z}_{\mathrm{r}}^{i}$. With a Gaussian noise assumption,

$$\phi_{\mathrm{r}}^{i} = l(\mathbf{T}_{\mathrm{CL}}^{t_i}; \mathbf{z}_{\mathrm{r}}^{i}) \propto \exp\{-\frac{1}{2}||\mathbf{e}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t}, \mathbf{z}_{\mathrm{r}}^{i})||^2\}.$$

Since ranging measurements are scalar values, the errors can be easily defined as the difference between the ranging measurement $\mathbf{z}_{\mathrm{r}}^{i}$ and the value computed with the model:

$$\mathbf{e}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t}, \mathbf{z}_{\mathrm{r}}^{i}) = \mathbf{z}_{\mathrm{r}}^{i} - \mathbf{h}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t}), \tag{4}$$

where $\mathbf{h}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t})$ is a ranging measurement model function:

$$\mathbf{h}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t}) = ||{}_{\mathrm{G}}\mathbf{p}_{\mathrm{AT}}^{t}|| = ||-\mathbf{p}_{\mathrm{GA}} + \mathbf{t}_{\mathrm{LG}} - \mathbf{R}_{\mathrm{GL}}\mathbf{R}_{\mathrm{LC}}^{t}\left(\mathbf{t}_{\mathrm{LC}}^{t} - \mathbf{p}_{\mathrm{CT}}\right)||. \tag{5}$$

3) Least-squares problem formulation: Assuming Gaussian noise on the measurements, the optimal solution maximizing the multiplication of all factors is equivalent to the least-squares solution that minimizes the squared errors between the measurements and the values computed with the following model functions:

$$\begin{aligned}\hat{\mathbf{X}} &= \underset{\mathbf{X}}{\operatorname{argmax}} \prod_{i,j} \phi_{\mathrm{vo}}^{ji} \prod_{i} \phi_{\rho}^{i} \\ &= \underset{\mathbf{X}}{\operatorname{argmin}} \sum_{i,j} (\mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}, \mathbf{z}_{\mathrm{vo}}^{ji}))^T \mathbf{\Omega}_{\mathrm{vo}} \mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j}, \mathbf{z}_{\mathrm{vo}}^{ji}) + \sum_{l} (\mathbf{e}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t_l}, \mathbf{z}_{\mathrm{r}}^{l}))^T \mathbf{\Omega}_{\mathrm{r}} \mathbf{e}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t_l}, \mathbf{z}_{\mathrm{r}}^{l}),\end{aligned} \tag{6}$$

where $\mathbf{\Omega}$ denotes the inverse of the measurement covariance matrix.

4) Iterative optimization: We apply the Levenberg-Marquardt algorithm to solve the formulated least-squares problem (6), which iteratively updates with $\Delta\hat{\mathbf{X}}^*$ computed as

$$(\mathbf{H} + \lambda \mathbf{I}) \Delta\hat{\mathbf{X}}^* = -\mathbf{b}, \tag{7}$$

where $\mathbf{H} = \sum_k \mathbf{J}_{\mathrm{k}}^T \mathbf{\Omega}_{\mathrm{k}} \mathbf{J}_{\mathrm{k}}$ and $\mathbf{b} = \sum_k \mathbf{J}_{\mathrm{k}}^T \mathbf{\Omega}_{\mathrm{k}} \mathbf{e}$ with Jacobian matrices $\mathbf{J}_{\mathrm{k}}$.

The Jacobian matrix of the odometry error function is

$$\mathbf{J}_{\mathrm{vo}}^{t_i,t_j} = \left[\mathbf{0} \dots \frac{\partial \mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j})}{\partial \mathbf{T}_{\mathrm{CL}}^{t_i}} \dots \frac{\partial \mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j})}{\partial \mathbf{T}_{\mathrm{CL}}^{t_j}} \dots \mathbf{0} \right], \tag{8}$$

where the partial deviations of the error function are

$$\frac{\partial \mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j})}{\partial \mathbf{T}_{\mathrm{CL}}^{t_i}} = \mathrm{adj}\left((\mathbf{T}_{\mathrm{CL}}^{t_j})^{-1} \mathbf{T}_{\mathrm{C^{t_j} C^{t_i}}} \right) \tag{9}$$

$$\frac{\partial \mathbf{e}_{\mathrm{vo}}(\mathbf{T}_{\mathrm{CL}}^{t_i}, \mathbf{T}_{\mathrm{CL}}^{t_j})}{\partial \mathbf{T}_{\mathrm{CL}}^{t_j}} = \mathrm{adj}\left(-(\mathbf{T}_{\mathrm{CL}}^{t_i})^{-1} (\mathbf{T}_{\mathrm{C^{t_j} C^{t_i}}})^{-1} \right), \tag{10}$$

in which $\mathrm{adj}(\mathbf{T})$ denotes an adjoint of a pose matrix $\mathbf{T}$ in SE(3) [16].

For a ranging error function, the Jacobian matrix is

$$\mathbf{J}_{\rho}^{t} = \left[\mathbf{0} \dots \frac{\partial \mathbf{e}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t_i})}{\partial \mathbf{T}_{\mathrm{CL}}^{t_i}} \dots \mathbf{0} \right], \tag{11}$$

where the error function's partial deviation is

$$\frac{\partial \mathbf{e}_{\mathrm{r}}(\mathbf{T}_{\mathrm{CL}}^{t_i})}{\partial \mathbf{T}_{\mathrm{CL}}^{t_i}} = -\frac{1}{d} (0, 0, 0, x, y, z)^T, \tag{12}$$

in which ${}_{\mathrm{C}}\mathbf{p}_{\mathrm{TA}}^{t} = (x, y, z)^T$ and $d = ||{}_{\mathrm{C}}\mathbf{p}_{\mathrm{TA}}^{t}||$.

For iterative computations, an available tool, *g2o* [19], is used.

EVALUATION

Simulations with the KITTI dataset

We analyze the proposed system's sensitivity to accuracy and data rate of ranging measurements, using the image sequence 07 of the KITTI dataset [10]. The ranging measurements are synthetically generated with zero-mean Gaussian noise. First, we conduct visual odometry only with the frame alignment step at initialization. Since we only use dead-reckoning process to estimate the trajectory, the vehicle's poses and scale drift as shown in Fig. 5 (the green line).

Then, we conduct the proposed data fusion using visual odometry and ranging measurements. Table I shows the Root Mean Square Error (RMSE) of the trajectory estimates in the radial, tangential, and normal directions, as well as for the position vectors. The RMSE values are reported for various settings of the ranging noise level and availability of ranging measurements. The former is defined as a zero-mean Gaussian noise, and its standard deviation (std) is given. The data rate of ranging measurements are given as ratio of the frame rate: 'nframes' denotes a ranging data rate equal to 1/n-th of the frame rate (one ranging available every n frames), whereas 'every' denotes a ranging data rate equal to the frame rate (ranging is available for every image frame).

First, we show the simulation results with ranges available at every image without noise: this is reported in the line 'no noise/every' in Table I. As expected, the proposed fusion algorithm significantly reduces the drift in the radial direction. The errors in the tangential and normal directions are not drastically reduced as the errors in the radial direction.

Then, we analyze the system sensitivity to the noise level using ranging measurements with different standard deviations and the same data rate of the image's: 'std=xm/every' in Table I. This table shows that the RMSE in the radial direction

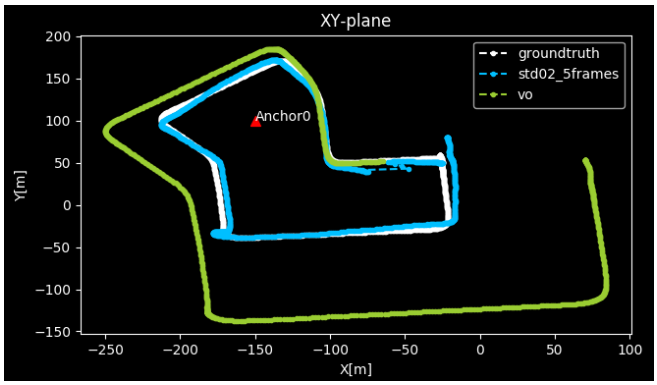


Fig. 5: Trajectories in the global frame's XY-plane: Ground truth (white), the trajectory estimated with visual odometry (green), and the trajectory estimated using the proposed data fusion method (skyblue) with ranging measurements that have Gaussian noise with zero mean and standard deviation 0.2m and 1/5 of the image's data rate (available for every five images). The red triangle is the anchor position.

TABLE I: The RMSE of the trajectory estimates in radial, tangential, normal directions, and position vectors with various noise levels and data rates of the ranging measurements [m]. 'std' denotes the standard deviation of Gaussian noise, and 'no noise' is written when std is 0m. 'nframes' denotes that ranging is available every n image frames (1/n data rate of images'), and 'every' is indicated when ranging is available at all image frames.

Settings	RMSE [m]			
	radial	tangential	normal	position
visual odometry	76.44	11.86	3.02	77.41
no noise/every	0.56	7.54	2.47	7.95
std=0.2m/every	0.93	8.59	3.03	9.16
std=0.5m/every	1.1	9.03	3.28	9.67
std=1.0m/every	1.12	9.03	3.28	9.67
std=1.5m/every	1.08	9.06	3.18	9.67
std=2.0m/every	1.11	8.65	3.0	9.22
no noise/2frames	0.67	7.33	2.42	7.75
no noise/5frames	0.85	7.26	2.53	7.73
no noise/10frames	1.03	7.47	2.55	7.96
no noise/20frames	1.79	7.49	2.6	8.13
no noise/30frames	2.45	7.92	2.62	8.7
std=0.2m/5frames	0.88	7.92	2.75	8.43

does not increase significantly as the noise level increases. The performance is robust to the noise level since ranging is assumed to be available at each frame, thus providing a large number of measurements.

In addition, we analyze the system sensitivity to the data rate, using ranging measurements with different data rates without noise, see line 'no noise/nframes' in Table I. In this table, we can observe that the RMSE in the radial direction increases more significantly as the data rate decreases, indicating that the proposed system is more sensitive to the data rate than to the noise level.

Lastly, we conduct a simulation using ranging measurements with a 1/5 rate and Gaussian noise with standard deviation equal to 0.2m. The blue line in Fig. 5 shows to what extent the proposed data fusion method can reduce the drift inhrent to visual odometry. Line 'std=0.2m/5frames' in Table I quantitatively assess the improvement in terms of RMSE when ranging measurements are integrated.

Experiments with a rover system

To analyze the proposed algorithm with real images and ranging data, we acquired a dataset in a gravel pit with a rover system developed by the Institute of Communications and Navigation of the German Aerospace Center (DLR). For ground truth, we use a real time kinematic (RTK) system that uses GNSS signals and inertial data. As shown in Fig. 6, a camera (Bumblebee2, pointgrey, 20fps) and UWB ranging module (Decawave, data rate $\sim$ 7.97Hz, $\sigma \sim 0.15\text{m}$) are mounted on a mobile rover, and another ranging module (anchor) is mounted on a static system. Both rovers have multisensor modules for the RTK system. The mobile rover traveled about 19 meters for about 3 minutes, and an image sample is presented in Fig. 1.

With the obtained real images and ranging measurements, we conducted the proposed data fusion with different sliding window sizes (the number of frames included in the formulated factor graph) whenever 10 new frames are stored in the database. The inverse covariance of ranging $\mathbf{\Omega}_{\text{r}}$ in the cost function (6) is set as $(1/0.15)^2$, using an experimentally obtained standard deviation of 0.15m. GNSS positioning was available only for the first 20 image frames, used for aligning the local and global frames during the initialization phase. To process the data, we use a Lenovo ThinkPad T460s with an

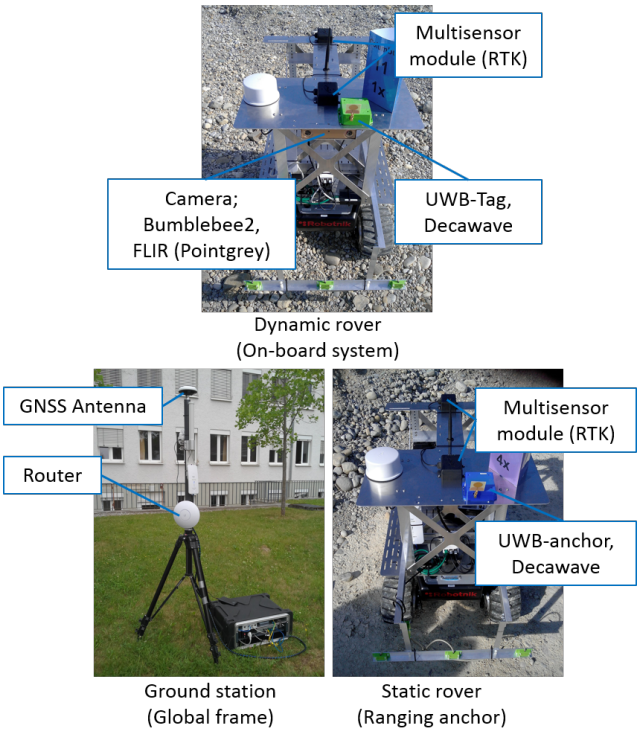


Fig. 6: The rover system developed by the Institute of Communications and Navigation of the German Aerospace Center (DLR). The mobile rover has a camera and ranging tag as on-board sensors, and multisensor modules for the RTK system. A static rover has a ranging anchor and multisensor modules for the RTK system. The ground station has a router for the network system and GNSS antenna for the RTK system.

TABLE II: The average, minimum, and maximum of the per-frame processing time [ms] and the RMSE of trajectory estimates in the radial, tangential, and normal directions, as well as the position vectors [m], with different sizes of the sliding window.

Settings	Proc. time [ms]			RMSE [m]			
	average	min	max	radial	tangential	normal	position
visual odometry	255	90	570	1.25	1.3	1.37	2.27
window size=10	366	130	770	1.29	1.24	1.15	2.13
window size=20	375	110	1040	0.84	1.26	1.01	1.82
window size=50	371	120	790	0.36	1.11	1.27	1.72
window size=100	381	120	870	0.48	0.96	1.31	1.70

Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz. Note that the proposed system is implemented as a multi-threading process, conducting visual odometry and the fusion process simultaneously on different threads.

Fig. 1 qualitatively shows that the trajectory estimated using the proposed method with the sliding window size 50 is closer to the ground truth compared to the one estimated with visual odometry.

Table II shows the average, minimum, and maximum of the processing time (CPU user time) per single frame, as well as the RMSE of the trajectory estimated with visual odometry and the proposed algorithm. The latter was obtained with various window sizes: 10, 20, 50, and 100. In this table, the proposed method shows improved results in positioning accuracy for all the test cases, without increasing the processing time significantly. The first test (window size=10) shows a slightly decreased accuracy in the radial component, but achieves nevertheless improved overall estimation. Comparing the 'window size=50' and 'window size=100' cases, we can observe that the RMSE in the radial direction is bigger, and the processing time is longer, when the window size case is larger. This result indicates that the positioning accuracy improves with a larger window size.

Moreover, we compare the proposed ranging fusion method with a fusion of visual odometry and absolute position measurements when GNSS-based positioning is available only for the first 50% of the entire trajectory (yellow line in Fig. 1). As shown in Fig. 1, the positions in the XY-plane estimated with the proposed ranging fusion method is closer to the ground truth compared to the one estimated with the GNSS fusion method when GNSS-based positioning is unavailable throughout the mission.

In addition, Fig. 7 shows errors in the radial direction of the trajectory estimated with the proposed method (the skyblue line), the one obtained with the fusion of visual odometry and absolute position measurements (the yellow line), and the

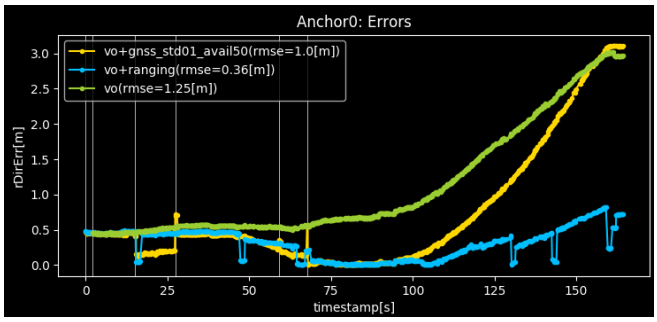


Fig. 7: The errors in the radial direction [m] of the trajectories estimated using visual odometry (green), the proposed ranging fusion algorithm (skyblue), and the fusion of visual odometry and absolute position measurements obtained with GNSS-based positioning (yellow). GNSS-based positioning is available only first 50% of the entire trajectory, and the standard deviation of measurement noise is 0.1m (for each direction). The vertical white lines indicate the timestamps that the fusion processes of odometry and 3D positions are conducted.

one estimated only with visual odometry (the green line). The vertical white lines indicate the timestamps when the GNSS data is used. After 3D position fusion is completed (yellow line, after the last vertical white line), the error starts to diverge, while the proposed ranging fusion algorithm reduces the drifts in the radial direction with ranging measurements that are available for the entire mission.

CONCLUSION

We proposed a data fusion of on-board visual odometry and ranging measurements between a vehicle and base station to mitigate accumulating pose errors in environments where GNSS-based positioning is untrustworthy or entirely unavailable after the initialization phase of a mission. With simulations using an available image dataset and synthetically generated ranging measurements, we analyze the system sensitivity to ranging noise and availability. Moreover, we evaluate the proposed algorithm using real data obtained in an outdoor environment, showing real-time performance and system complexity in terms of processing time.

REFERENCES

[1] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart, and M. Chli, "Gomsf: Graph-optimization based multi-sensor fusion for robust uav pose estimation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1421–1428.

[2] J. Surber, L. Teixeira, and M. Chli, "Robust visual-inertial localization with weak gps priors for repetitive uav flights," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 6300–6306.

[3] H. Oleynikova, M. Burri, S. Lynen, and R. Siegwart, "Real-time visual-inertial localization for aerial and ground robots," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 3079–3085.

[4] M. Schreiber, H. Königshof, A.-M. Hellmund, and C. Stiller, "Vehicle localization with tightly coupled gnss and visual odometry," in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2016, pp. 858–863.

[5] H. Nikookar and P. Oonincx, "An introduction to radio locationing with signals of opportunity," *Journal of Communication, Navigation, Sensing and Services (CONASENSE)*, vol. 2016, no. 1, pp. 1–10, 2016.

[6] A. Tabibiazar and O. Basir, "Radio-visual signal fusion for localization in cellular networks," in *2010 IEEE Conference on Multisensor Fusion and Integration*. IEEE, 2010, pp. 150–155.

[7] F. Perez-Grau, F. Fabresse, F. Caballero, A. Viguria, and A. Ollero, "Long-term aerial robot localization based on visual odometry and radio-based ranging," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2016, pp. 608–614.

[8] A. Benini, A. Mancini, and S. Longhi, "An imu/uwb/vision-based extended kalman filter for mini-uav localization in indoor environment using 802.15. 4a wireless sensor network," *Journal of Intelligent & Robotic Systems*, vol. 70, no. 1-4, pp. 461–476, 2013.

[9] Y.-H. Lee, C. Zhu, G. Giorgi, and C. Guenther, "Stereo vision-based simultaneous localization and mapping with ranging aid," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, 2018, pp. 404–409.

[10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[11] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 573–580.

[12] Y.-H. Lee, C. Zhu, G. Giorgi, and C. Günther, "Fusion of monocular vision and radio-based ranging for global scale estimation and drift mitigation," *arXiv preprint arXiv:1810.01346*, 2018.

[13] Q. Shi, X. Cui, W. Li, Y. Xia, and M. Lu, "Visual-uwb navigation system for unknown environments," 10 2018, pp. 3111–3121.

[14] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[15] C. Wang, H. Zhang, T.-M. Nguyen, and L. Xie, "Ultra-wideband aided fast localization and mapping system," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1602–1609.

[16] E. Eade, "Lie groups for 2d and 3d transformations."

[17] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[18] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Josa a*, vol. 4, no. 4, pp. 629–642, 1987.

[19] G. Grisetti, R. Kümmerle, H. Strasdat, and K. Konolige, "g2o: A general framework for (hyper) graph optimization," *Tech. Rep.*, 2011.