Cooperative Navigation and Guidance of a Micro-Scale Aerial Vehicle by an Accompanying UAV using 3D LiDAR Relative Localization

Václav Pritzl*, Matouš Vrba*, Petr Štěpán*, and Martin Saska*

***Abstract*— A novel approach for cooperative navigation and guidance of a micro-scale aerial vehicle by an accompanying Unmanned Aerial Vehicle (UAV) using 3D Light Detection and Ranging (LiDAR) relative localization is proposed in this paper. The use of 3D LiDARs represents a reliable way of environment perception and robust UAV self-localization in Global Navigation Satellite System (GNSS)-denied environments. However, 3D LiDARs are relatively heavy and they need to be carried by large UAV platforms. On the contrary, visual cameras are cheap, light-weight, and therefore ideal for small UAVs. However, visual self-localization methods suffer from loss of precision in texture-less environments, scale unobservability during certain maneuvers, and long-term drift with respect to the global frame of reference. Nevertheless, a micro-scale camera-equipped UAV is ideal for complementing a 3D LiDAR-equipped UAV as it can reach places inaccessible to a large UAV platform. To gain the advantages of both navigation approaches, we propose a cooperative navigation and guidance architecture utilizing a large LiDAR-equipped UAV accompanied by a small secondary UAV carrying a significantly lighter monocular camera. The primary UAV is localized by a robust LiDAR Simultaneous Localization and Mapping (SLAM) algorithm, while the secondary UAV utilizes a Visual-Inertial Odometry (VIO) approach with lower precision and reliability. The LiDAR data are used for markerless relative localization between the UAVs to enable precise guidance of the secondary UAV in the frame of reference of the LiDAR SLAM. The performance of the proposed approach has been extensively verified in simulations and real-world experiments with the algorithms running onboard the UAVs with no external localization infrastructure.**

Multimedia Attachment

`http://mrs.felk.cvut.cz/icuas2022-coop`

I. Introduction

The research of autonomous UAVs operating in indoor environments has recently experienced a tremendous surge in popularity. Possibilities for the employment of autonomous UAVs are being explored, e.g., in firefighting [1], [2], Search and Rescue (SAR) [3], [4], [5], [6], documentation of historical monuments [7], [8], inspection tasks, etc. As there is generally no GNSS or external localization system available in real-world indoor environments, the UAVs need to rely on their exteroceptive sensors to localize themselves and perceive the environment. Multiple options of the UAV sensory payload exist, each with its unique advantages.

This work was supported by CTU grant no SGS20/174/OHK3/3T/13, by the Technology Innovation Institute - Sole Proprietorship LLC, UAE, under the Research Project Contract No. TII/ATM/2032/2020, by project no. DG18P02OVV069 in program NAKI II, and by The Ministry of Education of the Czech Republic by OP VVV funded project CZ.02.1.01/0.0/0.0/16 019/0000765 "Research Center for Informatics".

*The authors are with the Multi-robot Systems Group, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic {vaclav.pritzl, matous.vrba, petr.stepan, martin.saska}@fel.cvut.cz

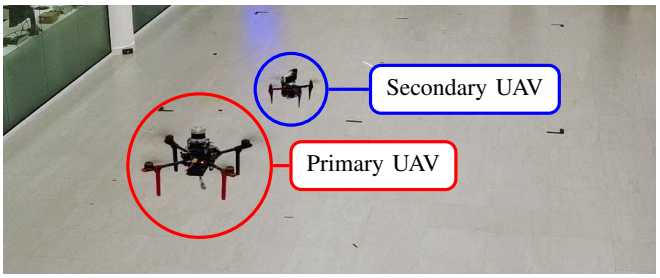

Fig. 1: The large primary UAV equipped with a 3D LiDAR is navigating a smaller secondary UAV carrying a monocular visual camera.

One of the most reliable choices of sensors are 3D LiDARs. LiDARs are precise, long-range, and exhibit good performance even under no illumination and low-visibility conditions [9]. However, they come with the cost of relatively large weight and size. Because of their weight, their use requires large UAV platforms to transport them, and such large platforms may not be able to fly through all the tight spaces present in real-world indoor environments. Furthermore, in many real-world applications it is advantageous to deploy a team of UAVs to, e.g., cover a larger area faster, but the high cost of 3D LiDARs significantly increases the overall cost of such a UAV fleet.

On the contrary, visual cameras are cheap, small, light-weight, with low power consumption. A visually-localized UAV can be very small and capable of flight through very narrow gaps. The overall price of such a UAV is much lower than the price of a large 3D LiDAR-equipped UAV. The use of visual self-localization also has its disadvantages, as it loses precision in texture-less environments and its long-term drift is larger than in the case of 3D LiDARs. Long-term flights then exhibit significant error in global positioning, which is a problem in the case of SAR missions as demonstrated in the DARPA Subterranean challenge where precise global positioning was required. Vision-aided Inertial Navigation Systems (VINS) also suffer from scale unobservability during constant-acceleration movement [10]. Additionally, 3D models obtained with a monocular camera using structure from motion approaches cannot compete with the accuracy and quality of a map constructed using a 3D LiDAR, which is crucial for mapping and planning in an unknown environment.

In this paper, we propose a cooperative navigation approach synergizing the advantages of both types of sensors. We devise a cooperative navigation architecture for a heterogeneous team of autonomous UAVs (see Fig. 1). The primary UAV is equipped with a 3D LiDAR and utilizes a LiDAR SLAM algorithm to localize itself within the surrounding environment. The secondary UAV carries a monocular visual camera and uses a VIO approach for its self-localization. The 3D LiDAR data obtained by the primary UAV are then also used for markerless relative localization of the secondary UAV. The relative localization enables the primary UAV to guide the secondary UAV in the frame of reference of its more accurate and reliable LiDAR SLAM algorithm. This approach enables building a team of robots in which one or several larger 3D LiDAR-equipped UAVs are guiding one or several micro-scale camera-equipped UAVs. These small UAVs may explore the environment inside and behind gaps too narrow for the larger UAV. Another strategy using the proposed method can be relying on a larger 3D LiDAR-equipped Unmanned Ground Vehicle (UGV) or UAV carrying the micro-scale UAVs and releasing them when needed. In both strategies, the 3D LiDAR-carrying vehicle ensures global localization of the entire team and reliable short-range deployment of smaller camera-equipped scouts.

In this work, we focus mainly on the design and evaluation of the cooperative navigation architecture itself, which enables guidance of the secondary UAV with unreliable self-localization using the accurate and reliable frame of reference of the primary UAV. It is worth mentioning that this approach also enables to utilize the 3D LiDAR information about the surrounding obstacles for guiding the secondary UAV, which has no internal knowledge about the obstacles. Such obstacle avoidance is beyond the scope of this paper.

A. Related work

1) UAV cooperative navigation: Cooperative navigation approaches utilizing pairs of robots have been proposed for improving navigation performance in outdoor GNSS-challenging environments. A cooperative navigation approach for multiple UAVs was proposed in [11] and [12]. One or more UAVs (called "fathers" or "deputies") are assumed to have reliable GNSS-based localization. The other UAV (called "son" or "chief") receives GNSS positioning information from the fathers and fuses it with relative position measurements and its own navigation state to improve its localization accuracy. Visual tracking is used to obtain relative position measurements between the UAVs.

A similar approach is to employ a UGV to improve UAV navigation in GNSS-challenging environments. In [13], a cooperative localization approach for a UAV-UGV pair was proposed. Ultra-wideband (UWB) localization was used to obtain the UAV position relative to the UGV and the obtained information was used to improve UAV localization accuracy. In [14], a UGV was employed as a mobile differential GNSS reference station, improving UAV localization accuracy. Similarly in [15], the authors focused on decentralized cooperative localization of a UAV in GNSS-challenging environment and a UGV capable of relative sensing.

In contrast to these works, we focus on the problem of cooperative navigation in GNSS-denied environments and we use a single sensor carried by the primary UAV for both navigation of the primary UAV and relative localization between the UAV pair.

2) VIO precision: We employ a monocular VIO method for self-localization and closed-loop stabilization of the secondary UAV in its local VIO frame. VIO methods suffer from several problems that may cause loss of precision. As shown in [16], the VIO has four unobservable directions under general movement, corresponding to the three Degrees of Freedom (DOFs) of global translation and the yaw orientation. This results in a long-term drift of the VIO estimates with respect to the global frame of reference. In [10], it was shown that VIO methods suffer from scale unobservability given constant or zero acceleration of the camera-Inertial Measurement Unit (IMU) pair and from global orientation unobservability given no rotational motion. In [17], the authors proposed incorporating wheel odometry measurements into the VIO estimator of a UGV and utilizing an assumption of movement on a planar surface to remove the unobservability and improve localization accuracy. The authors of [18] provided a detailed analysis of aided inertial navigation systems and arrived at a similar conclusion about the VIO observability. Although the rigid extrinsic transformation between the IMU and the camera allows us to observe the scale of the system, in practice the scale is still very close to unobservable at constant acceleration. This is often the case when exploring constrained environments where the UAV needs to fly very slowly, almost hovering. Moreover, the VIO can lose precision due to low amount of texture in the environment and due to feature degeneracy in some cases [19]. In this paper, we expect these VIO imprecisions to manifest mainly as long-term drift and as incorrect scale of the local VIO frame of reference. This leads to the UAV position control pipeline undershooting or overshooting the target position.

3) Relative localization approaches: Several different options for relative localization between UAVs exist. Due to the motivation of our research being indoor exploration of unknown environments and emergency response scenarios, we consider only approaches independent of external localization infrastructure. A common approach to UAV relative localization is using a vision-based detection method. Marker-based visual detectors utilize some sort of visually distinct markers placed on the targets that are easy to distinguish from the background, such as ultraviolet (UV) LEDs in combination with UV-sensitive cameras [20]. The use of UV LEDs mitigates issues with environmental illumination and allows estimation of the target's relative position and rough estimation of its relative orientation. Marker-less approaches typically rely on a Convolutional Neural Network (CNN) that is trained to detect the desired targets [21], [22]. This approach removes the need for placing markers on the target but typically requires knowledge of the target's physical

dimensions to estimate its relative 3D position. Machine learning-based approaches also require retraining if the visual appearance of the targets or the environment changes. In general, vision-based relative localization suffers from decreased performance under low visibility conditions and often struggles to provide precise estimate of 3D positions of the targets due to inaccurate distance estimation [22]. Furthermore, vision-based approaches may suffer from blind spots depending on the number and Fields of View (FOVs) of the employed cameras.

For relative localization, UAVs can also be equipped with UWB ranging modules [23]. However, this technique provides only the relative distance in the case of a pair of UAVs. Even in the case of a formation of multiple UAVs, the localization precision is highly dependent on the shape of the formation and on the UAV positions inside of it.

It is also possible to utilize a multi-robot SLAM algorithm and thus obtain the relative UAV positions through map merging and inter-robot loop closures. Various multi-robot SLAM approaches have been proposed utilizing, e.g., LiDAR data [24], stereo camera data [25], or monocular camera data [26]. However, the multi-robot SLAM approaches require sending a large amount of data (sensor data, partial maps, etc.) over the communication network and lack reliability. Some of the multi-robot SLAMs also require to employ a ground station for computations, but we aim for running all of the algorithms in real time onboard the computationally-constrained UAV platforms.

The relative localization approach employed in this paper is built upon a markerless 3D LiDAR-based localization solution originally designed in our team for detection of non-cooperating intruder UAVs [27]. Contrary to the vision-based approaches, it is able to accurately obtain the 3D relative position between the UAVs without knowing the exact size of the other UAV and without any markers on the secondary UAV. Furthermore, it is robust to low visibility conditions due to the nature of the LiDAR sensor. The approach does not require any additional hardware onboard either of the UAVs apart from the 3D LiDAR that is already used for SLAM of the primary UAV. Additionally, this approach does not pose any additional computational nor communicational requirements on the secondary UAV other than the ability to receive position commands. To the best of our knowledge, our technique presents the first UAV cooperative navigation and guidance approach based on 3D LiDAR relative localization for indoor environments.

B. Notations

Let $\boldsymbol{x}_{\mathcal{A}}^{\mathcal{B}}$ be the vector describing the position of the origin of the coordinate frame $\{\mathcal{A}\}$ in frame $\{\mathcal{B}\}$. Let $\mathbf{R}_{\mathcal{A}}^{\mathcal{B}} \in SO(3)$ be a rotation matrix describing the rotation from frame $\{\mathcal{A}\}$ to frame $\{\mathcal{B}\}$. We denote

$$\mathbf{T}_{\mathcal{A}}^{\mathcal{B}} = \begin{bmatrix} \mathbf{R}_{\mathcal{A}}^{\mathcal{B}} & \boldsymbol{x}_{\mathcal{A}}^{\mathcal{B}} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix} \in SE(3) \tag{1}$$

as the transformation matrix from frame $\{\mathcal{A}\}$ to frame $\{\mathcal{B}\}$. Let $\boldsymbol{g}^{\mathcal{A}}$ be a vector of the goal position in frame $\{\mathcal{A}\}$.

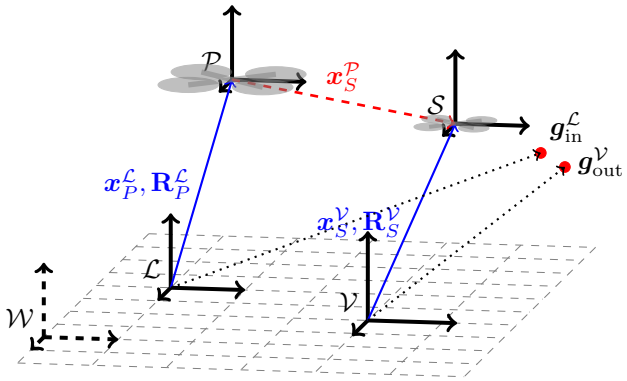

Fig. 2: Outline of the cooperative navigation and guidance problem. $\{\mathcal{W}\}$ denotes the global world frame of reference. Frames $\{\mathcal{P}\}$ and $\{\mathcal{S}\}$ correspond to the FCUs of the primary and secondary UAV, respectively. The primary UAV is localized by a LiDAR SLAM algorithm in the local frame $\{\mathcal{L}\}$, while the secondary UAV is localized by a VIO algorithm in the local frame $\{\mathcal{V}\}$. $\boldsymbol{g}_{\text{in}}^{\mathcal{L}}$ denotes the input goal position defined in the frame $\{\mathcal{L}\}$ and $\boldsymbol{g}_{\text{out}}^{\mathcal{V}}$ is the transformed goal position used in closed-loop control of the secondary UAV.

C. Problem statement

In this paper, we tackle the problem of cooperative navigation and guidance of a micro-scale aerial vehicle with limited sensory setup using onboard sensors of a better equipped UAV in the team. One of the UAVs, called the *primary* UAV, is equipped with a 3D LiDAR. The other UAV, called the *secondary* UAV, is smaller in size and is equipped with a monocular camera and an IMU. Both of these UAVs are equipped with an onboard computer, a wireless communication module, a Flight Controller Unit (FCU) with an embedded attitude controller, and an IMU. The UAVs operate in indoor GNSS-denied environment with no external localization system. All the algorithms run entirely onboard the UAVs with no external computational resources. The UAVs are able to communicate over a wireless network and the primary UAV is able to send position commands to the secondary UAV.

The navigation problem is illustrated in Fig. 2. The frame of reference $\{\mathcal{P}\}$ corresponds to the FCU of the primary UAV and frame $\{\mathcal{S}\}$ corresponds to the FCU of the secondary UAV. 3D pose of the primary UAV in the local frame $\{\mathcal{L}\}$ is obtained using a LiDAR SLAM algorithm running onboard the primary UAV. Similarly, 3D pose of the secondary UAV in $\{\mathcal{V}\}$ is obtained using a VIO algorithm running onboard the secondary UAV. The global world frame is denoted by $\{\mathcal{W}\}$.

The task tackled in this paper is to guide the secondary UAV $\{\mathcal{S}\}$ to the goal point $\boldsymbol{g}_{\text{in}}^{\mathcal{L}}$ given in the local LiDAR SLAM frame $\{\mathcal{L}\}$. The goal point is transformed to the point $\boldsymbol{g}_{\text{out}}^{\mathcal{V}}$ in the local VIO frame $\{\mathcal{V}\}$ using the output of the relative localization.

The relative transformation between the two UAVs is subject to the success of the detection, to delay and to

variable rate caused by the LiDAR data processing speed and wireless communication channel properties. The self-localization output of the secondary UAV $\{\mathcal{S}\}$ in the local VIO frame $\{\mathcal{V}\}$ utilized in the position control loop of the secondary UAV is subject to inaccuracies. The inaccuracies have the form of long-term drift of the frame $\{\mathcal{V}\}$ with respect to the world frame $\{\mathcal{W}\}$ and LiDAR SLAM frame $\{\mathcal{L}\}$ and scale of the frame $\{\mathcal{V}\}$ differing from the real-world metric scale of frame $\{\mathcal{W}\}$. The rate of the communication commands being sent from the primary UAV to the secondary UAV is subject to changes depending on the reliability of the wireless communication channel. Due to these real-world factors influencing the navigation output, the goal points $\boldsymbol{g}_{\text{in}}^{\mathcal{L}}$ and $\boldsymbol{g}_{\text{out}}^{\mathcal{V}}$ do not coincide in the global world frame $\{\mathcal{W}\}$ and the resulting control error needs to be mitigated through feedback control.

We assume that the orientation of the secondary UAV is a priori-known and constant over the course of the guiding process. The goal positions $\boldsymbol{g}_{\text{in}}^{\mathcal{L}}$ in the local LiDAR SLAM frame $\{\mathcal{L}\}$ are passed to the primary UAV by any high-level planning algorithm or a human operator.

II. MULTI-UAV SYSTEM ARCHITECTURE

A block diagram of the proposed cooperative navigation and guidance approach can be seen in Fig. 3. The primary UAV navigates through the environment based on the LiDAR data. Simultaneously, the primary UAV detects the secondary UAV from the LiDAR scans and tracks its movement. Goal points for the secondary UAV can be generated by a higher-level planning algorithm running on the primary UAV or by a human operator based on the LiDAR data. The primary UAV sends position commands to the secondary UAV over a wireless network. The primary and secondary UAVs together form a feedback loop controlling the position of the secondary UAV.

A. LiDAR-based relative UAV localization

The LiDAR-based relative localization approach is an extension of our previous work on UAV detection that was originally designed for autonomous interception of non-cooperating intruder UAVs [27]. The method consists of two modules: a *UAV detector* and a *UAV tracker*. A brief description of the detection and tracking algorithms is presented in this section, as a more detailed discussion is beyond the scope of this paper.

1) UAV detector: The detection algorithm utilizes the LiDAR data to estimate an occupancy voxel map of the environment where each voxel's value represents a confidence that the voxel contains either a background object or free air. This voxel map is estimated similarly as in common 3D mapping algorithms such as the UFOMap [29] but taking into account dynamic flying objects. Points from each LiDAR scan are raycasted from the LiDAR sensor's origin and the free-air likelihood of the intersected voxels is increased. Then, the points are separated into clusters based on their mutual Euclidean distance. Clusters closer to a high-confidence background voxel than a certain threshold are used to update the map by increasing the background likelihood of all voxels containing the corresponding points. Each remaining cluster is then either classified as a flying object if it is surrounded with high-confidence free-air voxels or it is discarded as unknown. 3D positions of the detected flying objects are the output of this algorithm.

2) UAV tracker: The *UAV tracker* serves to compensate the non-negligible delay of the *UAV detector*, improve position estimation of the target, and filter out potential sporadic false positives. For this purpose, the tracker keeps a buffer of N latest LiDAR scans and a set of current tracks that is updated using a Kalman Filter-based multi-target tracking approach with a motion model of the target.

Each input detection initializes a new track that is propagated from the corresponding scan in the buffer to the latest one. For each consecutive scan, the points in the scan are separated into Euclidean clusters. The latest voxel map from the *UAV detector* is used to reject clusters belonging to background objects. Then, the cluster closest to the track is used to update the track if it is closer than a certain threshold. Specifically, the cluster is used as a measurement for the Kalman Filter to update the track's position and velocity estimates. When the track is propagated to the latest LiDAR scan, it is either added to a set of current tracks or merged with a similar track from the set based on a similarity threshold.

The same approach is used to update the set of current tracks with new incoming LiDAR scans. After every update, the most confident track (with the smallest determinant of its covariance matrix) is outputted as the selected target. Similarly, low-confidence tracks are discarded. The output of the *UAV tracker* is the estimated position of the secondary UAV $\boldsymbol{x}_{\mathcal{S}}^{\mathcal{L}}$ in the LiDAR SLAM frame $\{\mathcal{L}\}$.

B. Target position transformation

The goal point $\boldsymbol{g}_{\text{in}}^{\mathcal{L}}$ is generated onboard the primary UAV by high-level planning. The *UAV guider* calculates the transformation $\mathbf{T}_{\mathcal{L}}^{\mathcal{S}}$ from the LiDAR SLAM frame $\{\mathcal{L}\}$ to the secondary UAV frame $\{\mathcal{S}\}$ as

$$\mathbf{T}_{\mathcal{L}}^{\mathcal{S}} = \begin{bmatrix} \mathbf{R}_{\mathcal{L}}^{\mathcal{S}} & \boldsymbol{x}_{\mathcal{L}}^{\mathcal{S}} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}, \tag{2}$$

where $\mathbf{R}_{\mathcal{L}}^{\mathcal{S}}$ is a rotation matrix calculated from zero pitch, zero roll, and known heading of the secondary UAV. $\boldsymbol{x}_{\mathcal{S}}^{\mathcal{L}}$ is the detected 3D position of the secondary UAV. The rate of the secondary UAV detections can vary, e.g., due to proximity of obstacles. Furthermore, the transformations can be delayed due to the processing time of the LiDAR data. Therefore, the transformations are stored in a buffer and are discarded if their age exceeds a predefined threshold. Transformations at the specific requested time between two available transformations in the buffer are calculated through linear interpolation by the Robot Operating System (ROS) *tf2* library[1].

[1] http://wiki.ros.org/tf2

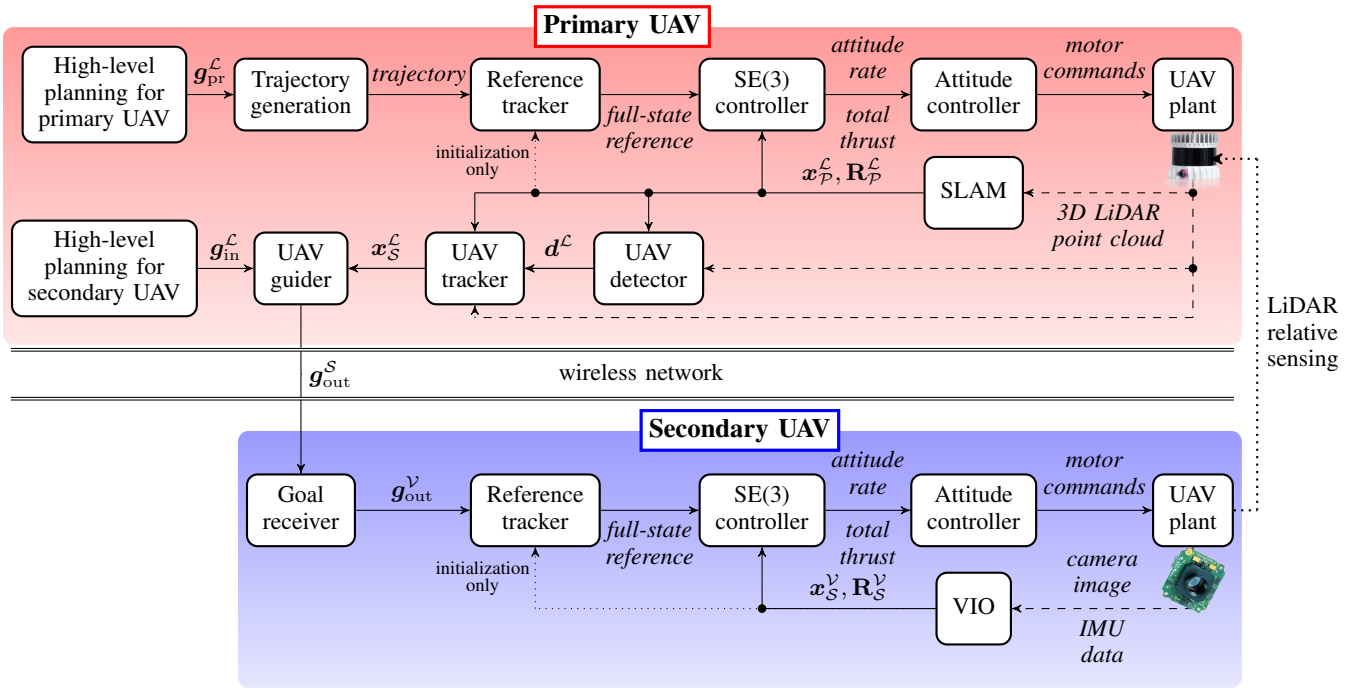

Fig. 3: Diagram of the proposed cooperative navigation approach. The primary UAV tracks the position of the secondary UAV from LiDAR data and sends position commands to the secondary UAV over wireless network. Movement of the UAV pair is planned onboard the primary UAV. Position control in the frame of each respective self-localization method is provided by the MRS UAV System [28].

The goal point is transformed to the frame of the secondary UAV as

$$\boldsymbol{g}^{\mathcal{S}}_{\text{out}} = \mathbf{T}^{\mathcal{S}}_{\mathcal{L}} \boldsymbol{g}^{\mathcal{L}}_{\text{in}} \tag{3}$$

and the resulting 3D position is sent to the secondary UAV over the wireless network. The goal point $\boldsymbol{g}^{\mathcal{S}}_{\text{out}}$ is repeatedly transmitted to the secondary UAV to account for possible loss of communication. At the secondary UAV, the goal point is received by the *Goal receiver* module, where the age of the goal point is checked to account for the goal point arriving out of order or arriving multiple times. Afterward, the goal point is transformed to the point $\boldsymbol{g}^{\mathcal{V}}_{\text{out}}$ based on the VIO self-localization output and is passed to the position control pipeline of the secondary UAV.

C. UAV position control pipeline

The position control pipeline used for guiding the UAVs to target positions specified in the frames of their respective self-localization method is provided by the MRS UAV System. This section briefly describes the specific components employed in our cooperative navigation approach. For more details, we refer the reader to the MRS UAV System paper [28].

The *trajectory generation* submodule generates a time-parametrized trajectory given a set of waypoints. The employed approach is based on the polynomial trajectory planning method described in [30]. The *Reference tracker* generates a smooth and feasible full-state reference for the feedback controllers using feedforward linear Model Predictive Control (MPC). The *SE(3) controller* is a geometric tracking feedback controller based on the method proposed in [31] that outputs desired angular rate and total thrust of the UAV. The *attitude controller* is implemented on the embedded flight controller and tracks the desired angular rate and total thrust and outputs commands to the electronic speed controllers of the UAV actuators.

The *SLAM* submodule encapsulates the LiDAR-based SLAM algorithm providing self-localization for the primary UAV. In the experimental evaluation shown in this paper, the LOAM SLAM algorithm [32] was employed. The *VIO* submodule represents the monocular VIO method used for onboard self-localization of the secondary UAV. In the experimental evaluation, the VINS-Mono algorithm [33] was utilized.

III. EXPERIMENTAL VERIFICATION

Video of the performed experiments is available online[2].

A. Simulations

The performance of the proposed cooperative navigation approach was extensively evaluated in computer simulations utilizing the Gazebo robotic simulator. The simulations were performed in the *Urban 2 Story*[3] indoor environment from OpenRobotics (see Fig. 4). The proposed approach was

[2] http://mrs.felk.cvut.cz/icuas2022-coop

[3] https://app.ignitionrobotics.org/OpenRobotics/fuel/models/Urban2Story

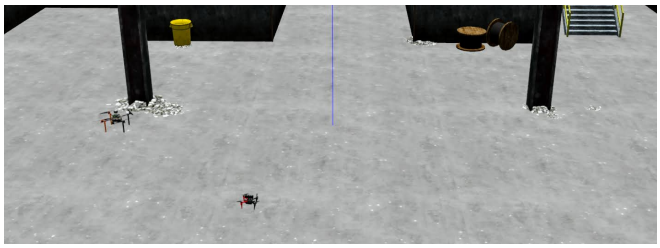

Fig. 4: The UAVs flying through the simulated environment. The primary UAV followed a line trajectory while simultaneously guiding the secondary UAV to complete a surrounding rectangular trajectory (see Fig. 5).

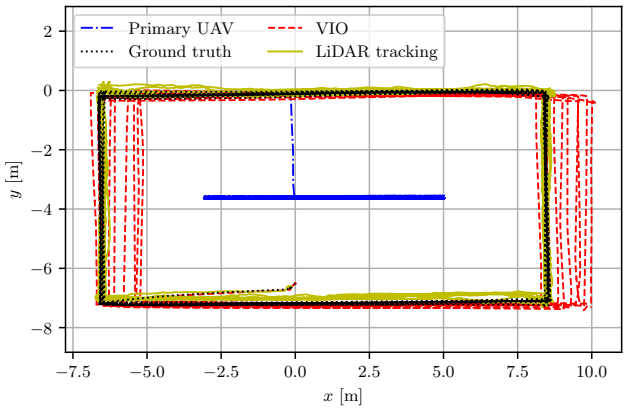

Fig. 5: Trajectories traversed by the UAVs during one of the simulations. The primary UAV followed the line trajectory in the center and guided the secondary UAV through waypoints on the surrounding rectangular trajectory. The LiDAR relative localization correctly tracked the secondary UAV while the VIO output exhibited long-term drift in the direction of the x-axis.

evaluated in its full extent. The primary UAV carried a simulated 3D LiDAR with the same parameters as the one employed in the real-world experimental evaluation. The LiDAR data were utilized both in closed-loop control of the primary UAV and for relative localization of the secondary UAV. The secondary UAV carried a simulated fisheye camera with an attached IMU. The camera data were utilized by the VINS-Mono algorithm for self-localization and consequently for use in the closed-loop control of the secondary UAV.

During each simulation, the primary UAV was flying back and forth in a straight-line trajectory. Simultaneously, the primary UAV guided the secondary UAV to fly through waypoints defined in the LiDAR SLAM frame, forming a rectangular trajectory around the primary UAV. The heading and altitude of the UAVs were kept constant during the flights. During each run of the simulation, the secondary UAV was tasked to complete the rectangular trajectory 10 times. The trajectory traversed by the UAVs during one of the simulation runs can be seen in Fig. 5. The initial positions of the trajectories in different frames of reference were aligned to the initial ground truth position. The LiDAR relative localization corresponded to the ground truth position without exhibiting any drift. On the contrary, the VIO output exhibited long-term drift in the direction of the x-axis. Nevertheless, the secondary UAV correctly followed the target rectangular trajectory without being influenced by the VIO drift as it was guided using the LiDAR relative localization. The simulated run was performed 100 times. The secondary UAV successfully performed the desired 10 loops of the rectangular trajectory in 95 runs. During 5 of the simulated runs, the simulation ended early due to a failure of the VIO self-localization.

Fig. 6 shows the progression of Root Mean Squared Error (RMSE) of the VIO and LiDAR relative localization over time. The RMSE was calculated from the 95 successful flights using Euclidean distance of the outputs of the respective localization methods from the ground truth. To calculate the RMSE value at a specific timestep, the Euclidean distances at such timestep were averaged over all the simulation runs. The 2D RMSE was calculated in the xy-plane only while the 3D RMSE utilized the full 3D Euclidean distance. The VIO-LiDAR 3D distance in the third plot was calculated as the Euclidean distance of VIO output from the LiDAR relative localization output averaged over all simulation runs at each respective timestep. The VIO output exhibited linearly-growing long-term drift while the RMSE of the LiDAR relative localization stayed approximately constant. The overall 2D RMSE of the LiDAR localization was $0.13\,\mathrm{m}$ and its overall 3D RMSE was $0.34\,\mathrm{m}$. The 3D RMSE of LiDAR relative localization was larger than the 2D RMSE as the tilting of the primary UAV, combined with the measurement delay caused by LiDAR data processing, decreased the localization precision along the z-axis.

B. Real-world experiments

The UAV platforms employed in the real-world experimental evaluation can be seen in Fig. 7. The primary UAV is built upon the Tarot 650 frame and carries the Intel NUC 10i7FNK onboard computer with the Intel Core i7 10710U CPU, 16 GB of RAM, and with a wi-fi module. The secondary UAV carries the Intel NUC 10i7FNH with the same CPU and RAM size. Both UAVs carry the Pixhawk 4 FCU which contains the embedded attitude controller along with a built-in IMU. The primary UAV carries the Ouster OS0-128 3D LiDAR. The LiDAR has a 360° horizontal and 90° vertical FOV. The LiDAR produces scans at the rate of $10\,\mathrm{Hz}$. The scans have a resolution of 1024×128 beams covering the entire FOV. The secondary UAV carries the front-facing Bluefox MLC200Wc camera with the DSL217 fisheye lens. The camera is rigidly connected to the ICM-42688-P IMU. The camera is set to produce images at the rate of $30\,\mathrm{Hz}$ at 752×480 resolution. The VINS-Mono algorithm is configured to process the images at the rate of $10\,\mathrm{Hz}$ and the IMU at the rate of $1000\,\mathrm{Hz}$.

The software running onboard the UAVs is based on Ubuntu 20.04, ROS middleware, and the MRS UAV sys-

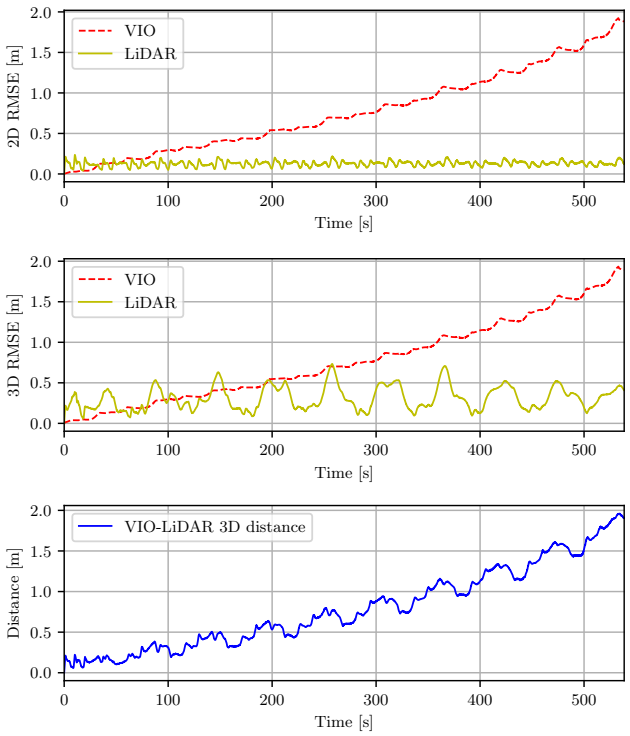

Fig. 6: Comparison of the RMSE progression over time of VIO and LiDAR relative localization calculated from the the successful simulated flights. The VIO exhibited linearly growing long-term drift while the RMSE of LiDAR relative localization stayed approximately constant.

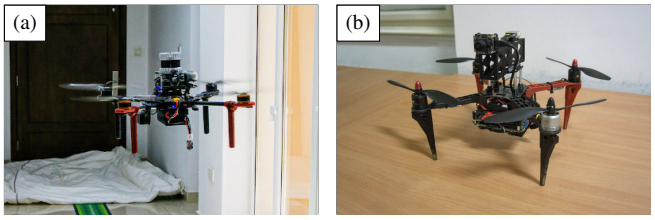

Fig. 7: The UAV platforms employed in the real-world experiments: (a) Primary UAV carrying the Ouster 3D LiDAR, (b) Secondary UAV with the monocular camera.

Fig. 8: Experimental evaluation of flight between obstacles. The primary UAV guided the secondary UAV to pass through a gap between two rectangular panels.

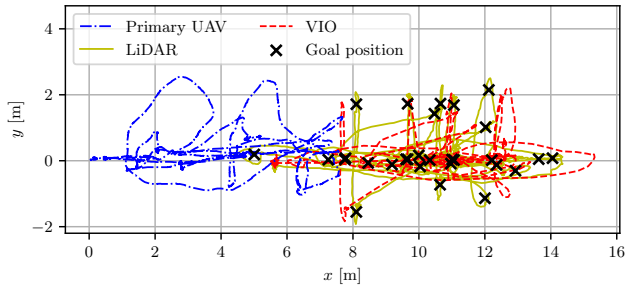

Fig. 9: UAV trajectories in an experiment with both UAVs moving. The primary UAV guided the secondary UAV to positions supplied by human operator and defined in the LiDAR SLAM frame.

tem[4] [28]. Nimbro network[5] is employed for transporting the ROS topics and services over the wireless network.

During the performed real-world experimental verification, all the algorithms ran completely onboard the UAVs and no external localization system was utilized. The experiments were performed in a large indoor arena as shown in Fig. 1 and Fig. 8. During the real-world experiments, the target heading and altitude of the secondary UAV were predefined in the local VIO frame.

1) Both UAVs moving: The first experimental trial was performed in an area without obstacles with both UAVs moving. Both UAVs were flying to target positions selected by human operators. The target positions for both UAVs were selected in the frame of the LiDAR SLAM algorithm of the primary UAV. The primary UAV simultaneously flew to its target positions, tracked the secondary UAV, and sent transformed target positions to the secondary UAV.

The trajectories traversed by the UAVs are shown in Fig. 9. During the flight, the primary UAV kept changing its position, altitude, and heading. The secondary UAV was successfully guided to the target positions. Fig. 10 shows a comparison of the VIO and LiDAR relative localization outputs along with the goal positions. The figure also shows the 3D distance between the outputs of VIO and LiDAR relative localization calculated at each point over time during the experiment. The VIO output gradually drifted away from the LiDAR localization output. The sharp peaks in the distance are caused by measurement delay due to LiDAR processing and mitigating them is a subject of future work.

2) Square trajectory at higher speed: The second experiment evaluated the ability to guide the secondary UAV along a square trajectory at higher velocity. The primary UAV was placed on the ground while it guided the secondary UAV to perform the flight. During this flight, the secondary UAV reached peak velocity of approximately $1.5\,\mathrm{m\,s^{-1}}$ while in the other real-world experiments, the peak velocity was approximately $1.1\,\mathrm{m\,s^{-1}}$ (calculated from the VIO self-localization output). Fig. 11 displays the traversed trajectory.

[4]https://github.com/ctu-mrs/mrs_uav_system
[5]https://github.com/AIS-Bonn/nimbro_network

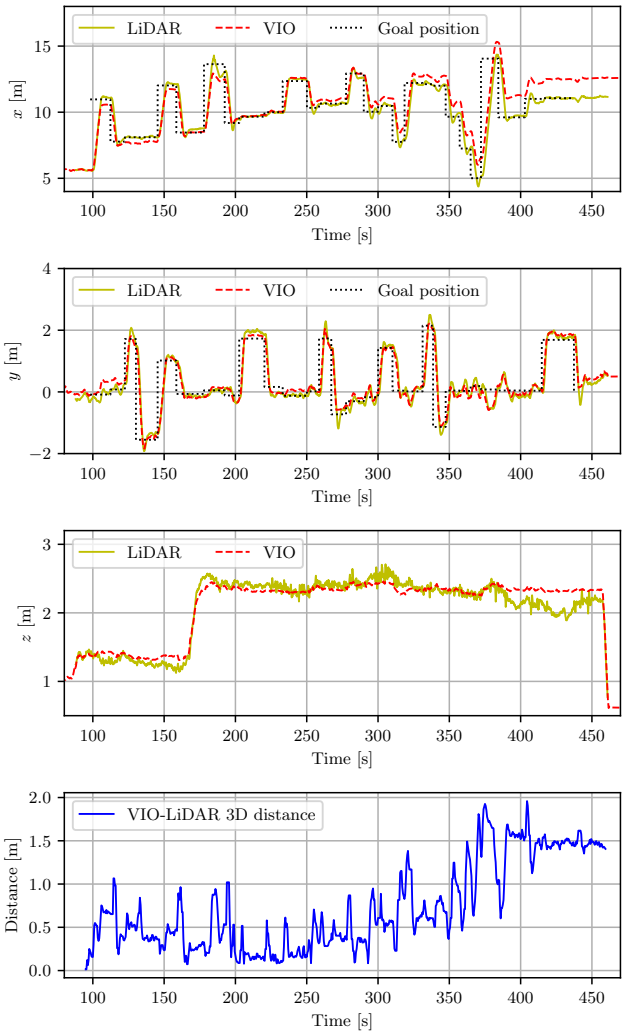

Fig. 10: Comparison of VIO and LiDAR relative localization outputs along with goal positions of the secondary UAV from the experiment with both UAVs moving. The primary UAV successfully guided the secondary UAV to the target positions even though the VIO output exhibited significant drift.

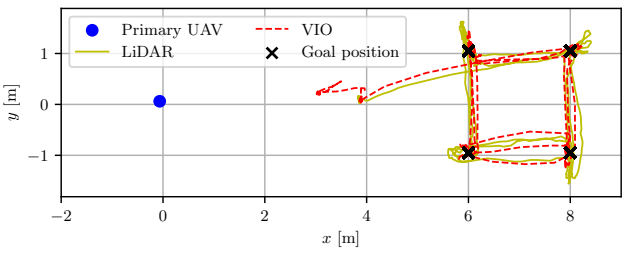

Fig. 11: UAV trajectory from the experiment evaluating square trajectory following at higher velocity. The primary UAV stood on the ground while the secondary UAV was flying in the air.

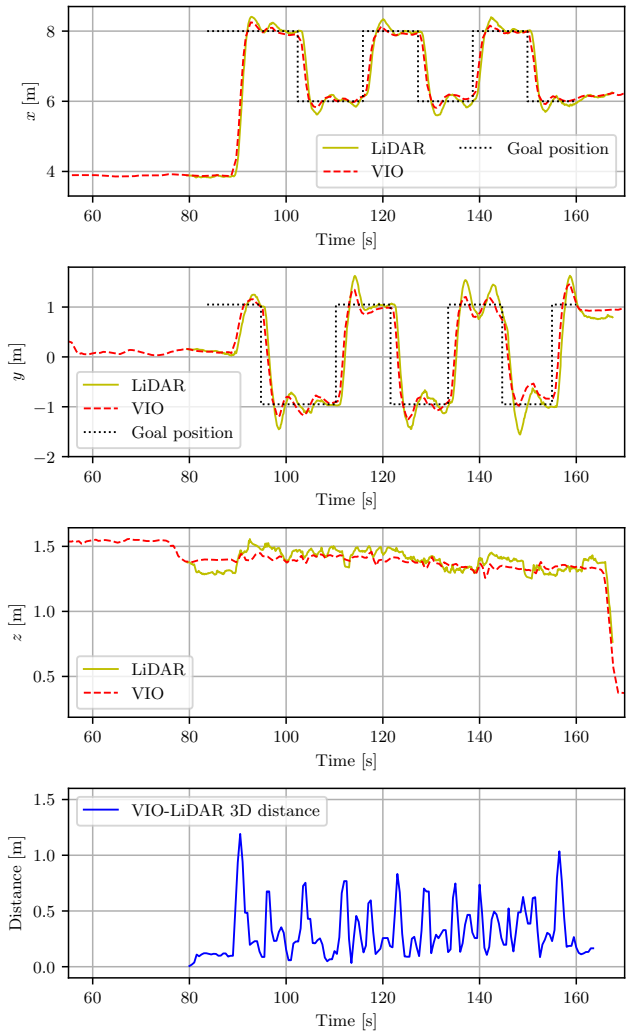

Fig. 12: Comparison of VIO and LiDAR relative localization outputs from the experiment evaluating square trajectory following at higher velocity. The secondary UAV successfully performed the target trajectory. The sharp peaks in the VIO-LiDAR 3D distance are caused by delay due to LiDAR processing time.

Fig. 12 displays the comparison of VIO and LiDAR relative localization outputs. The secondary UAV successfully performed the desired trajectory.

3) Flight between obstacles: In the final experiment, the primary UAV was hovering in place and simultaneously guided the secondary UAV to pass between a pair of obstacles three times (see Fig. 8). The target positions for the secondary UAV were predefined in the LiDAR SLAM frame of the primary UAV. Fig. 13 shows the trajectories traversed by the UAV during the experiment and Fig. 14 shows the comparison of localization outputs. The VIO and LiDAR relative localization outputs drifted apart during the flight between the obstacles. Nevertheless, the secondary UAV successfully performed the maneuver. This experiment evaluated the performance of the proposed approach in

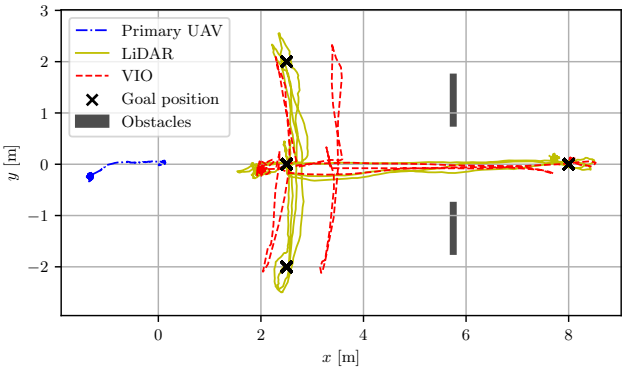

Fig. 13: UAV trajectories from the flight between obstacles. The primary UAV guided the secondary UAV to pass through a gap between two rectangular panels.

close proximity of obstacles. The LiDAR relative localization method successfully tracked the secondary UAV throughout the entire flight.

C. Results summary

Both simulations and real-world experiments have shown that the LiDAR relative localization outperforms the VIO self-localization in terms of accuracy, as it is not influenced by the detrimental factors inherent to visual self-localization methods. The relative localization was successfully utilized for precisely guiding the secondary UAV even when the primary UAV was moving and in close proximity of obstacles. The real-world data corresponded to the simulation results demonstrating the drift of VIO positions from the LiDAR relative localization output (see Fig. 6 and Fig. 10). In the flight near obstacles, the VIO output drifted more than 1 m away from the LiDAR position (see Fig. 13). Such significant drift may lead to a collision if the UAV was not guided by the proposed approach. In all real-world experiments, the proposed approach successfully tracked the secondary UAV and provided estimates of its position at the rate of 10 Hz, corresponding to the scanning rate of the 3D LiDAR. The accuracy of LiDAR relative localization was influenced by the measurement delay caused by LiDAR processing time. Mitigating the effects of this delay is a subject of future work. Future work will also focus on the estimation of relative orientation between the localization outputs and fusion of LiDAR detections with the VIO output to remove the simplifying assumptions of the current approach.

IV. CONCLUSIONS

A novel approach for cooperative navigation and guidance of a micro-scale aerial vehicle by an accompanying UAV using 3D LiDAR relative localization was proposed in this paper. The 3D LiDAR provides rich data for robust UAV self-localization and environment perception. However, it needs to be carried by a relatively large UAV platform due to its weight. On the contrary, a monocular visual camera is cheap, light-weight, and therefore ideal for a small payload-constrained UAV. The proposed approach combines the

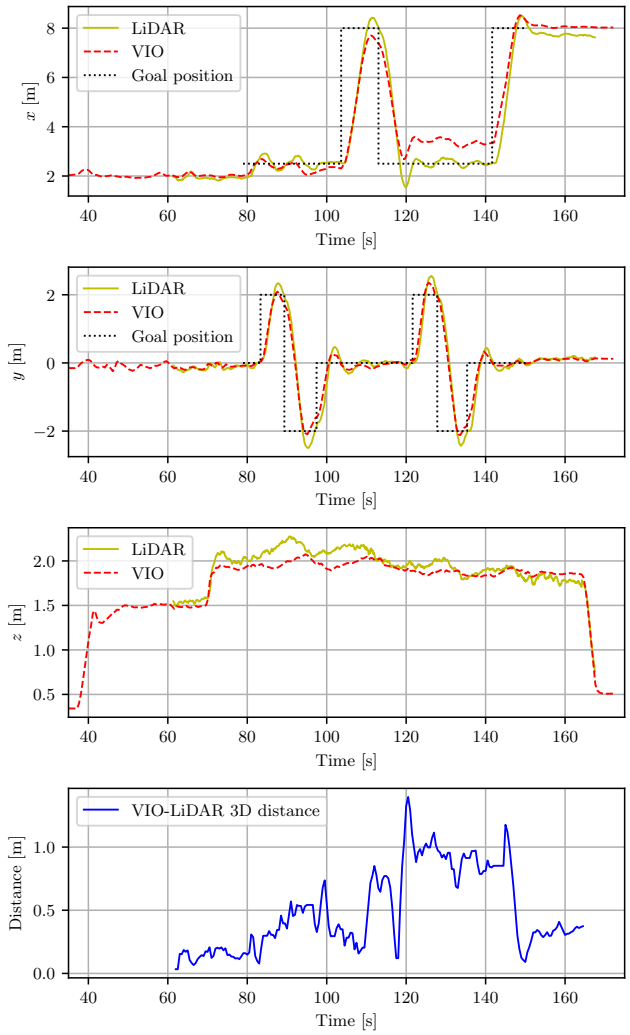

Fig. 14: Comparison of VIO and LiDAR relative localization outputs from the flight between obstacles. The secondary UAV successfully performed the desired trajectory even though the VIO output drifted during the flight.

advantages of both types of sensors to create a heterogeneous UAV team capable of robust navigation and at the same time capable of flight through constrained environments. 3D LiDAR data are used for markerless relative localization between the two UAVs to enable precise guidance of the secondary UAV in the frame of reference of the accurate LiDAR SLAM. The proposed approach was extensively evaluated in simulations and multiple real-world experiments with the algorithms running completely onboard the UAVs with no external localization infrastructure. The experiments showed the superior precision of the LiDAR relative localization when compared to the VIO outputs and successfully demonstrated the ability to guide the secondary UAV to target positions defined in the LiDAR SLAM frame of reference.

REFERENCES

[1] V. Spurny, V. Pritzl, V. Walter, M. Petrlik, T. Baca, P. Stepan, D. Zaitlik, and M. Saska, "Autonomous Firefighting Inside Buildings

by an Unmanned Aerial Vehicle," *IEEE Access*, vol. 9, pp. 15 872–15 890, 2021.

[2] V. Pritzl, P. Stepan, and M. Saska, "Autonomous Flying into Buildings in a Firefighting Scenario," in *2021 International Conference on Robotics and Automation (ICRA)*, 2021, pp. 239–245.

[3] J. Delmerico, S. Mintchev, A. Giusti, B. Gromov, K. Melo, T. Horvat, C. Cadena, M. Hutter, A. Ijspeert, D. Floreano, L. M. Gambardella, R. Siegwart, and D. Scaramuzza, "The current state and future outlook of rescue robotics," *Journal of Field Robotics*, vol. 36, no. 7, pp. 1171–1191, 2019.

[4] M. Petrlik, T. Baca, D. Hert, M. Vrba, T. Krajnik, and M. Saska, "A Robust UAV System for Operations in a Constrained Environment," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2169–2176, Apr. 2020.

[5] V. Krátký, P. Petráček, T. Báča, and M. Saska, "An autonomous unmanned aerial vehicle system for fast exploration of large complex indoor environments," *Journal of Field Robotics*, vol. 38, no. 8, pp. 1036–1058, 2021.

[6] P. Petráček, V. Krátký, M. Petrlík, T. Báča, R. Kratochvíl, and M. Saska, "Large-Scale Exploration of Cave Environments by Unmanned Aerial Vehicles," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7596–7603, Oct. 2021.

[7] P. Petráček, V. Krátký, and M. Saska, "Dronument: System for Reliable Deployment of Micro Aerial Vehicles in Dark Areas of Large Historical Monuments," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2078–2085, Apr. 2020.

[8] V. Krátký, P. Petráček, V. Spurný, and M. Saska, "Autonomous Reflectance Transformation Imaging by a Team of Unmanned Aerial Vehicles," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2302–2309, Apr. 2020.

[9] M. Petrlík, T. Krajník, and M. Saska, "LIDAR-based Stabilization, Navigation and Localization for UAVs Operating in Dark Indoor Environments," in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, Jun. 2021, pp. 243–251.

[10] K. J. Wu and S. I. Roumeliotis, "Unobservable Directions of VINS Under Special Motions," Dept. of Computer Science and Engineering, University of Minnesota, Tech. Rep., 2016. [Online]. Available: http://mars.cs.umn.edu/papers/tech_report_unobs_directs.pdf

[11] F. Causa, R. Opromolla, and G. Fasano, "Cooperative navigation and visual tracking with passive ranging for UAV flight in GNSS-challenging environments," in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, Jun. 2021, pp. 1538–1547.

[12] F. Causa and G. Fasano, "Adaptive Cooperative Navigation Strategies for Complex Environments," in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Apr. 2020, pp. 100–111.

[13] Y. Xianjia, L. Qingqing, J. P. Queralta, J. Heikkonen, and T. Westerlund, "Cooperative UWB-Based Localization for Outdoors Positioning and Navigation of UAVs aided by Ground Robots," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, Aug. 2021, pp. 1–5.

[14] V. O. Sivaneri and J. N. Gross, "Flight-testing of a cooperative UGV-to-UAV strategy for improved positioning in challenging GNSS environments," *Aerospace Science and Technology*, vol. 82-83, pp. 575–582, Nov. 2018.

[15] S. J. Dourmashkin, N. R. Ahmed, D. M. Akos, and W. W. Whitacre, "GPS-limited cooperative localization using scalable approximate decentralized data fusion," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Apr. 2018, pp. 1473–1484.

[16] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency Analysis and Improvement of Vision-aided Inertial Navigation," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, Feb. 2014.

[17] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "VINS on wheels," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 5155–5162.

[18] Y. Yang and G. Huang, "Observability Analysis of Aided INS With Heterogeneous Features of Points, Lines, and Planes," *IEEE Transactions on Robotics*, vol. 35, no. 6, pp. 1399–1418, Dec. 2019.

[19] H. Lim, Y. Kim, K. Jung, S. Hu, and H. Myung, "Avoiding Degeneracy for Monocular Visual SLAM with Point and Line Features," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, May 2021, pp. 11 675–11 681.

[20] V. Walter, N. Staub, A. Franchi, and M. Saska, "UVDAR System for Visual Relative Localization With Application to Leader–Follower Formations of Multirotor UAVs," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2637–2644, Jul. 2019.

[21] F. Schilling, F. Schiano, and D. Floreano, "Vision-Based Drone Flocking in Outdoor Environments," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2954–2961, Apr. 2021.

[22] M. Vrba and M. Saska, "Marker-Less Micro Aerial Vehicle Detection and Localization Using Convolutional Neural Networks," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2459–2466, Apr. 2020.

[23] K. Guo, X. Li, and L. Xie, "Ultra-Wideband and Odometry-Based Cooperative Relative Localization With Application to Multi-UAV Formation Control," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2590–2603, Jun. 2020.

[24] Y. Huang, T. Shan, F. Chen, and B. Englot, "DiSCo-SLAM: Distributed Scan Context-Enabled Multi-Robot LiDAR SLAM With Two-Stage Global-Local Graph Optimization," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1150–1157, Apr. 2022.

[25] Y. Tian, Y. Chang, F. H. Arias, C. Nieto-Granda, J. P. How, and L. Carlone, "Kimera-Multi: Robust, Distributed, Dense Metric-Semantic SLAM for Multi-Robot Systems," *IEEE Transactions on Robotics*, pp. 1–17, 2022.

[26] P. Schmuck and M. Chli, "Multi-UAV collaborative monocular SLAM," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 3863–3870.

[27] M. Vrba, Y. Stasinchuk, T. Báča, V. Spurný, M. Petrlík, D. Heřt, D. Žaitlík, and M. Saska, "Autonomous capture of agile flying objects using UAVs: The MBZIRC 2020 challenge," *Robotics and Autonomous Systems*, vol. 149, p. 103970, Mar. 2022.

[28] T. Baca, M. Petrlik, M. Vrba, V. Spurny, R. Penicka, D. Hert, and M. Saska, "The MRS UAV System: Pushing the Frontiers of Reproducible Research, Real-world Deployment, and Education with Autonomous Unmanned Aerial Vehicles," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 1, p. 26, Apr. 2021.

[29] D. Duberg and P. Jensfelt, "UFOMap: An efficient probabilistic 3D mapping framework that embraces the unknown," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6411–6418, Oct. 2020.

[30] C. Richter, A. Bry, and N. Roy, "Polynomial Trajectory Planning for Aggressive Quadrotor Flight in Dense Indoor Environments," in *Robotics Research*. Springer, Apr. 2016, pp. 649–666.

[31] T. Lee, M. Leok, and N. H. McClamroch, "Geometric tracking control of a quadrotor UAV on SE(3)," in *49th IEEE Conference on Decision and Control (CDC)*, Dec. 2010, pp. 5420–5425.

[32] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," in *Robotics: Science and Systems (RSS)*, Jul. 2014, pp. 109–111.

[33] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.