UAV-Assisted Collaborative Positioning in GNSS-Denied Environments

Andrea Masiero[1],* Charles Toth[2], Luca Morelli[3], Pawel Trybala[3], Fabio Remondino[3]

[1] CIRGEO Interdepartmental Research Center, TESAF Department, University of Padova, Italy - andrea.masiero@unipd.it
[2] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, US - toth.2@osu.edu
[3] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy - (lmorelli, ptrybala, remondino)@fbk.eu

Keywords: Tracking, SLAM, deep learning, vision, photogrammetry.

Abstract

Accurate and reliable positioning and navigation are fundamental for the development of a wide number of applications. Despite in most of the regular working conditions the use of a GNSS (Global Navigation Satellite System) receiver is sufficient for properly determining the absolute coordinates of such a device, determining a reliable solution of the positioning problem in challenging conditions, e.g. in environments where GNSS is not available or not reliable, can be difficult. In such conditions, exploiting the information shared by different sensors and platforms can be useful for reliably determining the platforms' positions. In this work, both ground and aerial platforms are considered: each platform is assumed to be provided with communication capabilities, which can be exploited to share its knowledge. In particular, since GNSS positioning is usually less effective at ground level than on a flying platform, here the aerial platforms are assumed to be provided with good GNSS-based positioning information. Instead, GNSS is assumed to be unavailable to the ground vehicles, which, instead, can use LiDAR/visual odometry for determining dead-reckoning solutions, Ultra Wide-Band (UWB) inter-platform ranging for relative positioning, and camera-based positions, provided by aerial platforms, for assessing their georeferenced positions. In particular, this work focuses on assessing the positioning performance when exploiting vision-based information about the georeferenced ground vehicle positions from a camera mounted on a Unmanned Aerial Vehicle (UAV). The camera acquired oblique views of the scene, while moving over the case study area during the test. YOLO network was used to detect cars from the image frames, whereas different techniques were used for obtaining the cars' georeferenced coordinates, including the extraction of vehicle coordinates from 3D reconstructions obtained from the MoGe-2 network. Average errors at meter level on the determined georeferenced coordinates, even on vehicles not visible in the image acquired by the UAV, were obtained when combining UWB vehicle-to-vehicle ranges with MoGe-2 reconstructions.

1. Introduction

Positioning and navigation systems on most of the devices on the market typically rely on GNSS (Global Navigation Satellite System): despite satellite positioning ensures good performance on standard outdoor scenarios, its accuracy and reliability are quite limited in non-open sky environments, and in several critical operating conditions, such as in urban canyons, tunnels, indoor spaces, underground areas, densely forested regions, or when dealing with jamming and spoofing issues (Zidan et al., 2020). Applications like self-driving cars, indoor navigation, autonomous search-and-rescue missions, and, more in general, tasks involving the use of autonomous agents, require precise positioning even in challenging environments. To address these scenarios, multi-sensor systems are often employed, integrating data from various sources to provide more robust positioning when GNSS is unreliable or unavailable.

Traditionally, Inertial Measurement Units (IMUs) played a crucial role for compensating GNSS temporary unavailability, e.g. bridging gaps in the GNSS signal (Grejner-Brzezinska et al., 2002, Li et al., 2017). IMU provides continuous, high-frequency data in any environmental condition, and, differently from GNSS, it is not susceptible to radio frequency interference or jamming, making IMU a very convenient solution to support the development of positioning methods alternative to GNSS. Unfortunately, when relying only on IMU data, the accumulation of IMU errors lead to unacceptable drifts in the positioning solution: this is usually very quickly visible when using low cost IMUs, while the use of high-end sensors can significantly increase the length of the time interval of stand-alone usability of IMU data.

Dedicated radio signals, and signals of opportunity (SoP), have also been deployed in many alternative positioning systems. The effective use of methods based on SoP typically requires the availability of prior information, usually in terms of long training datasets, collected ad hoc on the site of interest (Kuang et al., 2018, Davidson and Piché, 2016, Ban et al., 2015, Sakr and El-Sheimy, 2017). Often, more accurate results can be reached by the deployment of dedicated radio devices, such as Ultra Wide-Band (UWB) transceivers, which ensure both communication and Time-of-Flight-based ranging capabilities between any two of them (Retscher et al., 2019).

The availability of platforms provided with communication capabilities also enables the development of cooperative positioning approaches (Kealy et al., 2014): in this case, platforms, often referred to as agents, through communications can exchange positioning and navigation information, and any other kind of additional data that can be useful for either the localization problem or the successful completion of their tasks. In these kind of approaches, the system is typically seen as a network of platforms, either dynamic or static, where static platforms usually forms a static infrastructure, dedicated to supporting the positioning problem. A moving platform is usually capable to communicate with any other enabled device (V2X - vehicle-to-everything communication), which, in practice, usually leads to either communicating with static infrastructure nodes (vehicle-to-infrastructure (V2I)) or with other

* Corresponding author

moving platforms (vehicle-to-vehicle - (V2V)). In the context of autonomous vehicles, advancements in communication technologies like 5G are expected to enable wide-band real-time information sharing among interconnected platforms (Garcia et al., 2021). Cooperative positioning usually ensures the computation of more robust solutions, hence potentially reducing the effect of certain issues such as spoofing and jamming, or ensuring more effective positioning to agents not enabled to standalone robust localization.

Even though several sensors can be used to support the compensation of GNSS unreliability/unavailability, many recent systems are based on the use of SLAM (Simultaneous Localization and Mapping, (Leonard and Durrant-Whyte, 1991)) approaches. SLAM-based methods can be visual (Strasdat et al., 2012, Macario Barros et al., 2022), using cameras (e.g., ORB-SLAM3 (Mur-Artal et al., 2015, Campos et al., 2021) and COLMAP-SLAM (Morelli et al., 2023)), or LiDAR-based. In the latter case, it is often called LiDAR odometry (e.g., LOAM (Zhang and Singh, 2014) and LeGO-LOAM (Shan and Englot, 2018)), and, when inertial sensor data are also deployed, it is referred to as LiDAR Inertial Odometry (LIO), e.g. LIO-SAM (Shan et al., 2020), FAST-LIO2 (Xu et al., 2022), DeepLIO (Iwaszczuk et al., 2021) and Kinematic-ICP (Guadagnino et al., 2025).

Several works have already studied strategies to extend the use of vision, and in particular SLAM, to collaborative scenarios (Karrer et al., 2018, Tong et al., 2023, Zou et al., 2019), leading to the development of either centralized or distributed computation strategies (Schmuck et al., 2021, Cieslewski et al., 2018): centralized computations, on a master station, usually allow to obtain better performance, but they typically imply sharing high amount of data, which can be critical for the communication channel with the master station when the number of agents is large. Instead, decentralized strategies, if properly implemented, can ensure the scalability of the method even to large agent networks.

On the other hand, the recent availability of quite reliable networks for the generation of metric scale 3D point maps from single images (Yang et al., 2024b, Yang et al., 2024a, Wang et al., 2025b, Wang et al., 2025a) increases also the camera-based object tracking possibilities.

Even if the vision-based collaborative positioning problem has already been studied by several researchers, it is usually either limited to ground or aerial platforms, whereas the simultaneous use of both of them is quite uncommon. Motivated by this observation, this study focuses on the integration of multi-sensor data provided by aerial and ground platforms in a collaborative approach. Tests of the proposed approach are conducted on the CONTEST dataset[1] (Masiero et al., 2023a, Masiero et al., 2023b). This dataset includes reference trajectories and multisensor data from pedestrians and aerial and ground vehicles, collected in a parking lot in the campus of the Ohio State University (Figure 1).

2. Method

This paper is part of a research work that focuses on developing a collaborative strategy between ground and aerial vehicles, exploiting their peculiarities. Ground vehicles are usually more affected than aerial platforms by certain GNSS issues such as multi-path and signal loss. However, they are usually provided with a wider range of sensors, while the payload on UAVs (Unmanned Aerial Vehicles) is typically quite limited. Ground vehicles, in particular in applications involving autonomous platforms, are usually provided with several sensors such as LiDARs, cameras, and RADARs. In this case, each ground vehicle was provided with a LiDAR, a camera and an UWB transceiver. Instead, UAVs were provided with the in-built camera and an UWB transceiver. All the platforms were also provided with GNSS receivers: accurate solutions were computed in post processing kinematic (PPK) mode. In order to simulate a challenging working condition for the positioning system, in this work the GNSS solutions are assumed to be unavailable to certain of the involved platforms, in particular the ground vehicles. Instead, they do are used as reference data, for validating the obtained results. More details on platforms and sensors can be found in (Masiero et al., 2023a).

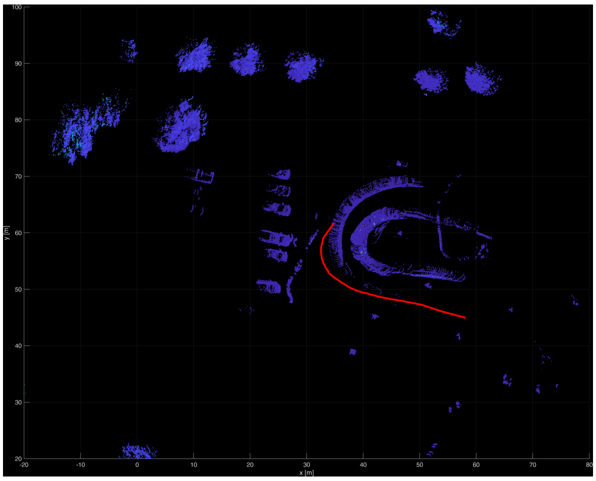
Figure 1. Aerial view of the test area.

In accordance with the working conditions mentioned above, GNSS measurements were assumed to be not available to the four ground platforms, which used their other sensors for positioning purposes. In particular, a LiDAR odometry approach has been implemented to determine their movements over time. Figure 2 shows the above-ground objects mapped, with 20 s of collected data, by one of the vehicles, the GPSVan of the Ohio State University, through a LiDAR odometry approach. The trajectories of certain other vehicles and pedestrians are quite easily visible in the mapped point cloud: segmenting, classifying and properly associating over time the clouds of the dynamic objects allow to assess their positions, in the navigation reference frame used to compute the LiDAR odometry solution.

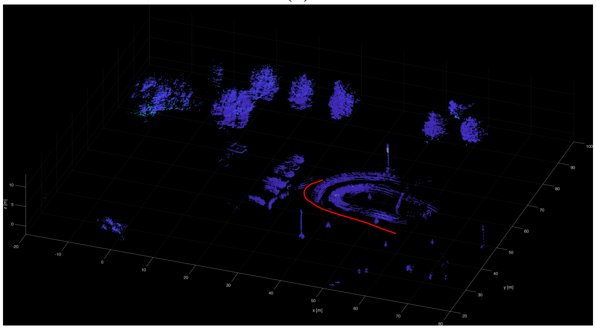
Even if LiDAR odometry can be used to determine the changes in the vehicle positions, the unavailability of GNSS measurements and of any other prior information imply that the ground vehicles cannot determine their absolute (georeferenced) positions with the information collected by their own sensors, which can be realistic for instance indoors, i.e. indoor parking lots, or in urban canyons. Instead, since UAVs usually fly at tens of meters above the ground, they are less affected by GNSS issues. This work considers the interaction between all the four ground vehicles of the CONTEST dataset with just one UAV, a DJI Phantom 4 Pro, which, in agreement with what mentioned above, is assumed to be provided with good positioning information, derived from PPK GNSS measurements and by processing its acquired images. Also, the camera of the UAV, is assumed to be geometrically calibrated.

Ground vehicles and pedestrians can be detected on the UAV image frames, thanks to a YOLO (You Only Look Once) network (Redmon et al., 2016, Terven et al., 2023). Figure 3 shows

[1] https://github.com/3DOM-FBK/Collaborative_Navigation



(a)



(b)

Figure 2. Top (a) and oblique view (b) of the above ground objects mapped by a ground vehicle, GPSVan, in approximately 20 s. The vehicle estimated trajectory is shown as a red line.

Figure 3. Sequence of UAV camera views: detected ground vehicles are shown with yellow boxes, whereas light blue boxes indicate the detected pedestrians.

a sequence of UAV camera views. The detected ground vehicles and pedestrians are shown with the corresponding bound boxes, yellow and light blue, respectively.

Three methods, named A, B and C in the following, were considered for determining camera-based georeferenced positions of the ground vehicles. The first basically rely on the knowledge of a terrain surface model, e.g. Digital Terrain Model (DTM). Since such a terrain model may not be available, two alternative methods for extracting the positions of the ground vehicles were also considered, where the last does not rely on a previously terrain model. Both such methods are based on the use of the MoGe-2 network (Wang et al., 2025b) and of UWB ranging, as explained below.

2.1 Method A

First, assuming the height of the ground vehicles to be approximately known, an approximation of the georeferenced 3D position of the centroid of the vehicle is obtained:

- determining the optical ray of the camera of the UAV passing through the 2D centroid of the region associated to the UAV in the considered image frame
- intersecting the mentioned optical ray with the ground surface, determined thanks to the points triangulated on different camera views, and moving the estimated position in the direction towards the UAV until reaching half of the vehicle height. In alternative, the intersection can be performed with any other available high resolution model of

the terrain. Half of the vehicle height can be added to the obtained position to determine the corresponding location on the top of the vehicle, which can be useful for comparing with the GNSS receiver locations, which were installed on the top of the vehicles.

Figure 9 shows an example of the locations of four ground vehicles estimated from the image frames shown in Figure 3.

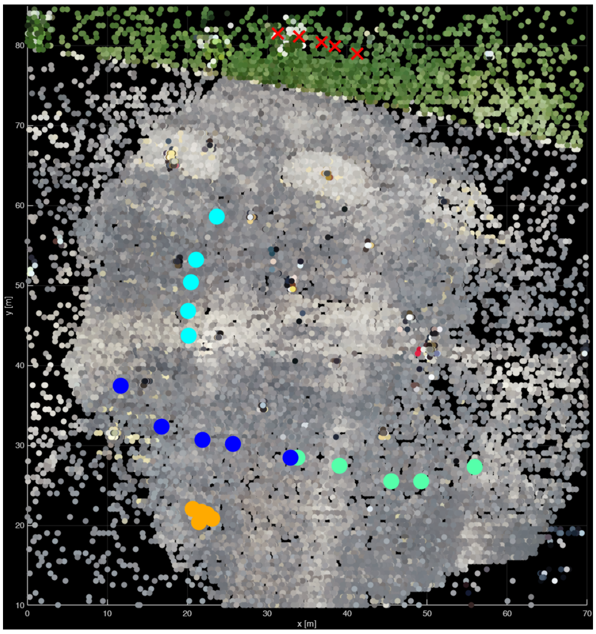
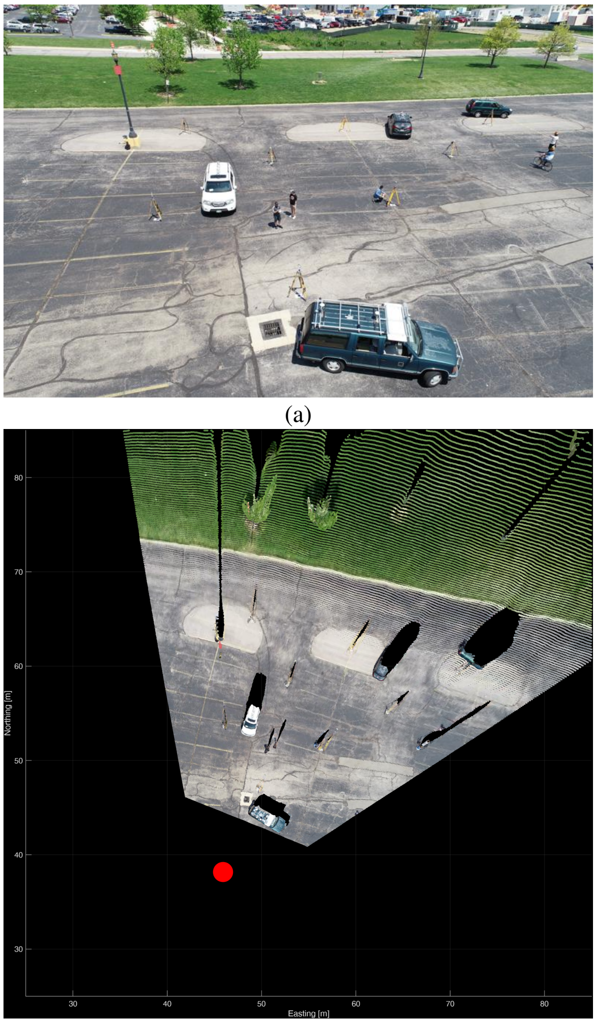


Figure 4. Example of ground vehicles tracked from a UAV: the UAV locations for the five frames of Figure 3 are shown with red x-marks, whereas the assessed positions for the four tracked ground vehicles are shown as solid circles (light blue, blue, geen, ocher).

2.2 Method B

MoGe-2 can be used to quickly obtain a low resolution 3D point cloud from a single image of the UAV. Figure 5(a) shows an example of image acquired by the UAV, whereas Figure 5(b) shows the corresponding georeferenced point cloud, obtained with the following steps:

- Running MoGe-2 on a low resolution undistorted version of the original image. For ensuring quite fast processing results, a 800×533 pixel image in inputted in MoGe-2.
- Since the camera model cannot be inputted in MoGe-2, the $x - y$ coordinates of the obtained output are re-scaled in order to match the field-of-view of the calibrated camera model.
- Interior and exterior camera parameters are used to obtain a georeferenced version of the MoGe-2 output.
- Ground points are removed by using a simple morphological filter (SMRF) (Pingel et al., 2013) and point sets of the vehicles can be segmented from the remaining point cloud by determining the areas in the MoGe-2 output corresponding to those detected by the YOLO network. To such aim, it is worth to notice that the MoGe-2 network also automatically determines the image distortion, without providing any output related to the distortion, which, hence, should also be assessed by comparison with the original image (in our work only one parameter of radial distortion has been considered).



(a)

(b)

Figure 5. Example of image acquired by the UAV (a) and the corresponding georeferenced point cloud obtained from MoGe-2 output (b). The position of the UAV during the acquisition of the image in (a) is shown as a red solid circle in (b).

Despite the point clouds outputted by MoGe-2 are approximately metric, the scaling factors computed by the network are only approximations of the correct ones. This can be experimentally proved by checking Figure 6, where the off-ground objects extracted from five consecutive frames, with the procedure mentioned above, are overlapped. Even if the tracks of different vehicles can be quite easily distinguished, a careful analysis of the image clearly proves the need for improving the scaling factors of the point clouds outputted by MoGe-2.

However, UWB ranging between the UAV and the ground vehicle (i.e. UAV2V ranges, hereafter) can be conveniently used to such aim: both aerial and ground platforms in the CONTEST dataset were provided with UWB ranging transceivers, hence, UAV-vehicle ranges can be used for properly scaling the corresponding distances in the MoGe-2 point cloud (a least squares solution, with outlier rejection, can be implemented if several range measurements are available).

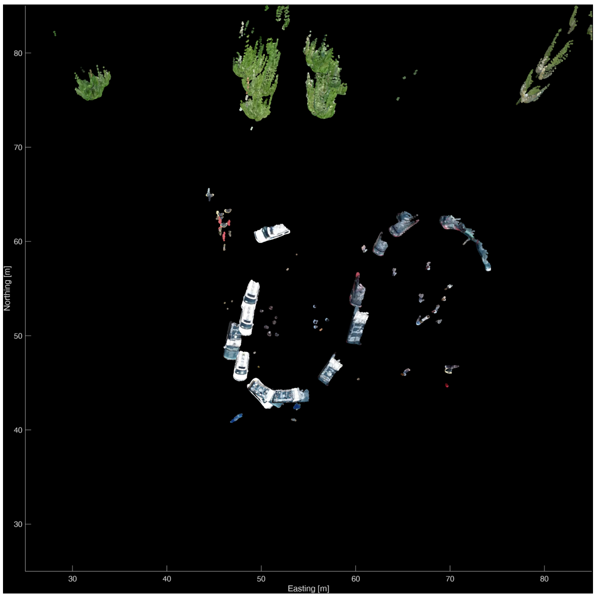


Figure 6. Overlapping of off-ground objects extracted from the georeferenced point clouds obtained from MoGe-2 applied to five consecutive images acquired by the UAV.

Once that the cloud corresponding to a vehicle has been extracted, two options have been considered to compute the corresponding coordinates. First, the silhouette corresponding to the bottom of the vehicle has been extracted. The longest line has been determined from the silhouette with a RANSAC (random sample consensus (Fischler and Bolles, 1981)) based approach. If the length line is longer than 2 m, than it is assumed to correspond to the side of the vehicle, otherwise to either the front or the rear. Then, the mean point along the detected line is considered, and the vehicle position in the cloud is determined from such a point by moving along the inner direction (i.e. inside the vehicle) conventionally by 0.8 m if the detected line corresponded to a vehicle side, by 2 m otherwise, and by $\bar{h}$, the vehicle height, in the up direction. $\bar{h}$ was determined from the maximum variation along the up direction of the points in the vehicle cloud. Then, the computed position is remapped on the image space, and the vehicle position is determined by intersecting the optical ray from the camera to the remapped point position with the terrain surface, similarly to the first considered case, in Method A.

2.3 Method C

Similarly to Method B, MoGe-2, SMRF and YOLO are used to obtain segmented point clouds associated to each vehicle. Then, the centroid of a vehicle point cloud, expressed in georeferenced coordinates, is simply considered instead, properly translated in the up direction by half of vehicle height, and taken as the estimated vehicle position.

2.4 Collaborative positioning based on V2V UWB ranging

Finally, thanks to the UWB V2V ranging between the ground vehicles, collaborative positioning is exploited to both improve the relative positions between the ground vehicles (Masiero et al., 2021), and to determine the positions of those vehicles when they are not visible in an UAV image.

3. Results

This section shows the results obtained by analyzing approximately 200 images acquired by the considered UAV. The corresponding UWB range measurements have been simulated, by considering the ground truth vehicle distances and a Gaussian white noise, with $\sigma = 0.25$ cm, as quite realistic in real applications (Masiero et al., 2021).

Figure 7 shows the distribution of the estimated scaling factor values for re-scaling the outputs of the MoGe-2 network, estimated for all the considered images.

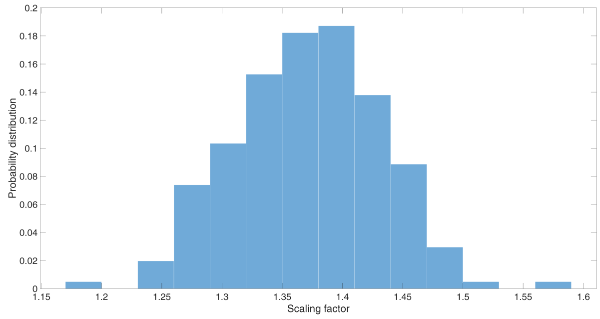


Figure 7. Distribution of the estimated scaling factor values for the MoGe-2 point clouds.

Then, Figure 8 compares the probability distributions of the relative distance between vehicles when not exploiting V2V UWB ranging (a) and when exploiting it (b).

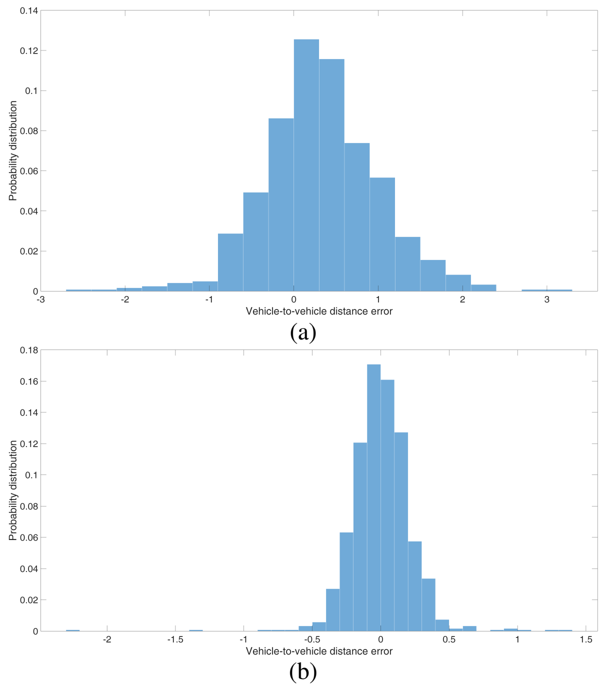


Figure 8. Probability distribution of the relative distance between vehicles when not exploiting V2V UWB ranging (a) and when exploiting it (b).

Table 1 compares the performance obtained in the following cases:

- UAV+YOLO+UAV2V: UWB UAV2V ranges have been used to determine the distance of the vehicles detected by the YOLO network, whose position is then determined by intersecting the corresponding optical from the UAV camera with the terrain DTM.

- UAV+YOLO+Moge2+silh: point clouds outputted by MoGe-2 are georeferenced thanks to the interior and exterior UAV camera parameters. Then, the silhouette of each vehicle is used to determine the corresponding center, which is remapped on the image frame and its corresponding optical ray intersected with the DTM.
- UAV+YOLO+Moge2+silh+UAV2V: similar as the previous one, but UAV2V ranging is used to re-scale the MoGe-2 output.
- UAV+YOLO+Moge2+silh+UAV2V+V2V: similar as the previous one, but V2V UWB ranging is used for cooperative positioning.
- UAV+YOLO+Moge2+centroid: point clouds outputted by MoGe-2 are georeferenced thanks to the interior and exterior UAV camera parameters. Then, the centroid the cloud of each vehicle is used to determine the corresponding vehicle position.
- UAV+YOLO+Moge2+centroid+UAV2V: similar as the previous one, but UAV2V ranging is used to re-scale the MoGe-2 output.
- UAV+YOLO+Moge2+centroid+UAV2V+V2V: similar as the previous one, but V2V UWB ranging is used for cooperative positioning.

The results shown in Table 1 corresponds to the mean m and standard deviations σ of the errors along the x and y directions (Easting and Northing), the 2D Root Mean Square (RMS) errors and the availability of the solution, computed as the number of the predicted vehicle positions divided by the total number of vehicle positions in the tests (the latter is 4, vehicles, multiplied by the number of considered images).

In order to make a fair comparison, in Table 1 all the methods mentioned above are compared on the prediction of the positions of the same instances of vehicles, even if some of them can make some additional vehicle position predictions, in particular thanks to the usage of V2V UWB collaborative positioning for determining positions even of vehicle not visible in the images. Differently, Table 2 provides the positioning results for all the possible predictions.

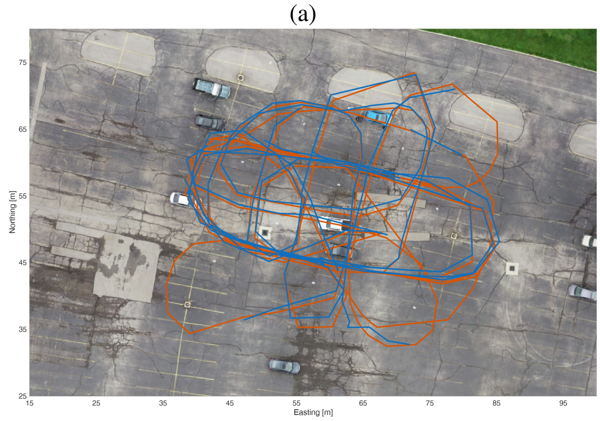
Figure 9 shows the ground truth (red solid lines) and the estimated (blue solid lines) positions for two of the ground vehicles (position estimates in the UAV+YOLO+Moge2+centroid+UAV2V+V2V case). The corresponding positioning errors, as functions of the UAV image index, are shown in Figure 10.

4. Conclusions

This paper compared different options to obtain georeferenced positions of ground vehicles when combining vision information provided by a UAV, assumed to be provided also with good GNSS positioning, with UWB ranging between ground-to-ground and aerial-to-ground vehicles.

The obtained results showed that the results obtained by exploiting MoGe-2 point clouds can be better than those obtained

(a)



(b)

Figure 9. Example of tracks for two ground vehicles: ground truth (red) and estimated (blue).

by obtaining the vehicle position by simply intersection of the corresponding optical ray with the terrain surface. This is probably due to the missing information of the exact vehicle height, and the approximation of the vehicle position with its centroid on the image, which is not necessarily close to its real center, nor to the GNSS receiver location, used as reference vehicle position.

The relative distance error between vehicles shows a clear advantage in the introduction of V2V UWB measurements: the reduction of the average error and of its dispersion is quite clear when comparing Figure 8(a) and (b). Also, they increased the estimated position availability from 77.0% to 86.2% in the MoGe-2 based cases.

On the other hand, the positive effect of using UAV-to-vehicle UWB ranging is also quite apparent in reducing the errors related to the scaling of MoGe-2 point clouds, as clear when comparing, in Table 1, the results of UAV+YOLO+MoGe2+silh with those of UAV+YOLO+MoGe2+silh+UAV2V, and those of UAV+YOLO+MoGe2+centroid with UAV+YOLO+MoGe2+centroid+UAV2V.

The use of the centroid of the vehicle point cloud extracted from the MoGe-2 output appears as the simplest method reaching the best, or close to the best, obtained results. This may not be the same when more information on the vehicles is available, but in this case study no prior information on the specific geometric shape of the vehicles has been introduced (just a generic approximate overall size, used for all the vehicles).

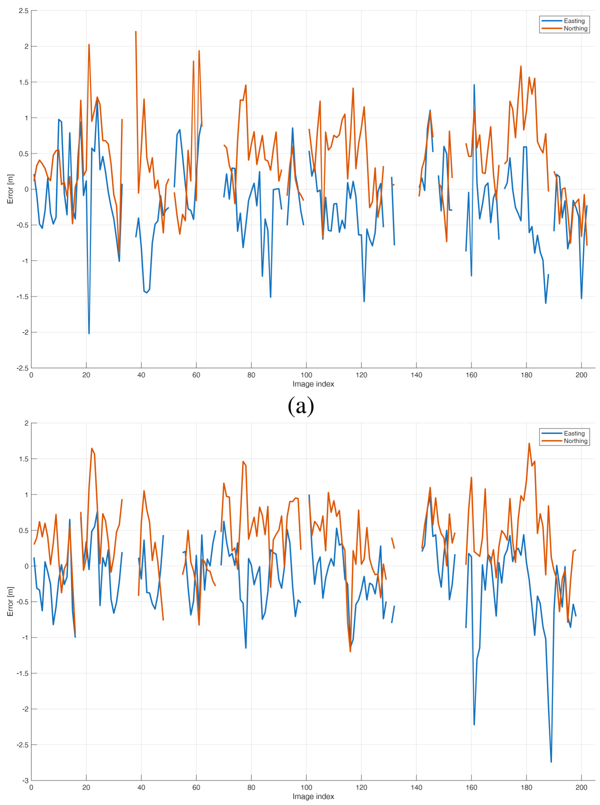
Given the results shown in Figure 8(b) and Table 1,

Approach	m_x [m]	m_y [m]	σ_x [m]	σ_y [m]	2D RMS [m]	availability %
UAV+YOLO+UAV2V	-0.19	0.27	0.91	0.78	1.25	80.2
UAV+YOLO+MoGe2+silh	-1.52	1.76	5.34	6.50	8.72	77.0
UAV+YOLO+MoGe2+silh+UAV2V	-0.12	0.37	0.79	0.75	1.16	77.0
UAV+YOLO+MoGe2+silh+UAV2V+V2V	-0.12	0.37	0.65	0.61	0.97	86.2
UAV+YOLO+MoGe2+centroid	-1.67	1.87	5.77	6.95	9.37	77.0
UAV+YOLO+MoGe2+centroid+UAV2V	-0.17	0.39	0.62	0.63	0.98	77.0
UAV+YOLO+MoGe2+centroid+UAV2V+V2V	-0.17	0.39	0.52	0.52	0.85	86.2

Table 1. Characteristics of the positioning error, computed on the same samples for all the approaches.

Approach	m_x [m]	m_y [m]	σ_x [m]	σ_y [m]	2D RMS [m]	availability %
UAV+YOLO+UAV2V	-0.17	0.32	0.93	0.83	1.29	80.2
UAV+YOLO+MoGe2+silh+UAV2V+V2V	-0.11	0.36	0.66	0.62	0.98	86.2
UAV+YOLO+MoGe2+centroid+UAV2V+V2V	-0.16	0.37	0.64	0.65	0.99	86.2

Table 2. Characteristics of the positioning error, computed on all the samples considerable for each approach.



(a)

(b)

Figure 10. Positioning error for the two vehicles in figure Figure 9(a) and (b).

it is clear that the absolute positioning error, in particular in the best case, 2D RMS=0.85 m for UAV+YOLO+MoGe2+centroid+UAV2V+V2V, is mostly caused by the georeferencing error. Actually, it is also worth to notice that the proposed method aims at estimating the (planimetric) position of the center of the vehicle, whereas the reference position is provided by GNSS receivers, which usually are not located in such center, hence a component of the error, typically of some decimeters, is due to such difference between the estimated position and the one used as reference. Such a difference will be taken into consideration in our future works.

Finally, our future work will also consider a more detailed examination when more geometric information about vehicles is provided, and to the use of real UWB measurements instead of simulated ones.

References

Ban, R., Kaji, K., Hiroi, K., Kawaguchi, N., 2015. Indoor positioning method integrating pedestrian dead reckoning with magnetic field and wifi fingerprints. *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, IEEE, 167–172.

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.

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

Davidson, P., Piché, R., 2016. A survey of selected indoor positioning methods for smartphones. *IEEE Communications Surveys & Tutorials*, 19(2), 1347–1370.

Fischler, M., Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Garcia, M. H. C., Molina-Galan, A., Boban, M., Gozalvez, J., Coll-Perales, B., Şahin, T., Kousaridas, A., 2021. A tutorial on 5G NR V2X communications. *IEEE Communications Surveys & Tutorials*, 23(3), 1972–2026.

Grejner-Brzezinska, D. A., Toth, C. K., Yi, Y., 2002. Bridging gps gaps in urban canyons: Can zupt really help? *Proceedings of the 58th Annual Meeting of The Institute of Navigation and CIGTF 21st Guidance Test Symposium (2002)*, 231–240.

Guadagnino, T., Mersch, B., Vizzo, I., Gupta, S., Malladi, M. V., Lobefaro, L., Doisy, G., Stachniss, C., 2025. Kinematic-icp: Enhancing lidar odometry with kinematic constraints for wheeled mobile robots moving on planar surfaces. *2025 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 11090–11096.

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.

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

Kealy, A., Alam, N., Efatmaneshnik, M., Toth, C., Dempster, A., Brzezinska, D., 2014. Collaborative positioning in gps-challenged environments. *Earth on the Edge: Science for a Sustainable Planet: Proceedings of the IAG General Assembly, Melbourne, Australia, June 28-July 2, 2011*, Springer, 493–500.

Kuang, J., Niu, X., Zhang, P., Chen, X., 2018. Indoor positioning based on pedestrian dead reckoning and magnetic field matching for smartphones. *Sensors*, 18(12), 4142.

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.

Li, Y., Zhuang, Y., Zhang, P., Lan, H., Niu, X., El-Sheimy, N., 2017. An improved inertial/wifi/magnetic fusion structure for indoor navigation. *Information Fusion*, 34, 101–119.

Macario Barros, A., Michel, M., Moline, Y., Corre, G., Carrel, F., 2022. A comprehensive survey of visual slam algorithms. *Robotics*, 11(1), 24.

Masiero, A., Morelli, L., Toth, C., Remondino, F., 2023a. Benchmarking Collaborative Positioning and Navigation Between Ground and Uas Platforms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 1127–1133.

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

Masiero, A., Toth, C., Remondino, F., 2023b. Vision and UWB-based collaborative positioning between ground and UAS platforms. *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, IEEE, 748–754.

Morelli, L., Ioli, F., Beber, R., Menna, F., Remondino, F., Vitti, A., 2023. COLMAP-SLAM: A framework for visual odometry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 317–324.

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.

Pingel, T. J., Clarke, K. C., McBride, W. A., 2013. An improved simple morphological filter for the terrain classification of airborne LIDAR data. *ISPRS journal of photogrammetry and remote sensing*, 77, 21–30.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Retscher, G., Gikas, V., Hofer, H., Perakis, H., Kealy, A., 2019. Range validation of UWB and Wi-Fi for integrated indoor positioning. *Applied Geomatics*, 11(2), 187–195.

Sakr, M., El-Sheimy, N., 2017. Efficient wi-fi signal strength maps using sparse gaussian process models. *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 1–8.

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.

Shan, T., Englot, B., 2018. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 4758–4765.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D., 2020. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 5135–5142.

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

Terven, J., Córdova-Esparza, D.-M., Romero-González, J.-A., 2023. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine learning and knowledge extraction*, 5(4), 1680–1716.

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.

Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., Novotny, D., 2025a. Vggt: Visual geometry grounded transformer. *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5294–5306.

Wang, R., Xu, S., Dong, Y., Deng, Y., Xiang, J., Lv, Z., Sun, G., Tong, X., Yang, J., 2025b. Moge-2: Accurate monocular geometry with metric scale and sharp details. *arXiv preprint arXiv:2507.02546*.

Xu, W., Cai, Y., He, D., Lin, J., Zhang, F., 2022. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics*, 38(4), 2053–2073.

Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H., 2024a. Depth anything: Unleashing the power of large-scale unlabeled data. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10371–10381.

Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., 2024b. Depth anything v2. *Advances in Neural Information Processing Systems*, 37, 21875–21911.

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

Zidan, J., Adegoke, E. I., Kampert, E., Birrell, S. A., Ford, C. R., Higgins, M. D., 2020. GNSS vulnerabilities and existing solutions: A review of the literature. *IEEE Access*, 9, 153960–153976.

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

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-41-2026 |