Vision-based Localization for Multi-rotor Aerial Vehicle in Outdoor Scenarios

Jan Bayer[0000−0003−1190−1085] 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 the experimental evaluation of the embedded visual localization system, the Intel RealSense T265, deployed on a multi-rotor unmanned aerial vehicle. The performed evaluation is targeted to examine the limits of the localization system and discover its weak points. The system has been deployed in outdoor rural scenarios at altitudes up to 20 m. The Absolute trajectory error measures the accuracy of the localization with the reference provided by the differential GPS with centimeter precision. Besides, the localization performance is compared to the state-of-the-art feature-based visual localization ORB-SLAM2 utilizing the Intel RealSense D435 depth camera. In both types of experimental scenarios, with the teleoperated and autonomous vehicle, the identified weak point of the system is a translation drift. However, taking into account all experimental trials, both examined localization systems provide competitive results.

1 Introduction

A precise and reliable localization system is necessary for many robotics and related applications, including mapping, augmented reality, and fully autonomous deployments of mobile robots. In this work, we consider the mobile robotics domain with the primary focus on localization systems for Unmanned Aerial Vehicles (UAVs) in applications such as autonomous navigation, exploration, or perimeter monitoring, where a full 6 DOF localization is required. The existing localization solutions can be broadly divided into two classes. The first class contains systems that require external infrastructures, such as the Global Navigation Satellite System (GNSS) [13] or optical systems operating on the line of sight [6,17,7]. The great advantage of these localization systems is the limited accumulation of localization errors, even in large-scale scenarios. However, these methods are limited by the needed infrastructure, particularly not suitable for large scale indoor environments without prior preparations or scenarios with many obstacles shading signal or line of sight. On the other hand, systems of the second class rely only on sensors mounted on the robot. These methods are much more suitable for unknown environments, where external infrastructure cannot

Fig. 1. The unmanned aerial vehicle DJI Matrice 600, which was deployed during the experimental evaluation.

be prepared in advance. Such localization systems mainly use Light Detection and Ranging (LiDAR) sensors [22] and different types of cameras [11].

In recent years, sensors for visual localization have been introduced, including the Intel RealSense T265 tracking camera [15] that is referred to as the T265 for short in the rest of the paper. The T265 provides image data, and it is also capable of processing the data by its embedded visual processing unit Movidius Myriad 2.0, capable of doing all the computations needed for visual localization. The power effectiveness is a great advantage of embedded solutions for the localization since the visual localization algorithm might run on dedicated computational resources of the camera itself, saving the main onboard computational resources of the UAV. The saved computational power can be thus utilized for navigation and real-time mapping tasks, or even more sophisticated tasks like autonomous exploration. The features and benefits of embedded solutions based on the T265 and recent successful deployments on the ground legged walking robots [4,16] motivate us to examine its performance for localization of small UAV, see Fig. 1. Thus, we evaluate the T265 as a vision-based localization system that uses affordable off-the-shelf lightweight cameras. Besides, we compare the performance to the state-of-the-art localization method ORB-SLAM2 [20].

The rest of the paper is organized as follows. Principles and related evaluations of the visual localization methods are presented in Section 2. The evaluation method utilized for the evaluation of the visual localization is overviewed in Section 3. The resulting localization precision measured during the deployments is reported in Section 4. The concluding remarks on the achieved localization precision and limits of the embedded localization system are presented in Section 5.

2 Related Work

Two recent compact sensors for vision-based localization are available on the market nowadays: the ZED mini [27] and Intel RealSense T265 [15], shown in Fig. 2a. Both of them are passive fisheye stereo cameras. ZED mini primary targets augmented reality applications, and it requires external computational power for GPU-based acceleration of the visual localization algorithms. On the other hand, the T265 runs all the computations onboard using a power-efficient visual processing unit, which can save the computational and power resources such, e.g., as reported in [4]. The T265 is thus a suitable choice for embedded localization for small robotic vehicles such as multi-rotor UAVs.

(a) Intel RealSense T265. (b) Intel RealSense D435.

Fig. 2. Cameras used for the visual localization.

The parameters of the T265 are promising, but the experimental evaluation is important for the verification of the localization performance in real-world scenarios because the real localization performance is affected by multiple factors that are related to the

- precision of the localization algorithm itself;
- properties of the environment;
- sensory equipment;
- computational resources (if limited);
- and motion of the platform with the sensors.

The experimental evaluations (to the best of the authors' knowledge) have been reported only for indoor conditions or carried by humans [23,2,18,5,1]. In these scenarios, the T265-based localization system is reported to with satisfactory performance. Thus, in the herein presented results, we aim to push the localization system to its limits. We deploy the system in real-world outdoor scenarios with the DJI Matrice 600 operating in a rural-like environment. The precision of the localization is measured as in [23] using the well-established metric of the Absolute Trajectory Error (ATE) [28].

Besides, a state-of-the-art vision-based localization method has been selected to provide a baseline solution for the selected scenarios using a traditional approach with cameras and data processing on dedicated standard computational resources. There are many different visual localization approaches mentioned in the literature. One of the differences between the methods is in the required sensory equipment. Techniques such as [9,10,21] use monocular cameras. Other

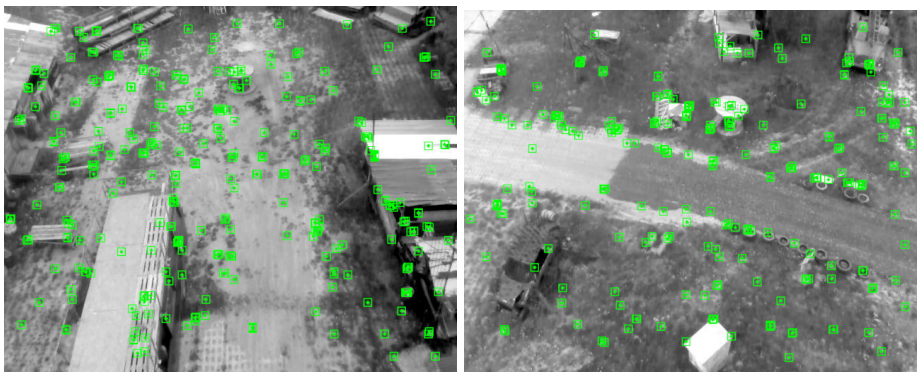


Fig. 3. Examples of image features detected by ORB-SLAM2 [20].

approaches use stereo cameras [29,24] and RGB-D cameras [8,30]. The stereo and RGB-D cameras can be considered more sophisticated than monocular cameras regarding the sensory equipment, and their advantage is the reduced drift of the map scale. Besides, the map initialization is easier as bootstrapping is avoided [12]. Thus, we focus on the stereo and RGB-D localization methods capable of being used with various sensors.

A representative approach is the ORB-SLAM2 [20], a state-of-the-art publicly available localization method, reported to perform well on datasets [19,28,26] and often used as the baseline for comparing different localization approaches [2]. Based on our previous work [4,3], we have chosen to use the feature-based localization method ORB-SLAM2 together with the RGB-D camera Intel RealSense D435 [14], shown in Fig. 2b. An example of the detected ORB features used by ORB-SLAM2 is in Fig. 3.

3 Evaluation Method for Localization Systems

In the literature, the localization systems are compared by the precision of trajectories obtained during an experiment. A comparison of the localization systems solely based on the trajectories allows us to compare black-box localization systems like the T265, where a detailed description of the particular algorithms is not provided. For such purposes, there is a well-established metric of the *Absolute Trajectory Error* (ATE) [28]. A ground truth trajectory and trajectories estimated by the localization systems under the examination are required to compute the ATE. Besides, it is also necessary to have the trajectory estimate and ground truth with poses that correspond to the same timestamps. Therefore, the trajectories have to be time-synchronized using interpolation or by finding the nearest neighbor [28] if the ground truth trajectory is provided with a different frequency than the trajectory estimate. In this work, the linear interpolation of the positions and *Linear Quaternion Interpolation* (LERP) of the orientation represented by the quaternions is utilized.

Once the trajectories are time-synchronized, they are processed by the ATE metrics. The ATE is defined in [28] 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) pose 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. According to [28], the transformation is obtained by minimizing the squared distances between the corresponding positions of the trajectory estimate and the ground truth.

The average value is used to generate a statistical indicator from the error for the whole trajectory

$$\overline{\mathrm{ATE}}_t = \frac{1}{n}\sum_{i=1}^{n} \parallel trans(\mathbf{F}_i) \parallel, \tag{2}$$

where $trans()$ computes the size of the translation from the SE(3) matrix. In the results reported in Section 4, we assume only the translation errors because the ground truth is provided in 3 DOF only.

4 Results

The UAV has been experimentally deployed in the outdoor environment of the rural deployment scenario shown in Fig. 4. The UAV was flying in the experimental setup at different altitudes: 5, 10, 15, and 20 m driven manually by a human operator and autonomously by the DJI autopilot, to verify the ability of the localization system to work with the absence of close objects while being exposed to different motions. Two pairs of the Intel RealSense cameras were mounted on the UAV. The first pair is pointed to the front tilted by 45°, the second pair is oriented to the rear side and tilted as well, see Fig. 5.

Data from all the sensors were collected using the ROS middleware [25]. The localization from tracking cameras is captured directly during the experiments together with the RGB-D data. The localization provided by the ORB-SLAM2 was generated by processing the RGB-D data from the rosbag dataset captured during the experiment. The ROS rosbag captures data incoming from the sensors as provided by the sensors. It also enables processing the data at different speeds to simulate different computational power. Thus, the RGB-D data were processed by the ORB-SLAM2 at two different speeds. The first processing speed is to simulate online processing, and it is denoted *online*-speed. The second speed is denoted *half*-speed, and it corresponds to two times more computational time for processing the captured images than processing them in real-time. The ORB-SLAM2 was run on the regular computer with the Intel i5-5257U processor running at 2.7 GHz with 4 GB of memory. The Differential GPS provided the ground truth trajectory with a centimeter precision at 10 Hz.

Fig. 4. The deployment scenario in the rural-like environment.

Fig. 5. The sensory equipment used during the experiments attached to the UAV; the first pair of the Intel RealSense cameras is pointed to the front. The second pair is pointed to the rear side.

4.1 Localization error

The mean ATE for all experimental trials is summarized in Table 1. Each experimental trial contains several circular flights that are illustrated on Trial 1 in Fig. 6. The results show that the ORB-SLAM2 provides the best results with the front camera in Trial 1 and Trial 3. On the other hand, the T265 provided better results in Trial 2 and Trial 4. In the case the UAV was teleoperated manually, the angular velocity of the helicopter was always under $30\,^\circ\,\mathrm{s}^{-1}$. In Trial 2, the UAV flew autonomously between preselected waypoints. On each waypoint, the UAV turned with an angular velocity above $30\,^\circ\,\mathrm{s}^{-1}$, which induced failure of the localization based on the ORB-SLAM2 in combination with the front D435 camera. Trial 2 is visualized in Fig. 7. It is possible to overcome the localization failure using more computational power, which can be observed for *half*-speed

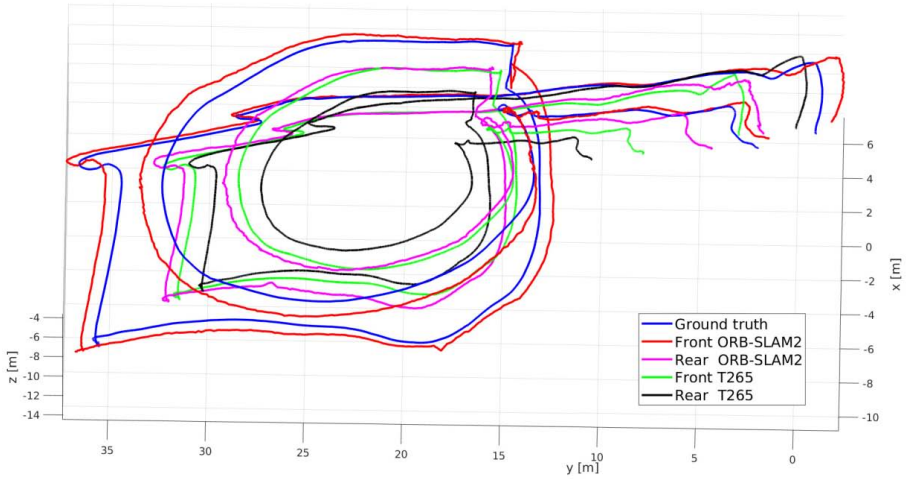


Fig. 6. Trial 1: Two circular flights.

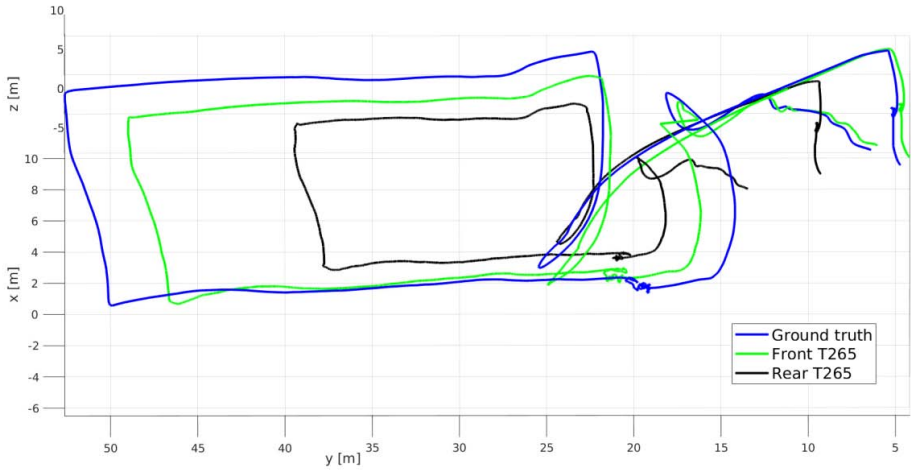
results. However, the drift at the corners of the trajectory is still very high; see Fig. 7b. Contrary, the effect of the high rotational speed on the localization quality is not observed for the T265.

The relatively small localization error provided by the T265 can be observed for Trial 4 shown in Fig. 8, where the UAV flew at altitudes of 10 m, 15 m, and 20 m. Based on the deformation of the trajectory estimated by the T265, shown in Fig. 7a and Fig. 8, it can be observed that the localization suffers mostly from the translation drift, not the orientation drift.

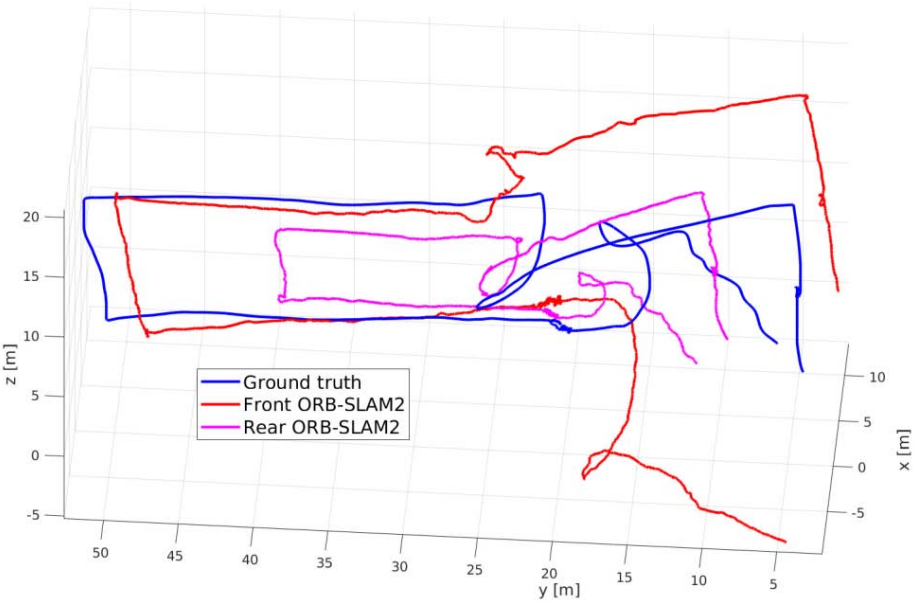
Table 1. The mean absolute trajectory error for front and rear setup in all four experimental trials.

Trial	**Flight mode**	**Length**	**T265**		**ORB-SLAM2**			
					online-speed		*half*-speed	
		[m]	Front	Rear	Front	Rear	Front	Rear
Trial 1	Manual	174	3.01	4.01	0.75	2.41	**0.69**	2.38
Trial 2	Autonomous	170	**1.57**	4.63	lost	4.12	7.12	4.58
Trial 3	Manual	596	3.84	6.37	1.92	4.09	**1.63**	3.07
Trial 4	Manual	1175	**0.91**	6.14	1.96	4.92	4.51	3.77

All results provided by the ORB-SLAM2 were evaluated on a computer with the Intel i5-5257U and 4 GB of memory. *Half*-speed means that the ORB-SLAM2 has twice more time to process the incoming data than in the *online* case.



(a) Intel RealSense T265

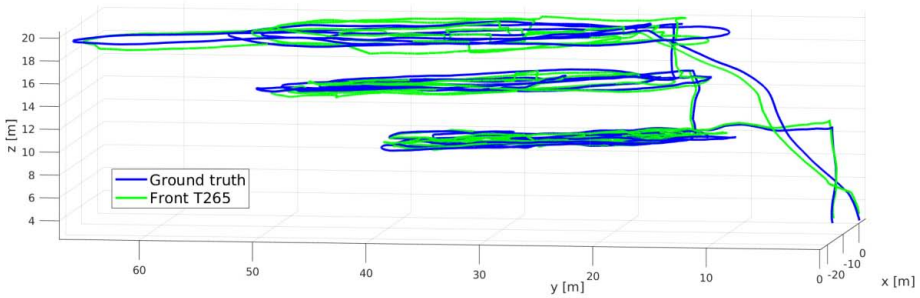


(b) ORB-SLAM2

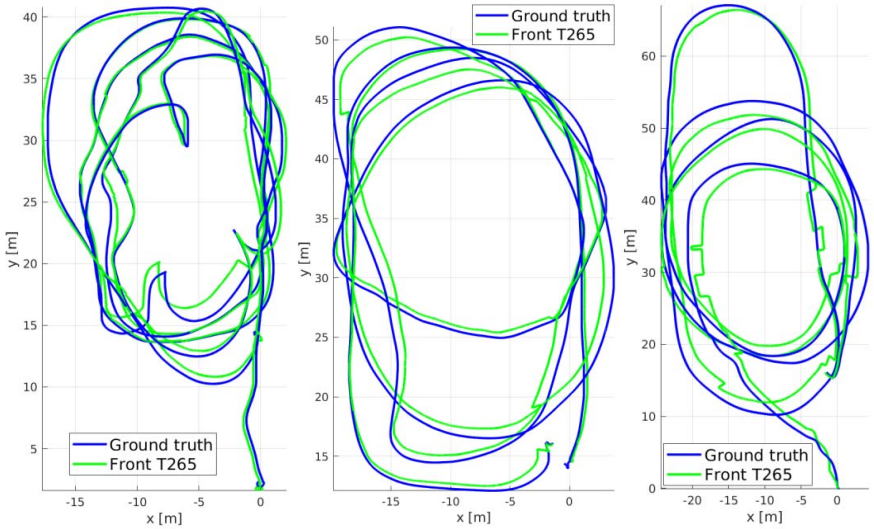
Fig. 7. Trial 2: UAV trajectories obtained during the autonomous flight.

4.2 Remarks on Practical Usage of the Intel RealSense Cameras

The deployed sensor system contained the two T265 and two D435 cameras, connected to a single Intel NUC computer. Generally, the USB 3 bandwidth of various NUC computers is sufficient to stream data from more than two pairs of Intel RealSense cameras. However, the camera driver for Ubuntu with the



(a) Side view of the whole trajectory.



(b) Altitude 10 m. (c) Altitude 15 m. (d) Altitude 20 m.

Fig. 8. Trial 4: Visualization of the localization drift for different altitudes during the long UAV flight. The places where the T265 induced fast changes of the estimated pose are visible at the altitude of 20 m; this phenomenon is observed from 10 m altitude for the rear T265 camera.

ROS[1] requires a particular launch sequence to detect all the cameras correctly. The cameras have been therefore launched one-by-one in a fixed order with more than 20 s delay. Besides, resetting USB hubs before starting the launch sequence has been found necessary.

5 Conclusion

The examined embedded localization system Intel RealSense T265 provided competitive results in the realistic outdoor deployment scenario to the state-of-the-

[1] Available at https://github.com/IntelRealSense/realsense-ros.

art localization method ORB-SLAM2. In half of the experimental trials, the T265 performed even better than the ORB-SLAM2. Especially in the trial, where the UAV was controlled autonomously with an angular velocity above $30\,^{\circ}\,s^{-1}$ at corners of the trajectory. The effect of the high angular velocity of the UAV has not been observed for the T265, which makes it superior to the ORB-SLAM2 in applications where such motions are required. Moreover, the flight altitude effect on the localization performance was for the front T265 camera observed only at altitudes above 15 m. On the other hand, in half of the trials, T265 suffered from translation drift. For both localization systems, the front UAV cameras provided better results in nearly all cases.

Most of the localization error is accumulated at the start of the trajectory. Therefore, we aim to investigate the impact of the takeoff on the localization system and the consequences of the camera orientation in future work.

Acknowledgement

The presented work has been supported by the Technology Agency of the Czech Republic (TAČR) under research Project No. TH03010362 and under the OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 "Research Center for Informatics". The support under grant No. SGS19/176/OHK3/3T/13 to Jan Bayer is also gratefully acknowledged.

References

1. Agarwal, A., Crouse, J.R., Johnson, E.N.: Evaluation of a commercially available autonomous visual inertial odometry solution for indoor navigation. In: 2020 International Conference on Unmanned Aircraft Systems (ICUAS). pp. 372–381 (2020). DOI: 10.1109/ICUAS48674.2020.9213962
2. Alapetite, A., Wang, Z., Hansen, J., Zajączkowski, M., Patalan, M.: Comparison of three off-the-shelf visual odometry systems. Robotics **9**, 56 (07 2020). DOI: 10.3390/robotics9030056
3. Bayer, J., Faigl, J.: Localization fusion for aerial vehicles in partially gnss denied environments. In: mesas. pp. 251–262 (2019). DOI: 10.1007/978-3-030-14984-0_20
4. Bayer, J., Faigl, J.: On Autonomous Spatial Exploration with Small Hexapod Walking Robot using Tracking Camera Intel RealSense T265. In: European Conference on Mobile Robots (ECMR). pp. 1–6 (2019). DOI: 10.1109/ECMR.2019.8870968
5. Bayer, J., Faigl, J.: Handheld localization device for indoor environments. In: International Conference on Automation, Control and Robots (ICACR) (2020)
6. Collective of authors: Vicon motion systems inc. https://www.vicon.com/ (accessed 1 8, 2020)
7. Collective of authors: Leica geosystems ag @ONLINE. http://leica-geosystems.com/products/total-stations (accessed 3 3, 2020)
8. Endres, F., Hess, J., Sturm, J., Cremers, D., Burgard, W.: 3-D Mapping with an RGB-D Camera. IEEE Transactions on Robotics **30**(1), 177–187 (2014). DOI: 10.1109/TRO.2013.2279412

9. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **PP** (07 2016). DOI: 10.1109/TPAMI.2017.2658577
10. 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). DOI: 10.1109/ICRA.2014.6906584
11. Fuentes-Pacheco, J., Ruiz-Ascencio, J., Rendón-Mancha, J.M.: Visual simultaneous localization and mapping: a survey. Artificial intelligence review **43**(1), 55–81 (2015). DOI: 10.1007/s10462-012-9365-8
12. Gauglitz, S., Sweeney, C., Ventura, J., Turk, M., Höllerer, 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). DOI: 10.1109/ISMAR.2012.6402532
13. Hofmann-Wellenhof, B., Lichtenegger, H., Wasle, E.: GNSS – Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more. Springer Vienna (2007), https://books.google.cz/books?id=Np7y43HU_m8C
14. Intel RealSense Depth Camera D435. https://www.intelrealsense.com/depth-camera-d435/, accessed Aug 4, 2020
15. Intel RealSense Tracking Camera T265. https://www.intelrealsense.com/tracking-camera-t265/, accessed Aug 4, 2020
16. Kim, D., Carballo, D., Di Carlo, J., Katz, B., Bledt, G., Lim, B., Kim, S.: Vision aided dynamic exploration of unstructured terrain with a small-scale quadruped robot. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 2464–2470 (2020). DOI: 10.1109/ICRA40945.2020.9196777
17. Lightbody, P., Krajník, T., Hanheide, M.: A versatile high-performance visual fiducial marker detection system with scalable identity encoding. In: Proceedings of the Symposium on Applied Computing. pp. 276–282. ACM (2017)
18. Mahmoud, A., Atia, M.M.: Hybrid imu-aided approach for optimized visual odometry. In: 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). pp. 1–5 (2019). DOI: 10.1109/GlobalSIP45357.2019.8969460
19. Menze, M., Geiger, A.: Object Scene Flow for Autonomous Vehicles. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3061–3070 (2015). DOI: 10.1109/CVPR.2015.7298925
20. 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). DOI: 10.1109/TRO.2017.2705103
21. 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). DOI: 10.1109/TRO.2015.2463671
22. 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). DOI: 10.1109/ICUAS.2016.7502580
23. Ouerghi., S., Ragot., N., Boutteau., R., Savatier., X.: Comparative study of a commercial tracking camera and orb-slam2 for person localization. In: Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP,. pp. 357–364. INSTICC, SciTePress (2020). DOI: 10.5220/0008980703570364
24. Pire, T., Fischer, T., Castro, G., Cristóforis, P.D., Civera, J., Berlies, J.J.: S-PTAM: Stereo parallel tracking and mapping. Robotics and Autonomous Systems **93**, 27–42 (2017), http://www.sciencedirect.com/science/article/pii/S0921889015302955

25. 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: ICRA Workshop on Open Source Software (2009)
26. Smith, M., Baldwin, I., Churchill, W., Paul, R., Newman, P.: The new college vision and laser data set. The International Journal of Robotics Research **28**(5), 595–599 (2009). DOI: 10.1177/0278364909103911
27. ZED Mini. https://www.stereolabs.com/zed-mini/, accessed Aug 4, 2020
28. 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). DOI: 10.1109/IROS.2012.6385773
29. Usenko, V., Engel, J., Stückler, J., Cremers, D.: Direct visual-inertial odometry with stereo cameras. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 1885–1892 (2016). DOI: 10.1109/ICRA.2016.7487335
30. Zhou, Y., Kneip, L., Li, H.: Semi-dense visual odometry for rgb-d cameras using approximate nearest neighbour fields. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp. 6261–6268 (2017). DOI: 10.1109/ICRA.2017.7989742