Localization Fusion for Aerial Vehicles in Partially GNSS Denied Environments

Jan Bayer and Jan Faigl[0000−0002−6193−0792]

Faculty of Electrical Engineering, Czech Technical University in Prague,
Technicka 2, 166 27, Prague, Czech Republic
{bayerja1,faiglj}@fel.cvut.cz
https://comrob.fel.cvut.cz

Abstract. In this paper, we report on early results of the experimental deployment of localization techniques for a multi-rotor Micro Aerial Vehicle (MAV). In particular, we focus on deployment scenarios where the Global Navigation Satellite System (GNSS) does not provide a reliable signal, and thus it is not desirable to rely solely on the GNSS. Therefore, we consider recent advancements in the visual localization, and we employ an onboard RGB-D camera to develop a robust and reliable solution for the MAV localization in partially GNSS denied operational environments. We consider a localization method based on Kalman filter for data fusion of the vision-based localization with the signal from the GNSS. Based on the reported experimental results, the proposed solution supports the localization of the MAV for the temporarily unavailable GNSS, but also improve the position estimation provided by the incremental vision-based localization system while it can run using onboard computational resources of the small vehicle.

1 Introduction

Accurate and reliable localization of a multi-rotor Micro Aerial Vehicle (MAV) is an essential prerequisite for its deployment not only in autonomous missions but also in semi-autonomous deployments where the vehicle is requested to follow a pre-designed path. The Global Navigation Satellite System (GNSS) is a natural choice and practical solution for outdoor missions, where it provides an estimation of the vehicle position. On the other hand, the GNSS system is not always available, and it becomes unreliable or inaccurate inside and also close to tall buildings or at places where not enough satellites are in the line of sight [27]. Therefore, we examine the properties of existing vision-based localization methods using onboard sensors to provide an estimation of the MAV position. In particular, we focus on the visual odometry and vision-based methods for Simultaneous Localization and Mapping (SLAM) [23] that recently exhibited significant improvements in the pose estimation [12].

The proposed method is based on a fusion of the unreliable GNSS localization and state-of-the-art vision-based localization. The used sensor fusion approach is based on Kalman filter [23] and the main contribution of this paper is in the

Fig. 1. The MAV used during the experiments.

presented results on the experimental deployment of the developed solution on the MAV in an outdoor scenario, with the recent Intel RealSense D435 sensor [3] depicted in Fig. 1.

The developed method itself represents a general fusion framework to combine various sources of the external localization like the GNSS with the incremental localization such as visual odometry. The herein reported results support the feasibility of the proposed solution in the particular experimental deployment where a precise pose estimation from the Global Positioning System (GPS), which provides the necessary ground truth for the evaluation, is modified to decrease its reliability by adding Gaussian noise and disabling the GNSS completely for a certain deployment period.

The rest of the paper is organized as follows. Principles of the existing localization methods, the most related sensory fusion methods, and metrics of localization precision assessment are overviewed in Section 2. The model for the proposed localization fusion is derived in Section 3. The evaluation results of the proposed solution are reported in Section 4. The concluding remarks on the achieved results and our future work are presented in Section 5.

2 Related Work

Two main classes of the methods to localize a robot in GNSS denied environments can be identified in the literature. The first class of the methods are solutions that rely on beacons and transmitters [10], or other sensors like cameras [2], [16] placed in the robot operational space. We do not consider these methods in this paper because these methods require an additional infrastructure in the robot operational space, which is not practical for our motivational deployment scenario.

The methods of the second class rely on exteroceptive sensors like LIDARs or cameras mounted on a robot. These sensors are used to compute a position of the robot incrementally from consecutive LIDAR scans or camera frames using some representative landmarks detected as visual feature points in the scans

and camera image. The concept of the incremental localization can be further extended to improve the localization estimation by simultaneously creating a map of the environment that is used in localization of the vehicle using new sensor measurements. Such a technique can be found in the literature as the SLAM [23]. A SLAM method can also be considered as an incremental localization system with loop closures that further improve the localization estimation once a robot returns to already visited places, and thus decrease the localization drift.

Methods for localization of the MAV using LIDAR scans have been successfully deployed in [27], [17], [4]. Although localization methods based on LIDAR sensors exhibit remarkable advancements during the last years, LIDARs are relatively expensive sensors in comparison to cameras. Besides, they are still mechanical systems, and they are more substantial than most of the conventional cameras. Therefore, we prefer vision-based localization systems that are using image processing from the onboard camera or especially new RGB-D cameras.

Several existing vision-based localization methods employ principles of the visual odometry from a single camera up to complex SLAM systems using several cameras of different types. The existing monocular camera localization methods include SVO [8], DSO [6], ORB-SLAM2 [14], and many others [18]. The main disadvantage of these methods is that they need several frames or additional information to obtain the scale of the scene, and thus the methods suffer from the map initialization problem, i.e., the bootstrapping problem [9]. Therefore, it is beneficial to use sensors that can estimate the scale of the scene from a single frame. It is the case of stereo, and RGB-B cameras and data from these sensors can be used in the existing methods such as the RGB-D SLAM [5], ORB-SLAM2 [13], and Stereo DSO [25] to name a few.

Beside exteroceptive sensors, the pose estimation can also be improved using data from the Inertial Measurement Unit (IMU). The authors of [27] and [4] report on fusion of IMU measurements with the localization based on the LIDARs data by Kalman filter. A complex navigation system for the six-legged walking robot is proposed in [21], where the authors use the information filter that is numerically equivalent to Kalman filter. Even though the authors of [21] report on lower computational requirements of the information filter in comparison to Kalman filter, the considered information filter is a more complex system than a regular Kalman filter.

In addition to the fusion of IMU measurements with other signals, the IMU data measurements can be directly utilized inside a localization method itself, for example as in a variant of the Stereo Parallel Tracking and Mapping (S-PTAM) method [19] proposed in [7]. However, such an approach requires changes to the visual localization method and thus developing a new localization method, which is not our main intention.

Regarding the presented overview of the existing methods and our previous work [15], we choose ORB-SLAM2 in combination with the RGB-D camera as the main visual localization system to overcome GNSS denied regions of the vehicle operational environment. Besides, we chose Kalman filter approach for data fusion because it is more straightforward than an information filter.

2.1 Localization Metrics

In our work, we aim to improve the localization system, and therefore, we need a method to measure and compare the precision of different localization systems. The general framework for measuring localization precision is described in [22], where the authors measure the error of the localization from the trajectories obtained during the robotic experimental trial in a particular scenario. The compared trajectories are the ground truth, which is provided by a reference localization system, e.g., D-GPS, and a trajectory estimate provided by the examined localization system. The metrics to measure the error of the estimated trajectory are the *Absolute Trajectory Error* (ATE) and *Relative Pose Error* (RPE) [22], which are well-established and used in the literature.

The ATE is given by the equation:

$$\mathbf{F}_i = \mathbf{Q}_i^{-1}\,\mathbf{S}\,\mathbf{P}_i, \tag{1}$$

where the matrices $\mathbf{Q}_i$ and $\mathbf{P}_i$ are SE(3) positions of the ground truth and estimated trajectory, respectively. The matrix $\mathbf{S}$ is a transformation between the coordinate frames of the ground truth and trajectory estimate. The transformation is obtained from the minimization of the squared distances between the corresponding positions of the trajectory estimate and the ground truth [22].

The RPE is given by the equation:

$$\mathbf{E}_i = (\mathbf{Q}_i^{-1}\mathbf{Q}_{i+\Delta})^{-1}(\mathbf{P}_i^{-1}\mathbf{P}_{i+\Delta}), \tag{2}$$

where Δ represents a fixed distance between two positions that are used during the calculation [22] and it is $\Delta = 1$ in our case.

Once the ATE and RPE values are computed for the whole trajectory, the computed error values are used to get a statistical indicator to evaluate the precision of the localization based on the whole trajectory. In [22], the authors suggest to use the average value and the Root Mean Square (RMS) to the determined statistical indicators as

$$\begin{aligned}
\overline{\mathrm{ATE}}_t &= \frac{1}{n}\sum_{i=1}^{n} \parallel trans(\mathbf{F}_i) \parallel, \quad RMS(\mathrm{ATE}_t) = \left(\frac{1}{n}\sum_{i=1}^{n} \parallel trans(\mathbf{F}_i) \parallel\right)^{\frac{1}{2}}, \\
\overline{\mathrm{RPE}}_t &= \frac{1}{n}\sum_{i=1}^{n} \parallel trans(\mathbf{E}_i) \parallel, \quad RMS(\mathrm{RPE}_t) = \left(\frac{1}{n}\sum_{i=1}^{n} \parallel trans(\mathbf{E}_i) \parallel\right)^{\frac{1}{2}},
\end{aligned} \tag{3}$$

where $trans()$ computes the size of the translation from the SE(3) matrix. In the herein reported results, we assume only the translation errors, because the used fusion method fuses only the positions of the robot. Besides, the statistical indicators that compute the average ATE and average RPE are utilized because the RMS does not provide any additional information in our particular case.

3 Fusion of the GNSS with Vision-based Localization

We propose to use linear Kalman filter to overcome a loss of the GNSS signal that is substituted by the incremental vision-based localization in the GNSS denied

environments [11]. Kalman filter is a standard technique for a relatively computationally inexpensive fusion of two or more sources of measurements, which allows developing a system to uninterruptedly provide the requested estimation of the vehicle position when a particular source is not available. Contrary to the systems with SLAM based localization [26], we select an incremental vision-based localization because we assume the GNSS signal is only temporarily unavailable. Thus, it is not expected a full loop-closure is necessary, and therefore, we can choose an existing SLAM system with low computational requirements.

The developed localization system is based on a linear Kalman filter to combine an external GNSS localization with the incremental vision-based localization. The system consists of two principal parts: Kalman filter, and a model of the vehicle, which is derived from the motion equation of the robot body. Both parts are described in the following paragraphs to make the paper self-contained.

Kalman filter for a linear and discrete-time model of the system described by the state equation

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \tag{4}$$

can be defined by the set of the following equations [23]:

$$\overline{\boldsymbol{\mu}}_k = \mathbf{A}_k\boldsymbol{\mu}_{k-1} + \mathbf{B}_k\boldsymbol{u}_k \tag{5}$$
$$\overline{\boldsymbol{\Sigma}}_k = \mathbf{A}_k\boldsymbol{\Sigma}_{k-1}\mathbf{A}_k^T + \mathbf{R}_k \tag{6}$$

$$\mathbf{K}_k = \overline{\boldsymbol{\Sigma}}_k\mathbf{H}_k^T(\mathbf{H}_k\overline{\boldsymbol{\Sigma}}_k\mathbf{H}_k^T + \mathbf{Q}_k)^{-1} \tag{7}$$
$$\boldsymbol{\mu}_k = \overline{\boldsymbol{\mu}}_k + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\overline{\boldsymbol{\mu}}_k) \tag{8}$$
$$\boldsymbol{\Sigma}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\overline{\boldsymbol{\Sigma}}_k, \tag{9}$$

where $\overline{\boldsymbol{\mu}}_k$, $\overline{\boldsymbol{\Sigma}}_k$ are mean, and covariance of the system states predicted based on the last estimated mean $\boldsymbol{\mu}_k$ of the system states $\mathbf{x}_k$. The predictions can be used asynchronously with the correction phase represented by (7), (8), and (9) in the case we need the robot position in an arbitrary time. However, in this paper, we extract the position of the robot synchronously with the correction phase from $\boldsymbol{\mu}_k$. $\boldsymbol{\Sigma}_k$ represents the covariance of the system states $\mathbf{x}_k$ and the matrix $\mathbf{R}_k$ represents the uncertainty of the model of the system. The vector $\mathbf{z}_k$ is the measurement obtained according to

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\delta}_k, \tag{10}$$

where $\boldsymbol{\delta}_k$ is the Gaussian noise with the zero mean and covariance $\mathbf{Q}_k$. The more detailed description of the vehicle model and the related matrices is as follows.

The model of the vehicle is based on the body motion model that can be described as

$$\begin{bmatrix}\mathbf{p}_{k+1}\\ \mathbf{v}_{k+1}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_{3\times3} & t_s\cdot\mathbf{I}_{3\times3}\\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3}\end{bmatrix}\begin{bmatrix}\mathbf{p}_k\\ \mathbf{v}_k\end{bmatrix} + \begin{bmatrix}\frac{t_s^2}{2m}\cdot\mathbf{I}_{3\times3}\\ \frac{t_s}{m}\cdot\mathbf{I}_{3\times3}\end{bmatrix}\mathbf{f}_k, \tag{11}$$

where $\mathbf{f}_k$ is the force that moves the vehicle, $\mathbf{p}_k$ and $\mathbf{v}_k$ are the position and velocity of the vehicle, respectively, t_s is the time, and m is the vehicle weight.

Let suppose that we use the position from the incremental localization as the position $\mathbf{p}_k$ during the filtration. The problem is that the incremental localization drifts, and thus its uncertainty grows to infinity. Moreover, if the uncertainty of the incremental localization increases to a certain level, the contribution of the incremental localization would vanish. On the other hand, the incremental localization provides relatively precise estimates of the transformations between the consecutive robot positions, and a precision of these transformations is not affected by the position drift. Therefore, we can incorporate differences of the consecutive positions to the vehicle model. The straightforward way to incorporate these differences is to substitute the vehicle velocity using the equation

$$\mathbf{v}_k = \frac{\mathbf{q}_k - \mathbf{q}_{k-1}}{t_s} = \frac{\Delta\mathbf{q_k}}{t_s}. \tag{12}$$

Then, the model of the system (11) changes to

$$\begin{bmatrix} \mathbf{p}_{k+1} \\ \Delta\mathbf{q}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix} \begin{bmatrix} \mathbf{p}_k \\ \Delta\mathbf{q}_k \end{bmatrix} + \begin{bmatrix} \frac{t_s^2}{2m} \cdot \mathbf{I}_{3\times3} \\ \frac{t_s}{m} \cdot \mathbf{I}_{3\times3} \end{bmatrix} \mathbf{f}_k. \tag{13}$$

In the case of the used MAV, the force $\mathbf{f}_k$, which moves the vehicle, is usually not hard to obtain. In other cases, when the force is laborious to obtain, it is possible to follow the tracking scenario [24], which is considered in the rest of this section. The model of the vehicle can be expressed without the force as

$$\begin{bmatrix} \mathbf{p}_{k+1} \\ \Delta\mathbf{q}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix} \begin{bmatrix} \mathbf{p}_k \\ \Delta\mathbf{q}_k \end{bmatrix}, \tag{14}$$

where the matrices of the state equation (4) used in Kalman filter are

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{p}_k \\ \Delta\mathbf{q}_k \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix}, \ \mathbf{B} = \mathbf{0}_{6\times1}.$$

There is not an input to the model (14) because the model only updates the position of the robot by the last value of $\Delta\mathbf{q}_k$. Thus, the position of the robot provided by the model is uncertain and has to be corrected by new measurements. The model uncertainty is described by the covariance matrix $\mathbf{R}_k$, see (6).

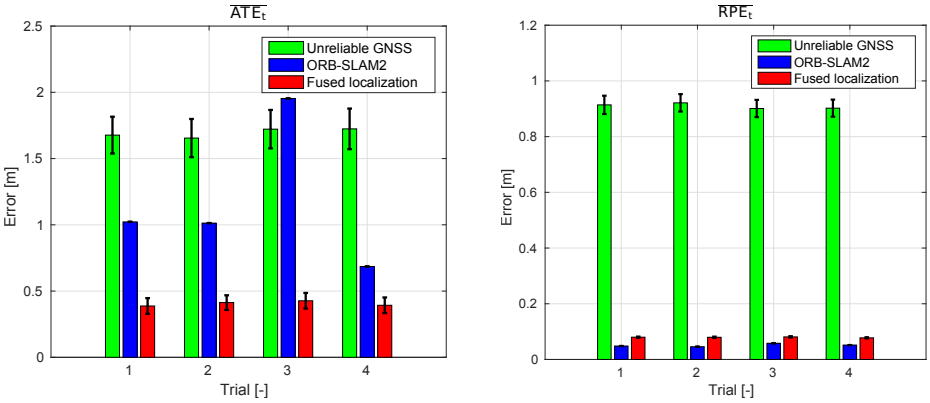
The measurements used to correct the position of the robot estimated by the model (14) are: (i) the position data provided by the external localization that corrects the state $\mathbf{p}_k$; and (ii) differences of the consecutive measurements provided by the incremental localization correct the states $\Delta\mathbf{q}_k$. The precision of the measurements is described by the covariance matrix $\mathbf{Q}_k$, which represents Gaussian noise $\boldsymbol{\delta}_k$ in the measurement equation (10). The matrices $\mathbf{H}_k$ and $\mathbf{Q}_k$ of the measurement equation are composed as follows.

$$\mathbf{H} = \mathbf{H}_{k=1,2,\ldots} = \mathbf{I}_{6\times6}, \quad \mathbf{Q}_k = \begin{bmatrix} \mathbf{Q}_{p,k} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{Q}_{\Delta q,k} \end{bmatrix},$$

where $\mathbf{Q}_{p,k}$ and $\mathbf{Q}_{\Delta q,k}$ represent the uncertainty of the external localization and the uncertainty of the incremental localization, respectively.

4 Experimental Results

The developed localization system has been experimentally verified with the real MAV shown in Fig. 1 in a practical field experimental deployment. The experiments took place in the environment where the GNSS was available because we used a precise GPS as the reference localization system to evaluate the performance of the developed solution with the available ground truth. Thus, the ground truth and RGB-D data from the visual localization have been collected using ROS [20] in four flights with the vehicle operating at the height between 5 and 10 meters above the terrain. The reference localization has been provided by the localization system that combines a precise GPS with a laser altimeter, and the IMU onboard of the MAV.



(a) Average ATE (b) Average RPE

Fig. 2. Accuracy of the provided trajectory localization for the unreliable GNSS, ORB-SLAM2, and the localization fusion.

The RGB-D data provided by the Intel RealSense D435 [3] have been processed by the ORB-SLAM2 [13] to get trajectory estimation solely based on the visual localization. The captured data have been processed at the same frequency as they have been captured by the onboard RGB-D camera using ROS. In particular, the ORB-SLAM2 was running on the computer with the dual-core Intel i5-5257U CPU running at 2.7 GHz with 4 GB of memory. The used computational environment has similar computational power as the onboard computer at the used MAV to reflect requirements on the onboard processing. The conditions under we used the visual localization are almost identical to the on-line deployment on the vehicle, and the ORB-SLAM2 provides localization at the average frequency of 6.67 Hz. Notice that even though the ORB-SLAM2 can perform a large loop closing, which may improve the localization precision significantly, this feature has been disabled to demonstrate the ability of the proposed filter to address the localization drift.

An unreliable GNSS-based localization has been substituted by adding the Gaussian noise with the standard deviation of 0.5 m in all three axes to the ground truth trajectory. Moreover, we selected approximately 20% of the trajectory, where the localization signal was disabled entirely to simulate temporal unavailability of the GNSS. The developed Kalman filter was used for fusing the localization provided by the ORB-SLAM2 and the simulated unreliable GNSS.

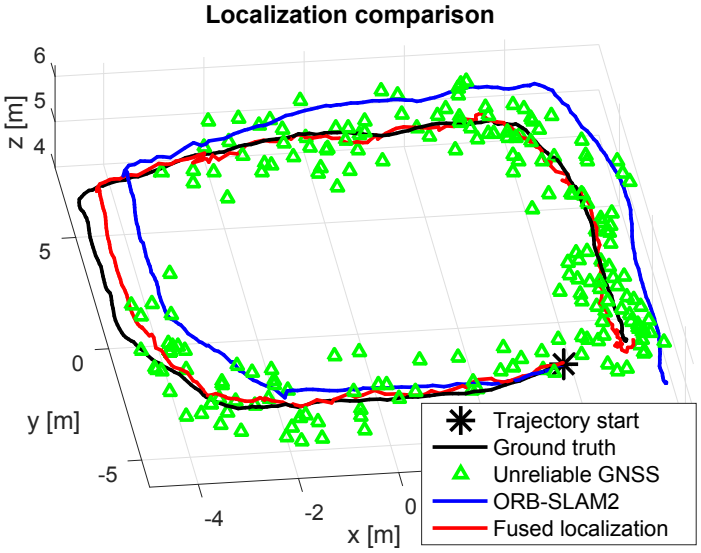


Fig. 3. Trajectory localization provided by the evaluated localization systems: the ground truth, unreliable GNSS localization, the ORB-SLAM2, and the developed fusion of the vision-based localization. Notice the gap in the GNSS poses in the left part of the plot where the green triangles are missing. For this part of the trajectory, the GNSS is temporarily unavailable.

Before the fusion, we estimate the matrices $\mathbf{Q}_{p,k}$, $\mathbf{Q}_{\Delta q,k}$, and $\mathbf{R}_k$ that describe the uncertainty of the localization systems and the uncertainty of the MAV model. The matrix $\mathbf{Q}_{p,k}$ has been estimated from the parameters of the unreliable GNSS localization with the added noise, and it reflects the GNSS availability as well. If the unreliable GNSS is disabled, $\mathbf{Q}_{p,k}$ is increased 40 times. The matrix $\mathbf{Q}_{\Delta q,k}$ has been estimated so that it models the uncertainty of the position error between two consecutive positions. Finally, the matrix $\mathbf{R}_k$ has been tuned as the diagonal matrix, which forces the filter to not rely on the velocity estimates provided by the model but on the robot position.

Four flight trials have been performed by the vehicle which follows approximately square shaped trajectory using the model predictive trajectory tracking [1]. For each trial, we obtained one trajectory by the ORB-SLAM2, and 500 different trajectories from the artificially created unreliable GNSS signal. Thus, the localization fusion has been performed 500 times per each experimental trial. All fused trajectories have been evaluated using the average of the errors com-

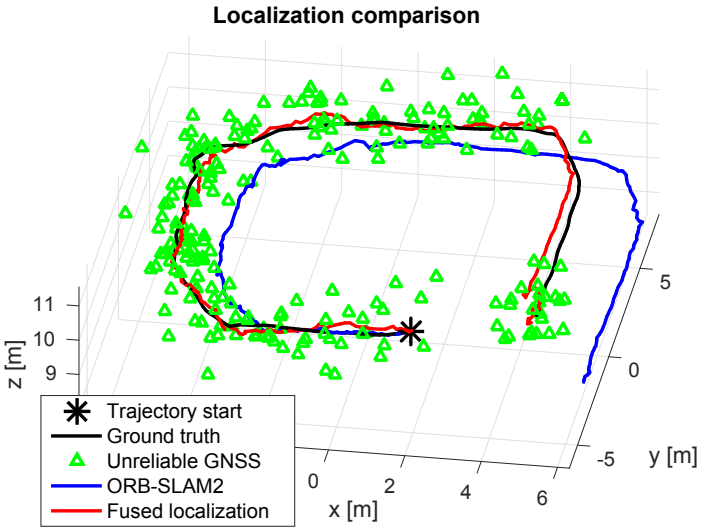


Fig. 4. Detail of the localization drift observed for the ORB-SLAM2.

puted by (3). In particular, we suppose the orientation of the vehicle is known at the beginning of each trial, which is utilized to synchronize the coordinate frame of the localization provided by the ORB-SLAM2 with the other localization systems. Since the fusion method fuses only the position of the robot and not its orientation, we add the orientation to the resulting fused trajectory from the trajectory provided by the ORB-SLAM2; so, we can use the evaluation metrics (1) and (2). The results are summarized in Fig. 2 and the ground truth, unreliable GNSS, ORB-SLAM2, and the achieved trajectory estimations provided by the fusion method are depicted in Fig. 3.

Fig. 2a indicates that the method performed as expected because the error of the localization method that employs fusion is lower than the error for both input sources of the localization, which in fact is the expected behaviour. However, the ATE of the third trial shows that the error of the resulting localization is affected by the huge drift of the incremental localization, which is further detailed in Fig. 4. On the other hand, the localization obtained by the proposed fusion method has almost 50% higher RPE than the trajectory provided by the ORB-SLAM2. Nevertheless, the RPE of the developed localization is still significantly lower than the RPE of the unreliable GNSS.

5 Conclusion

The presented results indicate the developed solution of the localization fusion provides an improved estimation of the robot pose with the ATE lower than the localization provided by the unreliable GNSS and also the pose estimation solely based on the visual localization. On the other hand, the localization obtained by the fusion method has higher the RPE than the trajectory provided by the visual

localization method. The reported results support the feasibility of the method that seems to be a vital approach to deal with a temporary absent GNSS and also with the drift of the visual localization method.

In the presented experimental scenario, the drift of the visual localization is high mainly in the position estimation. However, we expect that for a very high drift in the robot orientation, the fusion method would probably fail. Therefore, we plan to extend the solution by an additional sensor for measuring the vehicle orientation to handle the orientation drift of the visual localization.

Acknowledgement

This work has been supported by the Technology Agency of the Czech Republic (TAČR) under research Project No. TH03010362.

References

1. Báča, T., Heřt, D., Loianno, G., Saska, M., Kumar, V.: Model predictive trajectory tracking and collision avoidance for reliable outdoor deployment of unmanned aerial vehicles, `http://mrs.felk.cvut.cz/data/papers/ral2018mpctracker.pdf`, citied on 2018-10-31
2. Collective of authors: Vicon Motion Systems Inc., `https://www.vicon.com`, cited on 2018-08-05
3. Collective of autors: Intel RealSense Depth Camera D435, `https://click.intel.com/intelr-realsensetm-depth-camera-d435.html`, cited on 2018-08-04
4. Cui, J.Q., Lai, S., Dong, X., Liu, P., Chen, B.M., Lee, T.H.: Autonomous navigation of uav in forest. In: International Conference on Unmanned Aircraft Systems (ICUAS). pp. 726–733 (2014)
5. Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., Burgard, W.: An Evaluation of the RGB-D SLAM System. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 1691–1696 (2012)
6. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(3), 611–625 (March 2018)
7. Fischer, T., Pire, T., Čížek, P., De Cristóforis, P., Faigl, J.: Stereo Vision-based Localization for Hexapod Walking Robots Operating in Rough Terrains. In: IEEE International Conference on Intelligent Robots and Systems (IROS). pp. 2492–2497 (2016)
8. Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: Fast semi-direct monocular visual odometry. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 15–22 (2014)
9. Gauglitz, S., Sweeney, C., Ventura, J., Turk, M., Hllerer, T.: Live Tracking and Mapping from Both General and Rotation-only Camera Motion. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 13–22 (2012)
10. Khoury, H.M., Kamat, V.R.: Evaluation of position tracking technologies for user localization in indoor construction environments. Automation in Construction **18**(4), 649–656 (2009)
11. Lu, Y., Xue, Z., Xia, G.S., Zhang, L.: A survey on vision-based uav navigation. Geo-spatial Information Science **21**(1), 21–32 (2018). https://doi.org/10.1080/10095020.2017.1420509

12. Menze, M., Geiger, A.: Object Scene Flow for Autonomous Vehicles. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3061–3070 (2015)
13. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. IEEE Transactions on Robotics **33**(5), 1255–1262 (2017)
14. Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: a Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics **31**(5), 1147–1163 (2015)
15. Nowicki, M., Belter, D., Kostusiak, A., Čížek, P., Faigl, J., Skrzypczynski, P.: An Experimental Study on Feature-based SLAM for Multi-legged Robots with RGB-D sensors. Industrial Robot: An International Journal **44**(4), 320–328 (2017)
16. Olson, E.: AprilTag: A Robust and Flexible Visual Fiducial System. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 3400–3407 (May 2011)
17. Opromolla, R., Fasano, G., Rufino, G., Grassi, M., Savvaris, A.: Lidar-inertial integration for uav localization and mapping in complex environments. In: International Conference on Unmanned Aircraft Systems (ICUAS). pp. 444–457 (2016)
18. Piasco, N., Sidib, D., Demonceaux, C., Gouet-Brunet, V.: A survey on visual-based localization: On the benefit of heterogeneous data. Pattern Recognition **74**, 90–109 (2018)
19. Pire, T., Fischer, T., Civera, J., De Cristóforis, P., Jacobo Berlles, J.: Stereo Parallel Tracking and Mapping for Robot Localization. In: IEEE International Conference on Intelligent Robots and Systems (IROS). pp. 1373–1378 (2015)
20. Quigley, M., Conley, K., Gerkey, B.P., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y.: ROS: an Open-source Robot Operating System. In: IEEE International Conference on Robotics and Automation (ICRA): Workshop on Open Source Software (2009)
21. Stelzer, A., Hirschmller, H., Grner, M.: Stereo-vision-based navigation of a six-legged walking robot in unknown rough terrain. The International Journal of Robotics Research **31**(4), 381–402 (2012)
22. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A Benchmark for the Evaluation of RGB-D SLAM Systems. In: IEEE International Conference on Intelligent Robots and Systems (IROS). pp. 573–580 (2012)
23. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT press (2005)
24. Toloei, A., Niazi, S.: State estimation for target tracking problems with nonlinear kalman filter algorithms. International Journal of Computer Applications **98**(17), 30–36 (July 2014), full text available
25. Usenko, V., Engel, J., Stckler, J., Cremers, D.: Direct visual-inertial odometry with stereo cameras. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 1885–1892 (2016)
26. Wang, C., Wang, T., Liang, J., Chen, Y., Zhang, Y., Wang, C.: Monocular visual slam for small uavs in gps-denied environments. In: IEEE International Conference on Robotics and Biomimetics (ROBIO). pp. 896–901 (2012). https://doi.org/10.1109/ROBIO.2012.6491082
27. Wang, F., Cui, J.Q., Chen, B.M., Lee, T.H.: A comprehensive uav indoor navigation system based on vision optical flow and laser fastslam. Acta Automatica Sinica **39**(11), 1889–1899 (2013)