A TEST ON COLLABORATIVE VISION AND UWB-BASED POSITIONING

A TEST ON COLLABORATIVE VISION AND UWB-BASED POSITIONING

M. Gurturk[1], A. Masiero[2]*, C. Toth[3], P. Dabove[4], V. Di Pietra[4], A. Vettore[5], A. Guarnieri[5], I. Cortesi[2], E. Pellis[2], M. Soycan[1]

[1] Yildiz Technical University, Istanbul, Turkey - mgurturk@yildiz.edu.tr, soycan@yildiz.edu.tr
[2] University of Florence, Italy - andrea.masiero@unifi.it, irene.cortesi@unifi.it, eugenio.pellis@unifi.it
[3] The Ohio State University, US - toth.2@osu.edu
[4] Polytechnic of Turin, Italy - paolo.dabove@polito.it, vincenzo.dipietra@polito.it
[5] University of Padua, Italy - alberto.guarnieri@unipd.it, antonio.vettore@unipd.it

KEY WORDS: Collaborative positioning, SLAM, Vision, UWB

ABSTRACT:

Despite GNSS (Global Navigation Satellite System) enables positioning, navigation and timing (PNT) almost everywhere, the development of applications like self-driving vehicles and indoor navigation requires extending accurate positioning to scenarios where GNSS either is not reliable or does garantee a sufficiently precise solution. Integrating inforamtion provided by different sensors is commonly accepted to be a quite viable way for such extension. In particular this work is part of a project aiming at investigating the positioning performance that can be obtained by integrating vision with radio-based systems and inertial sensors, which are commonly installed on many smart devices, such as smartphones. Furthermore, this work considers positioning in a collaborative scenario, where different interconnected platforms, i.e. unmanned aerial vehicles and pedestrians provided with smartphones, are moving on the same area. The results obtained in the considered tests show a good potential (submetric 2D positioning error) for what concerns the implemented strategies, where the integration of different technologies can ensure decent performance in a wider range of working cases.

1. INTRODUCTION

This work is part of a project conducted as a collaboration between different research groups participating to the IAG Working Group 4.1.4 "Computer Vision in Navigation". In particular, the goal is that of investigating the performance of a multi-sensor system approach in collaborative positioning, focusing in particular on the role and potential of sensor integration and vision in order to compensate the unreliability and/or unavailability of GNSS.

Nowadays, vision-based positioning methods are quite commonly used, in particular, when dealing with real-time applications, for what concerns visual simultaneous localization and mapping (SLAM) techniques (Mur-Artal et al., 2015, Leonard and Durrant-Whyte, 1991, Strasdat et al., 2012, Whelan et al., 2016, Kukko et al., 2017). In off-line applications such methods typically lead to slightly different approaches, such as Structure from Motion photogrammetry (Förstner and Wrobel, 2016). Furthermore, LiDAR has also been employed in SLAM-like approaches, i.e. to determine the device ego-motion (Zhang and Singh, 2014, Zhang and Singh, 2017) and to map the environment in the neighborhood of the device.

Despite vision has already been widely exploited in positioning applications, it is still a very active research field, mostly in order to make its performance more robust with respect to the operating conditions, e.g. overcoming the issues related to illumination and weather conditions, reducing the drift. Furthermore, a number of works recently considered the collaborative case: the cooperating agents are supposed to be interconnected, e.g. provided with some communication abilities, typically radio-based. Thanks to such communication abilities, each agent can exploit the information shared by the others when computing its own solution (Tong et al., 2023). For instance, sharing (at least part) of their knowledge the multiple agents increase the loop-closure probability (intra-camera and/or inter-camera loop closure), hence potentially reducing the overall mapping and localization error (Zou et al., 2019). Such collaborative approaches can either be centralized (Karrer et al., 2018, Schmuck and Chli, 2017, Schmuck et al., 2021) or decentralized (Cieslewski et al., 2018). Despite centralized solutions are expected to generally perform better, decentralized ones ensure a better scalability, i.e. they can be used also when the number of interacting agents is large. In a number of works, deep learning methods have also been employed in order to improve vision based tracking (Gurturk et al., 2021) (also with LiDAR (Chiang et al., 2020, Iwaszczuk et al., 2021)).

Among the sensors that can be installed on platforms and smart devices, inertial ones are very convenient in terms of cost, for what concerns MEMS ones, and availability of the solution. Indeed, differently from most of the other sensors, inertial ones, provided by an IMU (Inertial Measurement Unit), can be used in practically any working condition, hence always guaranteeing the availability of a solution. Nevertheless, solutions based only on the use of affordable inertial sensors usually quickly drift from the correct positions. In this work, a pedestrian dead reckoning approach has been implemented. The reader is referred to (Masiero and Vettore, 2016) and the reference therein in order to have a more complete overview on pedestrian dead reckoning.

In addition to vision and IMU, radio based measurements, and in particular Ultra-Wide Band (UWB) ones, are considered as well. While IMU and vision are often used in order to determine the platform ego-motion, in terms of relative movements with respect to its previous position, UWB ranging can be used both for position fixing, when a static architecture of UWB nodes of known position is available, or for ranging between different platforms. The first case will be named vehicle-to-infrastructure

* Corresponding author

hereafter, and, in particular, pedestrian-to-infrastructure when the considered agent is just a pedestrian, whereas the latter will be named vehicle-to-vehicle ranging, or, more precisely, either pedestrian-to-pedestrian or pedestrian-to-drone in the cases of interest for this paper.

Hence, this work aims at exploiting IMU, vision and UWB ranging in order to compute a robust solution. The combined use of vision and UWB ranging has already been partially investigated in (Masiero et al., 2021b), for what concerns the collaborative positioning of terrestrial vehicles. Instead, this work mostly focuses on pedestrian collaborative positioning, similarly to (Retscher et al., 2020). Nevertheless, differently from the latter, this paper considers the simultaneous availability also of measurements from Unmanned Aerial Vehicles (UAVs).

2. CASE STUDY

In the considered test case three pedestrians and two UAVs moved simultaneously on the same area, mostly visible on Figure 1. Despite the overall test lasted for around twenty minutes, this paper focuses on the analysis of a 2-minute interval extracted from such dataset.

Table 1 summarizes the platforms and the data acquired during the test. All the pedestrians were provided with a geodetic GNSS receiver (see Figure 2) of an UWB Pozyx transceiver and of a smartphone, collecting a video. Pedestrian 1 also collected inertial data by means of the affordable IMU mounted on his Pozyx rover. An example of video frame collected by the smartphone of one of the pedestrians is shown on Figure 3.

Platform	Data acquired
Pedestrian 1	GNSS, IMU, video, UWB
Pedestrian 2	GNSS, video, UWB UWB
Pedestrian 3	GNSS, video, UWB UWB
UAV 1 (DJI Matrice 210)	GNSS, video, UWB
UAV 2 (DJI Phantom 4 Pro)	GNSS, video

Table 1. Summary of the platforms and data acquired during the data collection campaign.

Figure 1. Top view of the central part of the test area: pedestrians (red circles), targets (light blue circles).

Both the UASs collected GNSS measurements (in particular, an external GNSS receiver was mounted on the DJI Matrice 210, as shown in Figure 4) and videos (Figure 1 shows a drone view of the area, where pedestrians (red circles) and targets (light blue circles) are also visible). DJI Matrice 210 was also provided with a Pozyx UWB transceiver.

Figure 2. Sensors carried by pedestrians.

Figure 3. Example of pedestrian's smartphone view.

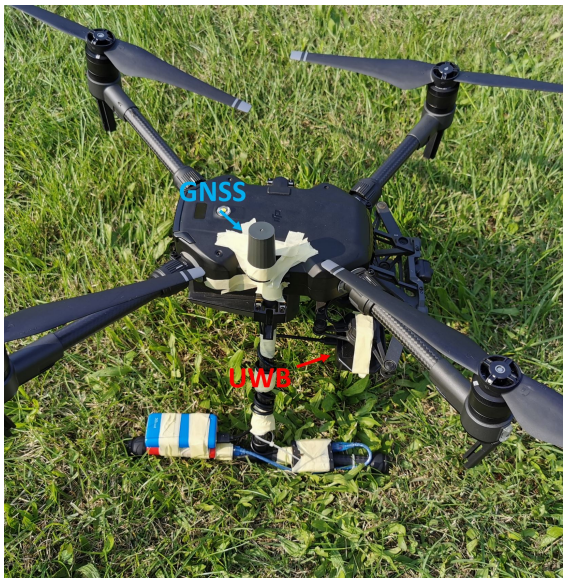

Figure 4. Sensors on DJI Matrice 210.

In addition to the UWB transceivers mounted on the moving agents a static infrastructure of eight UWB nodes (also

named anchors), with known positions, has also been deployed. Hence, three types of UWB ranging measurements can be distinguished: pedestrian-to-static infrastructure of UWB anchors (P2I), pedestrian-to-pedestrian (P2P) and pedestrian-to-aerial drone (P2A). Figure 5 shows how the percentage of UWB measurements were distributed among P2I, P2P and P2A. The different distribution of the UWB measurements among P2I, P2P and P2A is motivated by several factors. Among them, two major ones are: (i) the different number of pedestrians, UASs and anchors and (ii) the ranging success rate, which varies depending on the working conditions.

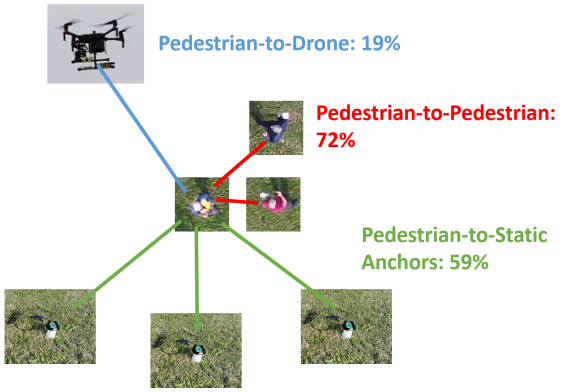

Figure 5. success rate of UWB ranging measurements among pedestrian-to-infrastructure, pedestrian-to-pedestrian and pedestrian-to-drone.

Indeed, while GNSS, video and IMU data were acquired at constant rate during the all test, the UWB ranging system, which was also active during the all test, provided ranging at a variable success rate. In particular, the UWB ranging success rate depends on the distance between the involved devices, as shown in Figure 6 and in (Masiero et al., 2021b). Furthermore, Figure 7 shows the distribution of the UWB ranging error while varying the distance between the transceivers. The figure also shows that the ranging uncertainty tends to slightly increase with the distance, probably also due to the higher chance of the presence of obstructions on the line of sight between two devices increasing the distance between the two of them.

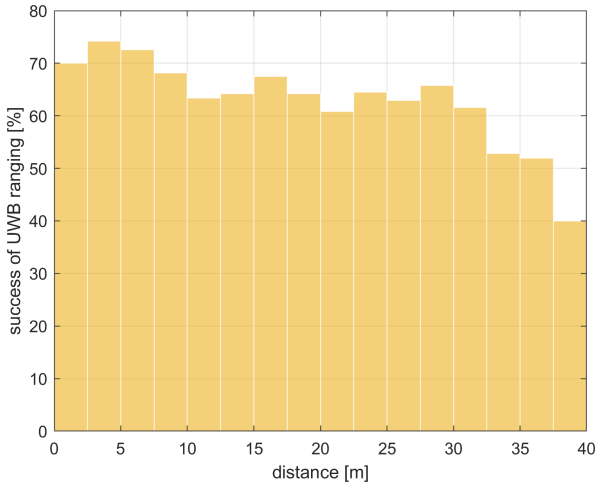

Figure 6. Success rate of Pedestrian-to-Infrastructure UWB ranging measurements varying the distance (2-minute dataset).

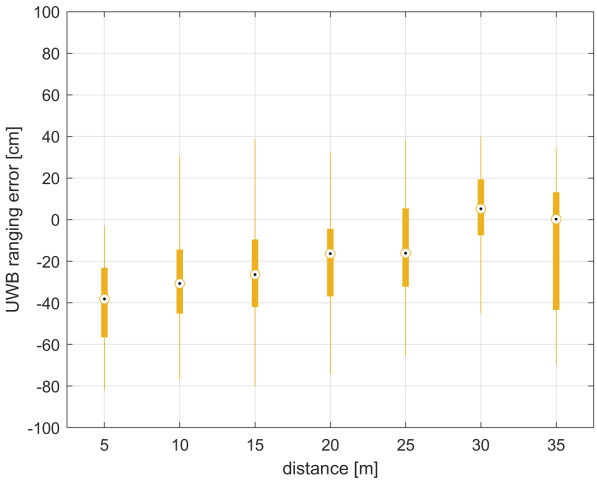

Figure 7. Ranging error of Pedestrian-to-Infrastructure UWB measurements varying the distance (2-minute dataset).

3. METHODS

This paper compares three different strategies to assess the pedestrian positions:

1. UWB P2I: pedestrian positions are based on the deployment of an Extended Kalman filter on each agent. This approach can extended to a collaborative one if considering P2P ranging as well (and deploying a proper centralized/distributed approach).

2. UWB + IMU/Smartphone vision: dead reckoning solution can be derived from UWB/vision, then a distributed Extended Kalman filter is employed for integrating UWB ranges from the infrastructure and from the other agents in such a way to track agents in the desired coordinate reference system.

3. UAS vision: in this approach pedestrians are tracked on the UAS video frames and their position in a properly defined reference system is determined thanks to the targets positioned in the considered area (Figure 1).

Approach 1 is based on the implementation of an Extended Kalman filter (EKF), where one of its instances runs separately for each pedestrian. The approach is briefly described below, however, the reader is referred to for a general introduction on EKF (Anderson and Moore, 2012, Jazwinski, 2007, Kailath et al., 2000). The state vector for each agent is defined as the current position $\mathbf{p}$ and velocity $\mathbf{v}$. The state evolution is described as $\mathbf{x}_{k+1} = F_k \mathbf{x}_k + \omega_k$, with $F_k = \begin{bmatrix} I & \Delta t_k I \\ 0 & I \end{bmatrix}$, and being $\Delta t_k = t_k - t_{k-1}$, the time difference between two iterations of the Kalman filter, ω_k a zero-mean Gaussian-distributed white noise, with covariance matrix Q_k.

The measurement equation at time t_k for the UWB ranging between the considered agent and the j-th anchor, with position $\mathbf{p}^j$, is

$$z_k^j = \left| \mathbf{p}(t_k) - \mathbf{p}^j \right| + \xi_k \tag{1}$$

where ξ_k is assumed to be a Gaussian-distributed zero-mean white noise process, with variance σ_ξ^2. Once linearized, the

above equation leads to the following line to be inserted in the observation matrix H_k at time t_k:

$$\mathbf{h}_k^j = \frac{\left(\mathbf{p}_{k|k-1}^j - \mathbf{p}^j\right)^\top}{z_k^j} \tag{2}$$

Then, EKF is implemented straightforwardly when dealing with P2I measurements only.

The considered collaborative implementation, as in Approach 2, inter-agent UWB ranging is considered as well. In this case, distributed solution is computed by allowing communications between each agent and those in its neighborhood, which at every step are supposed to share their current state value (and clearly participating to the UWB ranging measurement process). Furthermore, prediction step is now obtained by considering the state evolution determined by exploiting IMU and/or SLAM estimates (in particular ORB-SLAM3, (Mur-Artal et al., 2015, Campos et al., 2021)). It is worth to notice that SLAM intrinsically provides a scale-free movement assessment. For what concerns IMU-based pedestrian dead reckoning, despite the movement scale (e.g. the step length) can be assessed, errors accumulating due to such uncertainties on such estimates usually quite quickly becomes large when alternative measurements are not available to reset them. Consequently, updates provided by UWB measurements play a fundamental role for drift reduction and for assessing the movement scale.

The UAS vision-based approach (Approach 3) is grabbed from (Masiero et al., 2021a): pedestrian detection on each frame is based on the implementation of a background subtraction method combined with a correlation maximization, based on the availability of a previously derived template for the pedestrians. Once that pedestrians are found on a UAS video frame, then, their coordinates in the adopted coordinate system are obtained thanks to the use of a set of target in the area of interest. The reader is referred to (Masiero et al., 2021a) for more details on such approach.

4. RESULTS AND DISCUSSION

The pedestrian 2D positioning results obtained in a 2-minute portion of the collected dataset are shown in Figure 8, 9 and 10 for what concerns approach 1, 2 and 3, respectively.

First, from the obtained results it is quite apparent that all the considered strategies allowed to obtain submetric 2D error in the considered test. The performance of the first two approaches is quite similar: the median error in approach 2 is just 1 cm smaller than that of approach 1, but overall the error statistics are quite similar, showing the predominant role in this test of the UWB static network, which in this case covered quite well the area of interest. Nevertheless, the integration of additional sensors (e.g. IMU, camera) can be very useful in order to make the overall system more robust when the UWB network coverage is less good, allowing to temporarily compensate its unreliability or unavailability.

Figure 10 shows that the UAS vision approach allows to obtain the best results among the tested strategies. This is mostly due to the fact that almost all the test area was visible in most of the DJI Matrice 210 video frames, hence allowing to continuously track the pedestrians. However, despite such good

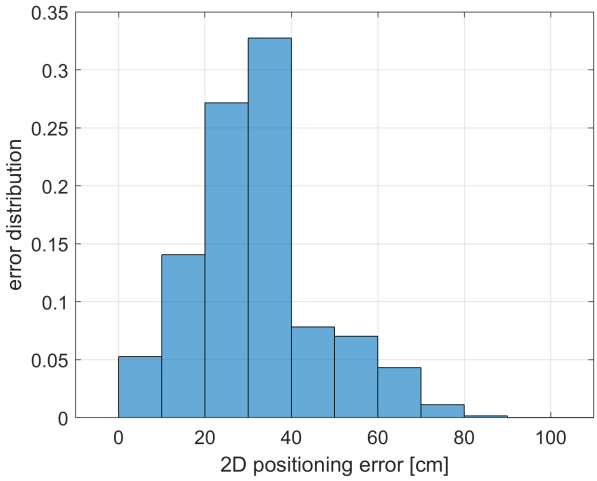

Figure 8. UWB pedestrian-to-infrastrucure based positioning 2D error distribution (2-minute test).

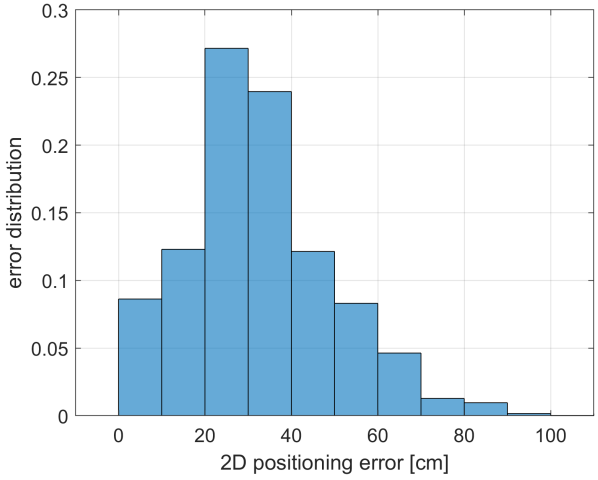

Figure 9. Positioning 2D error distribution of the UWB with sensors onboard of the smartphone (2-minute test).

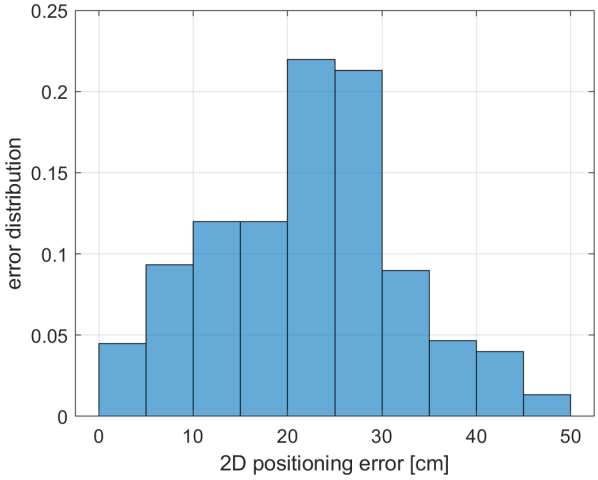

Figure 10. UAV imagery-based positioning 2D error distribution (2-minute test).

coverage is ensured during most of this test, continuity with a UAS-vision only approach cannot be ensured in a more general scenario, where a portion of the area of interest may be not vis-

ible (obstructed, or out of the field of view) from the UAV camera. Hence, integration with the other proposed strategies shall be useful in order to guarantee good operational conditions in a wider range of scenarios.

5. CONCLUSIONS

This work presented the results obtained in a test aiming at assessing the performance of pedestrian collaborative positioning, involving the use of several sensors, including UWB transceivers, cameras and IMU, and the use of additional information provided by a UAV.

The results obtained in the considered test show that UWB allow to obtain sub-metric error in good coverage conditions. The integration of UWB with IMU/smartphone camera allow to obtain a quite minor improvement in the performance obtained in the considered test. Nevertheless, such integration should improve the overall performance of the system in terms of continuity in different working conditions, in particular when the coverage provided by the static UWB architecture is less good.

UAS vision also proved to be very effective when the tracks of the agents are not obstructed and within the camera field of view. Nevertheless, since such working condition cannot always be ensured, the integration with other sensors shall make the overall solution more robust.

REFERENCES

Anderson, B. D., Moore, J. B., 2012. *Optimal filtering*. Courier Corporation.

Campos, C., Elvira, R., Rodríguez, J. J. G., M. Montiel, J. M., D. Tardós, J., 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, 37(6), 1874-1890.

Chiang, K.-W., Tsai, G.-J., Li, Y.-H., Li, Y., El-Sheimy, N., 2020. Navigation engine design for automated driving using INS/GNSS/3D LiDAR-SLAM and integrity assessment. *Remote Sensing*, 12(10), 1564.

Cieslewski, T., Choudhary, S., Scaramuzza, D., 2018. Data-efficient decentralized visual slam. *2018 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2466–2473.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric Computer Vision*. Springer.

Gurturk, M., Yusefi, A., Aslan, M. F., Soycan, M., Durdu, A., Masiero, A., 2021. The YTU dataset and recurrent neural network based visual-inertial odometry. *Measurement*, 184, 109878.

Iwaszczuk, D., Roth, S. et al., 2021. DeepLIO: deep LIDAR inertial sensor fusion for odometry estimation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 47–54.

Jazwinski, A. H., 2007. *Stochastic processes and filtering theory*. Courier Corporation.

Kailath, T., Sayed, A., Hassibi, B., 2000. *Linear Estimation*. Prentice-Hall.

Karrer, M., Schmuck, P., Chli, M., 2018. CVI-SLAM—collaborative visual-inertial SLAM. *IEEE Robotics and Automation Letters*, 3(4), 2762–2769.

Kukko, A., Kaijaluoto, R., Kaartinen, H., Lehtola, V. V., Jaakkola, A., Hyyppä, J., 2017. Graph SLAM correction for single scanner MLS forest data under boreal forest canopy. *ISPRS journal of photogrammetry and remote sensing*, 132, 199–209.

Leonard, J., Durrant-Whyte, H., 1991. Simultaneous map building and localization for an autonomous mobile robot. *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*, 1442–1447 vol.3.

Masiero, A., Dabove, P., Di Pietra, V., Piragnolo, M., Vettore, A., Cucchiaro, S., Guarnieri, A., Tarolli, P., Toth, C., Gikas, V., Perakis, H., Chiang, K.-W., Ruotsalainen, L., Goel, S., Gabela, J., 2021a. A case study of pedestrian positioning with UWB and UAV cameras. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1, 111–116.

Masiero, A., Toth, C., Gabela, J., Retscher, G., Kealy, A., Perakis, H., Gikas, V., Grejner-Brzezinska, D., 2021b. Experimental Assessment of UWB and Vision-Based Car Cooperative Positioning System. *Remote Sensing*, 13(23), 4858.

Masiero, A., Vettore, A., 2016. Improved Feature Matching for Mobile Devices with IMU. *Sensors*, 16(8), 1243.

Mur-Artal, R., Montiel, J. M. M., Tardos, J. D., 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, 31(5), 1147–1163.

Retscher, G., Kealy, A., Gabela, J., Li, Y., Goel, S., Toth, C. K., Masiero, A., Blaszczak-Bak, W., Gikas, V., Perakis, H., Koppanyi, Z., Grejner-Brzezinska, D. A., 2020. A benchmarking measurement campaign in GNSS-denied/challenged indoor/outdoor and transitional environments. *Journal of Applied Geodesy*, 14(2), 215–229.

Schmuck, P., Chli, M., 2017. Multi-uav collaborative monocular slam. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 3863–3870.

Schmuck, P., Ziegler, T., Karrer, M., Perraudin, J., Chli, M., 2021. Covins: Visual-inertial slam for centralized collaboration. *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, IEEE, 171–176.

Strasdat, H., Montiel, J. M., Davison, A. J., 2012. Visual SLAM: why filter? *Image and Vision Computing*, 30(2), 65–77.

Tong, P., Yang, X., Yang, Y., Liu, W., Wu, P., 2023. Multi-UAV Collaborative Absolute Vision Positioning and Navigation: A Survey and Discussion. *Drones*, 7(4), 261.

Whelan, T., Salas-Moreno, R. F., Glocker, B., Davison, A. J., Leutenegger, S., 2016. ElasticFusion: Real-time dense SLAM and light source estimation. *The International Journal of Robotics Research*, 35(14), 1697–1716.

Zhang, J., Singh, S., 2014. Loam: Lidar odometry and mapping in real-time. *Robotics: Science and Systems*, 2number 9.

Zhang, J., Singh, S., 2017. Low-drift and real-time lidar odometry and mapping. *Autonomous Robots*, 41(2), 401–416.

Zou, D., Tan, P., Yu, W., 2019. Collaborative visual SLAM for multiple agents: A brief survey. *Virtual Reality & Intelligent Hardware*, 1(5), 461–482.