ON THE ESTIMATION OF VEHICLE TRAJECTORIES WITH A MINI-UAS

F. Mugnai[1], A. Masiero[1,*], B. Ciuffo[2]

[1] Dept. of Civil and Environmental Engineering, University of Florence,
via di Santa Marta 3, Florence 50139, Italy - (francesco.mugnai, andrea.masiero)@unifi.it
[2] Directorate for Energy, Transport and Climate Change, European Commission, Joint Research Center
via E. Fermi 2749, Ispra 21027, Italy - biagio.ciuffo@ec.europa.eu

KEY WORDS: Car tracking, Traffic monitoring, Mini-UAV, Photogrammetry, Vision, Structure from Motion

ABSTRACT:

The development of effective solutions for automated vehicles ensuring high levels of safety in any scenario have to be tested in many real world conditions. Large datasets describing real user behaviors in many conditions should be collected to this aim. To this purpose, a quite attractive option is that of using Unmanned Aerial System (UAS) imagery, which can be effectively used to extract real world driver trajectories, and to extract information of interest concerning both their interactions and the context. This work aims at providing an assessment of the accuracy of the geometric information that can be extracted on users' trajectories and to present a strategy for the estimation of the 3D shape of the involved vehicles. The obtained results show a quite remarkable performance in terms of ability of determining a proper description of the characteristics of the vehicle trajectories, i.e. speed and track.

1. INTRODUCTION

The realization of safe and effective solutions for automated vehicles requires the availability of large real-scenario datasets in order to properly check and test the implemented methods. To such aim, the use of Unmanned Aerial System (UAS) imagery appears as an attractive option to collect quite long sequences of video data, which can be used to extract real world driver trajectories, to model their behavior and their interactions (Puri et al., 2007, Khan et al., 2017, Kanistras et al., 2013, Coifman et al., 2006). Automated vehicles shall be appropriately integrated in real scenarios, hence the availability of the above mentioned datasets, along with additional information on the context (e.g. road types, maps, traffic signs), is of remarkable importance to check via simulation their behavior in a number of real-like conditions.

Despite some datasets of UAS imagery are already available for the purpose of properly describing real car user behaviors (Krajewski et al., 2018, Bock et al., n.d., Krajewski et al., 2020), the goal of our work is:

- providing an assessment of the reliability of the geometric information computed by the UAS imagery;
- providing, in addition to the user trajectories, high resolution geometric information of the test area (e.g. HD maps (Chiang et al., 2022)) and of the involved vehicles;
- presenting an approach for computing 3D information from the UAS imagery, in particular for what concerns the moving vehicles.

In addition to the previous considerations, a real-time analysis of the UAS imagery could also be used to support positioning and navigation in challenging environments (e.g. urban canyons, GNSS-denied areas, (Masiero et al., 2021a)), if the obtained solutions are immediately made available to the vehicles, which could be possible for future interconnected smart vehicles.

This paper focuses on the first and last item of the previous list, whereas the rest will be considered in our future investigations.

Similarly to (Masiero et al., 2022), this work aims at exploiting vision tools in order to properly track moving objects/persons on the ground. In particular, the rationale of the method used here for determining an initial assessment of the driver trajectories is similar to that in (Kurz et al., 2022), i.e. via forward ray intersection. Nevertheless, this work extends the approach in (Kurz et al., 2022) by also considering a partial assessment of the vehicle shape.

Section 2 introduces the case study considered in this paper. Section 3 provides a description of the proposed method: first, summarizing the implemented approach to determine the vehicle trajectories, and then providing a more detailed description on a procedure to assess also the 3D shapes of the moving vehicles from the imagery acquired by the (moving) UAS. Section 4 shows the obtained results. Finally, some discussion and conclusions are drawn in Section 5 and 6.

2. CASE STUDY

This paper presents the results obtained in a portion of the data acquired within a collaboration between the Joint Research Center, the University of Florence and the IAG Working Group 4.1.4 (Computer Vision in Navigation), during the SARA project test campaign.

A DJI Mini 2 UAS, shown in Fig. 1, was used to acquire the imagery to be used to track car positions in a urban environment.

Geodetic GNSS receivers were fixed on the top of the vehicles, to collect the reference vehicle trajectories. Furthermore, in order to make a proper validation of the vision-based tracking

* Corresponding author

method, the GNSS receiver position has been apparently pointed out with a black x-mark on the car ceilings (Fig. 2): this ensured the possibility of comparing the reference solutions with the estimated trajectories on the same points on the cars.

Figure 1. DJI Mini 2 drone.

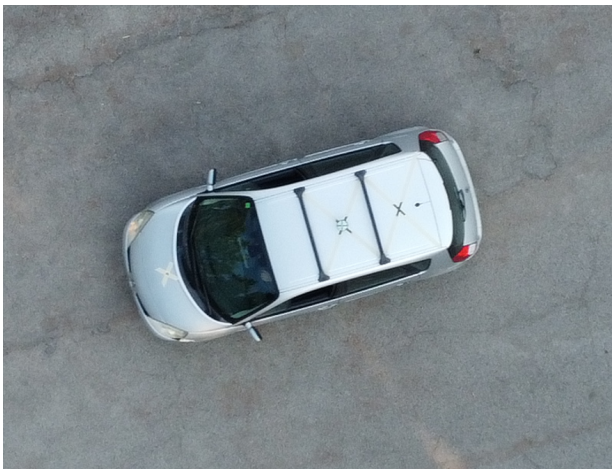


Figure 2. Car top view: a geodetic GNSS receiver was fixed on the top of the vehicle, and its position apparently pointed out with a black x-mark.

3. METHOD

Since mini-UASs are usually provided with navigation systems able to determine the vehicle geographic position with errors at meter level, first, a vision-based UAS positioning procedure has been implemented.

A set of targets, surveyed with geodetic GNSS receivers, have been distributed over the area of interest. Alternatively, any other point of known position, easy to be identified in the UAS images, can be used as landmark.

The UAS camera is assumed to be pre-calibrated (Remondino and Fraser, 2006, Luhmann et al., 2015, Fraser, 2018, Habib and Morgan, 2003, Balletti et al., 2014), hence tracking on the UAS images the above mentioned target positions (for instance with the Kanade-Lucas-Tomasi feature tracker (Lucas et al., 1981, Shi and Tomasi, 1994)) allow to determine frame by frame the camera exterior orientation by spatial resection (Förstner and Wrobel, 2016, Kraus, 2007, Mikhail et al., 2001).

Furthermore, so long as the UAS moves over all the area of interest while acquiring images, it is well known that in the conditions mentioned above it is possible to obtain, via trinagulation (Mikhail et al., 2001, Hartley and Sturm, 1997), a georeferenced 3D reconstruction of the static parts of the scene visible in the UAS imagery (commercial software can be used to this aim, e.g. Agisoft Metashape, Pix4D).

Then, as long as a vehicle point is identified on the UAS imagery, for instance by determining the vehicle region by means of background subtraction or by using a deep-learning based method (Masiero et al., 2021b), its 3D position can be estimated via forward ray intersection (assuming the vehicle height approximately known, and that a digital terrain model of the area is available, e.g. obtained via photogrammetric reconstruction using the UAS imagery).

Since different parts of a vehicle have different heights, the above mentioned method could be used to determine an initial approximation, used also in the following to assess the vehicle motion, whereas the approach reported below is used in order to determine the 3D positions of the tracked points of the vehicle.

First, it is assumed that the vehicle movements can be model as a rigid body motion. Let M_t be the position of a point on the vehicle at time t, and $R_{t,c}$ and $t_{t,c}$ the rotation matrix and the translation vector representing the vehicle orientation and position (of its centroid, for instance) at time t. Then, M_t can be expressed as a function of its original position M_0 at time 0 and of $\{R_{t,c},\ t_{t,c}\}$ as follows:

$$M_t = R_{t,c} M_0 + t_{t,c} \tag{1}$$

where estimates of $R_{t,c}$ and $t_{t,c}$ are assumed to be available, as previously explained.

Then, the following procedure, similar to the Direct Linear Transformation (DLT) (Hartley and Zisserman, 2003, Abdel-Aziz et al., 2015), can be used in order to assess M_0 (and consequently the point position during all the time instants when it has been tracked).

First, without loss of generality assume that the lens distortion has already been corrected, and hence that the camera can be modeled as a pinhole camera. Then, the relation between the measured pixel coordinates (u_t, v_t) and the 3D point M_t is:

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \approx P_t \begin{bmatrix} M_t \\ 1 \end{bmatrix} \tag{2}$$

where $\approx$ stands for an equality up to a scaling factor.

Let $p_{t,\cdot}$, $\rho_{t,\cdot}$ and $w_{t,\cdot}$ be defined as it can be derived from the following:

$$P_t = \begin{bmatrix} p_{t,1}^\top \\ p_{t,2}^\top \\ p_{t,3}^\top \end{bmatrix} = \begin{bmatrix} \rho_{t,1}^\top & w_{t,1} \\ \rho_{t,2}^\top & w_{t,2} \\ \rho_{t,3}^\top & w_{t,3} \end{bmatrix} \tag{3}$$

then, from (2) and the equation above:

$$u_t p_{t,3}^\top \begin{bmatrix} M_t \\ 1 \end{bmatrix} = p_{t,1}^\top \begin{bmatrix} M_t \\ 1 \end{bmatrix} \tag{4}$$

$$v_t p_{t,3}^\top \begin{bmatrix} M_t \\ 1 \end{bmatrix} = p_{t,2}^\top \begin{bmatrix} M_t \\ 1 \end{bmatrix} \tag{5}$$

Given the similarity of the two equations above, let's focus just on the first one, substituting $p_{t,\cdot}^\top$ with $\rho_{t,\cdot}^\top$ and $w_{t,\cdot}$:

$$u_t \begin{bmatrix} \rho_{t,3}^\top & w_{t,3} \end{bmatrix} \begin{bmatrix} M_t \\ 1 \end{bmatrix} = \begin{bmatrix} \rho_{t,1}^\top & w_{t,1} \end{bmatrix} \begin{bmatrix} M_t \\ 1 \end{bmatrix} \quad (6)$$

Then, after some trivial computations:

$$(u_t\rho_{t,3}^\top - \rho_{t,1}^\top)R_{t,c}M_0 = (\rho_{t,1}^\top - u_t\rho_{t,3}^\top)t_{t,c} + w_{t,1} - u_t w_{t,3} \quad (7)$$

$$(v_t\rho_{t,3}^\top - \rho_{t,2}^\top)R_{t,c}M_0 = (\rho_{t,2}^\top - v_t\rho_{t,3}^\top)t_{t,c} + w_{t,2} - u_t w_{t,3} \quad (8)$$

Finally, let A and b as defined below:

$$A = \begin{bmatrix} (u_t\rho_{t,3}^\top - \rho_{t,1}^\top)R_{t,c} \\ (v_t\rho_{t,3}^\top - \rho_{t,2}^\top)R_{t,c} \end{bmatrix} \quad (9)$$

$$b = \begin{bmatrix} (\rho_{t,1}^\top - u_t\rho_{t,3}^\top)t_{t,c} + w_{t,1} - u_t w_{t,3} \\ (\rho_{t,2}^\top - v_t\rho_{t,3}^\top)t_{t,c} + w_{t,2} - u_t w_{t,3} \end{bmatrix} \quad (10)$$

Hence, M_0 can be estimated as:

$$M_0 = A^\dagger b \quad (11)$$

where $A^\dagger$ is the pseudo-inverse of A.

Clearly, this method allows to obtain just an approximate value of M_0, which could be optimized as the solution of a proper bundle adjustment step.

4. RESULTS

An initial assessment of the vision-based trajectory estimation is provided in Fig. 3 for the car in Fig. 2 in a 60 s time interval. Fig. 3 compares the vision-based estimated trajectory (shown with blue dot marks) with the GNSS-based reference solution (yellow line). Table summarizes the characteristics of the positioning error.

Fig. 4 shows the comparison, on a 30 s interval, between the vision-based estimated car speed (red dot marks) and the GNSS-based reference velocity (blue solid line). It is worth to notice that the two solutions are provided at 5 Hz in the GNSS case, whereas at 30 Hz in the vision one. Fig. 4 apparently shows the similarity between the two solutions. The numerical comparison between the assessed speed (in the previously mentioned 60 s interval) and the reference one leads to a median error of 0.00 m/s, and a median absolute deviation between the two of 0.05 m/s. The speed error distribution is shown in Fig. 5.

m_x [cm]	D_x [cm]	m_y [cm]	D_y [cm]
5	25	0	30

Table 1. Positioning error (median m and median absolute deviation D) along x and y directions.

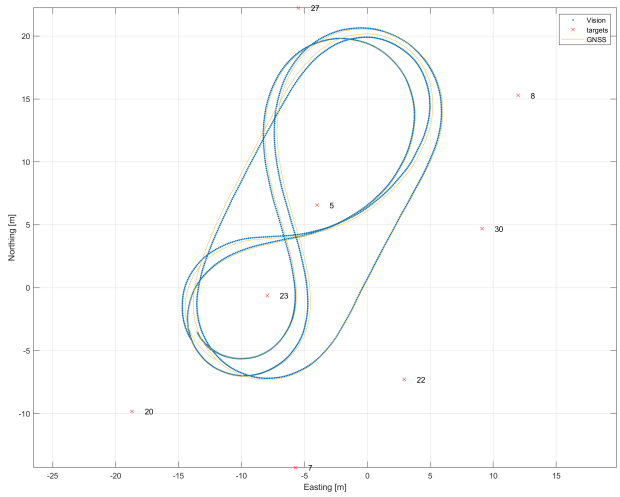


Figure 3. Assessment of the estimated car trajectory: comparison of the vision-based estimated trajectory (blue dot marks) with the GNSS-based reference solution (yellow line).

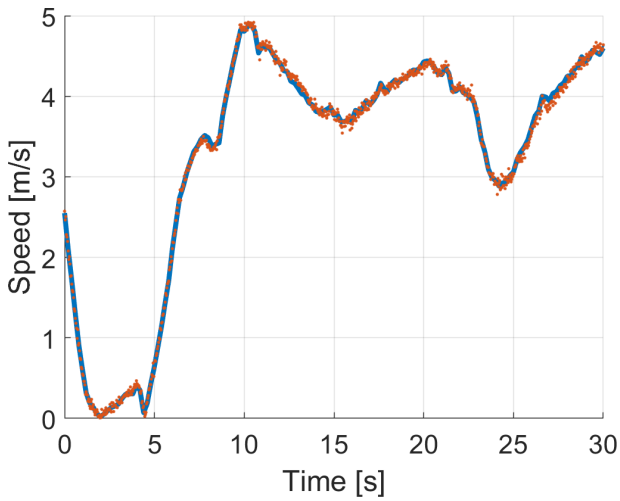


Figure 4. Comparison between the vision-based estimated car speed (red dot marks) and the GNSS-based reference velocity (blue solid line).

The procedure to determine an approximate vehicle 3D shape (e.g. the values of M_0 for a set of feature points) is tested on a synthetic example: 400 points have been sampled from a LiDAR point cloud of the vehicle and synthetic noisy measurements (affected by Gaussian noise of standard deviation of 3 pixels, on both the x and y image pixel coordinates) were simulated on 150 frames, with the vehicle moving on a portion of the trajectory shown in Fig. 3.

Fig. 6 and 7 compare, from two different points of view, the LiDAR point cloud with the 3D point positions assessed as described in Section 3, showing a reasonable compatibility between the shapes of the two clouds.

Finally, Fig. 8 compares (cloud-to-cloud point distance distribution) the LiDAR scan of the vehicle and vision-based point cloud.

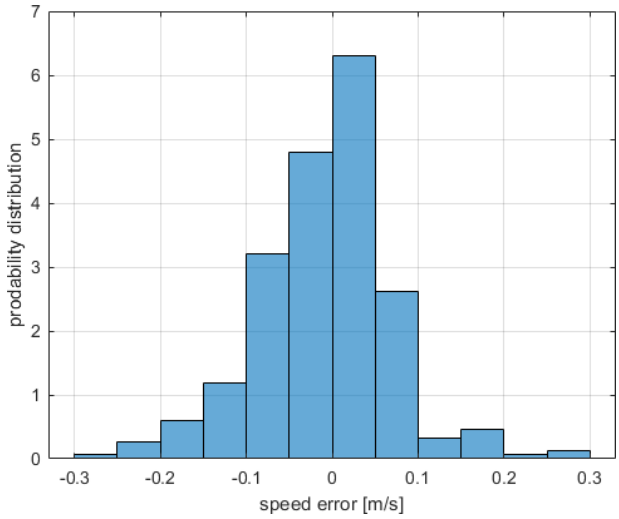
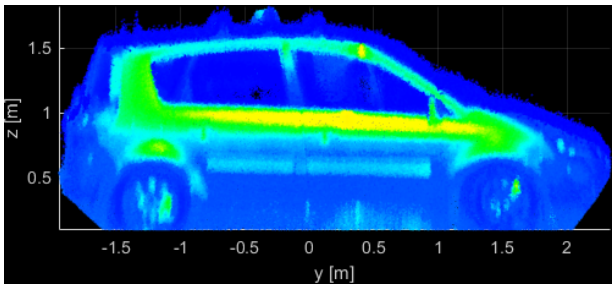
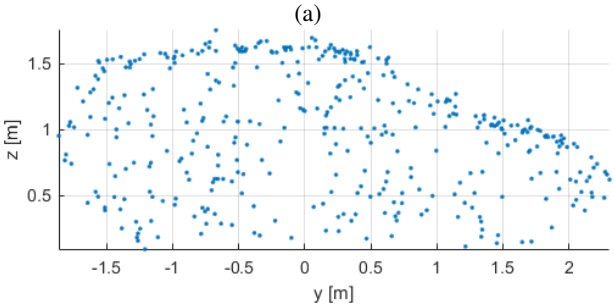


Figure 5. Speed error distribution.



(a)



(a)

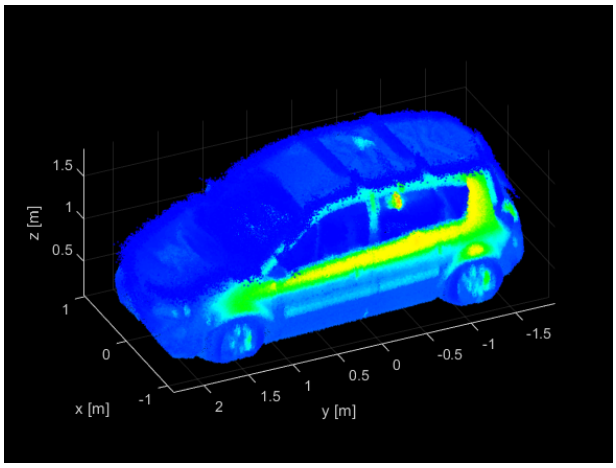
Figure 6. Car shape (view 1): (a) LiDAR scan, (b) vision-based assessment.

5. DISCUSSION

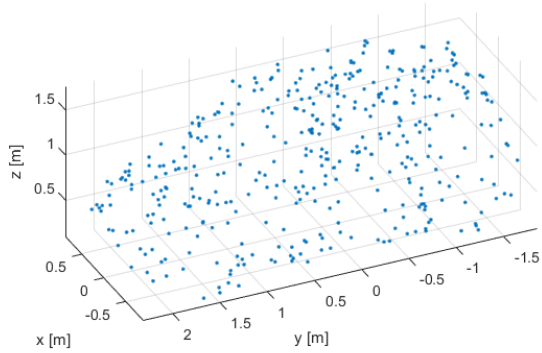
The obtained results show that the driver trajectories can be determined with the proposed method with a median error of few centimeters (with a variability of the error at few decimeter level (Table 1)).

In addition to the vehicle location, other parameters characterizing the vehicle dynamics can be computed, such as the vehicle speed (half width at half maximum of the speed error distribution is less than 0.1 m/s, as shown in Fig. 5).

Furthermore, the proposed method to assess the 3D shape of the vehicle provided quite reasonable results in the considered synthetic example, with most of vision-based reconstructed 3D points at less than 5 cm from the LiDAR point cloud, as shown in Fig. 8. A more in depth investigation on the performance of the proposed method on real data shall be considered in our future investigations.



(a)



(a)

Figure 7. Car shape (view 2): (a) LiDAR scan, (b) vision-based assessment.

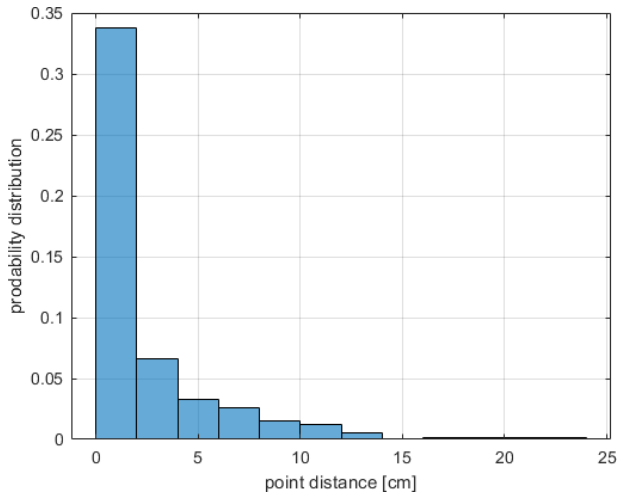


Figure 8. Comparison of LiDAR scan of the vehicle and vision-based assessment: cloud-to-cloud point distance distribution.

Since varying the flight altitude influences the UAS imagery spatial resolution and the visible area, it shall have a great impact on the obtained level of tracking accuracy and, on the other, on the number of vehicles visible in the image. Hence, our fu-

ture works shall also include an analysis of the proposed system performance variation as a function of the flight altitude.

6. CONCLUSIONS

The results presented in this paper show that mini-UAS imagery can be quite effectively used in order to track vehicle trajectories in urban environments, and to reliably extract certain information of interest (e.g. vehicle speeds).

Furthermore, the proposed method for assessing the 3D shape of the tracked vehicles led to quite satisfactory results in the considered synthetic example. Nevertheless, a more in-depth investigation should be carried out on real data.

A more detailed analysis of the obtained results varying the UAS altitude and the vehicle speed shall also be considered in our future works, along with an investigation on the main factors influencing the obtained estimation errors.

REFERENCES

Abdel-Aziz, Y. I., Karara, H. M., Hauck, M., 2015. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Photogrammetric engineering & remote sensing*, 81(2), 103–107.

Balletti, C., Guerra, F., Tsioukas, V., Vernier, P., 2014. Calibration of Action Cameras for Photogrammetric Purposes. *Sensors*, 14(9), 17471–17490.

Bock, J., Krajewski, R., Moers, T., Runde, S., Vater, L., Eckstein, L., n.d. The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. *2020 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 1929–1934.

Chiang, K.-W., Pai, H.-Y., Zeng, J.-C., Tsai, M.-L., El-Sheimy, N., 2022. Automated Modeling of Road Networks for High-Definition Maps in OpenDRIVE Format Using Mobile Mapping Measurements. *Geomatics*, 2(2), 221–235.

Coifman, B., McCord, M., Mishalani, R. G., Iswalt, M., Ji, Y., 2006. Roadway traffic monitoring from an unmanned aerial vehicle. *IEE Proceedings-Intelligent Transport Systems*, 153number 1, IET, 11–20.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric computer vision*. Springer.

Fraser, C., 2018. Camera calibration considerations for UAV photogrammetry. *ISPRS TC II Symposium: Towards Photogrammetry 2020*.

Habib, A., Morgan, M., 2003. Automatic calibration of low-cost digital cameras. *Optical Engineering*, 42(4), 948–955.

Hartley, R., Sturm, P., 1997. Triangulation. *Computer Vision and Image Understanding*, 68(2), 146–157.

Hartley, R., Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Kanistras, K., Martins, G., Rutherford, M. J., Valavanis, K. P., 2013. A survey of unmanned aerial vehicles (uavs) for traffic monitoring. *2013 International Conference on Unmanned Aircraft Systems (ICUAS)*, IEEE, 221–234.

Khan, M. A., Ectors, W., Bellemans, T., Janssens, D., Wets, G., 2017. UAV-based traffic analysis: A universal guiding framework based on literature survey. *Transportation research procedia*, 22, 541–550.

Krajewski, R., Bock, J., Kloeker, L., Eckstein, L., 2018. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2118–2125.

Krajewski, R., Moers, T., Bock, J., Vater, L., Eckstein, L., 2020. The round dataset: A drone dataset of road user trajectories at roundabouts in germany. *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 1–6.

Kraus, K., 2007. *Photogrammetry: geometry from images and laser scans*. 1, Walter de Gruyter.

Kurz, F., Mendes, P., Gstaiger, V., Bahmanyar, R., d'Angelo, P., Azimi, S. M., Auer, S., Merkle, N. M., Henry, C., Rosenbaum, D. et al., 2022. Generation of Reference Vehicle Trajectories in real-world Situations using Aerial Imagery from a Helicopter. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Lucas, B. D., Kanade, T. et al., 1981. An iterative image registration technique with an application to stereo vision. *Proceedings of the International Joint Conference on Artificial Intelligence*, 674—-679.

Luhmann, T., Fraser, C., Maas, H.-G., 2015. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 37–46.

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., Dabove, P., Di Pietra, V., Piragnolo, M., Vettore, A., Guarnieri, A., Toth, C., Gikas, V., Perakis, H., Chiang, K.-W. et al., 2022. A comparison between UWB and laser-based pedestrian tracking. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

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.

Mikhail, E. M., Bethel, J. S., McGlone, J. C., 2001. *Introduction to modern photogrammetry*. John Wiley & Sons.

Puri, A., Valavanis, K., Kontitsis, M., 2007. Statistical profile generation for traffic monitoring using real-time uav based video data. *2007 Mediterranean Conference on Control & Automation*, IEEE, 1–6.

Remondino, F., Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-5, 266–272.

Shi, J., Tomasi, C., 1994. Good features to track. *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, IEEE, 593–600.