The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W18-2025
Symposium on GeoSpatial Technologies: Visions and Horizons 2025 (GeoVisions2025), 8–10 October 2025, Çanakkale, Türkiye

Synthetic and real data for training and testing ground camera-based UAV tracking

Andrea Masiero[1,*], Paolo Dabove[2], Vincenzo Di Pietra[2], Marco Piragnolo[1], Alberto Guarnieri[1], Antonio Vettore[1], Charles Toth[3]

[1] University of Padova, Italy - (andrea.masiero, marco.piragnolo, antonio.vettore, alberto.guarnieri)@unipd.it
[2] Politecnico di Torino, Italy - (paolo.dabove, vincenzo.dipietra)@polito.it
[3] The Ohio State University, US - toth.2@osu.edu

Keywords: Tracking, camera, photogrammetry, Kalman filter

Abstract

The high versatility and affordable price made Unmanned Aerial Vehicles (UAVs) very popular in many civil applications. Despite their usage often takes advantage of automatic flight mode, this flight modeality is usually limited to the good working conditions of Global Navigation Satellite Systems (GNSS) positioning. Extending their operability to cases when GNSS is not available or reliable, including for instance departure and landing in not open-sky areas, is clearly of great importance for the drone market. This paper aims at investigating the performance of a vision-based system to monitor, and support, if needed, UAV movements, within some tens of meters from a ground camera, used to track the UAV. In this work, a uncalibrated camera is used to track the UAV: UAV detection on the camera frames is implented with a background subtraction approach, even if neural network-based approaches can be used as well. Then, the UAV centroid on the camera frames along with its area in pixels are used as inputs for the machine learning predictors. While using an uncalibrated camera is clearly suboptimal in terms of performance, it eases the usage of the proposed method in for non-expert operators. The proposed approach is tested both on a synthetic and a real dataset, collected at the Agripolis campus of the University of Padua, in order to determine whether the performance is limited by the size of available training dataset. The results reported in this work show that the usage of tree ensamble regression can lead to submeter errors in tracking a UAV with a ground camera, when the UAV is at less than approximately 100 meters from the camera.

1. Introduction

Unmanned Aerial Vehicles (UAVs) have become widely popular for outdoor applications due to their versatility and cost-effectiveness, particularly when Global Navigation Satellite Systems (GNSS) positioning is reliable. However, the need to expand their usability in areas with poor GNSS coverage has led to the development of alternative positioning systems, able to compensate the GNSS unreliability in certain critical conditions. These systems often integrate different kind of sensors, including cameras (Nabavi-Chashmi et al., 2023), LiDAR (Opromolla et al., 2016), ultrasound (Paredes et al., 2017), UWB (Masiero et al., 2017, Zahran et al., 2019), and RADAR (Barra et al., 2022, Zahran et al., 2018), sometimes even utilizing multi-platform communication and localization (Masiero et al., 2023a, Masiero et al., 2023b).

Cameras have emerged as a popular choice among these alternative sensors, thanks to their affordability and usability in different contexts. For instance, when mounted on a UAV, they can serve for both navigation and mapping purposes. This capability is exploited in Simultaneous Localization and Mapping (SLAM, (Leonard and Durrant-Whyte, 1991, Strasdat et al., 2012)) approaches (Macario Barros et al., 2022, Kazerouni et al., 2022), which may be further enhanced with supplementary sensor data to improve localization accuracy (Mateos-Ramirez et al., 2024).

Achieving absolute positioning using a drone-mounted camera typically requires the usage of additional reference points, such as landmarks, targets, or ground control points in aerial photogrammetry (Förstner and Wrobel, 2016, Kraus, 2011). Alternatively, cameras installed in the flying area can be used to monitor and localize UAVs, similarly to motion capture systems (Masiero et al., 2019), though this usually necessitates multiple cameras for accurate tracking (Islam et al., 2020, Mustafah et al., 2012).

Among the cases of interest for alternative positioning systems for UAVs, it is worth to mention precise navigation for landing in small areas without reliable GNSS, which is clearly a crucial application that demands alternative sensor support. This scenario also applies to UAV monitoring in surveillance applications, where various sensors, including RADAR, acoustic arrays, and cameras, have been employed. Actually, the rising incidence of UAV-related aerial incidents, particularly near airports, has also motivated numerous studies on drone detection and monitoring (Wang et al., 2020, Shanliang et al., 2022).

Camera-based UAV detection primarily relies on background-foreground segmentation methods (assuming a relatively static background, (Seidaliyeva et al., 2020)), or, more recently, on deep learning techniques, often utilizing YOLO (You Only Look Once, (Redmon et al., 2016, Redmon and Farhadi, 2017, Hussain, 2023)) networks (Jiang et al., 2022, Aydin and Singha, 2023). Artificial intelligence methods can also be applied to various other UAV-related tasks (Rahman et al., 2024, Yan et al., 2023).

This study focuses on using a single (external) ground camera installed in the flying area to track UAV movements. While the vision-based detection method is similar to those mentioned earlier, this approach employs just one uncalibrated camera, unlike multi-camera systems, hence making the system easier to use to non-highly qualified operators. To be more specific, the proposed method is based on the usage of a simple background-subtraction approach for determining the foreground-background separation, to a Kalman filter, running

* Corresponding author

on image frames, for 2D tracking and data association, and, finally, on the usage of machine learning tools to estimate the UAV position based on some geometric features extracted from image frames, e.g. the centroid and the area of the blob on the image frames corresponding to the tracked UAV.

This paper presents both some experimental results, obtained for a dataset collected at the University of Padua (Italy), and some on a synthetic dataset, in order to just test the machine learning predictors, independently of the influence of errors coming from the vision-based detection and the uncalibrated camera.

The paper is organized as follows: first, the proposed method is introduced in Section 2, then Section 3 presents the case study considered to test the proposed method, 4 shows the obtained results, and, finally, some conclusions are drawn in Section 5.

2. UAV tracking with a ground-camera

First, it is assumed that the UAV is flying in an area visible by the camera: while the method is independent of being an outdoor or indoor scenario, the UAV visibility is a key working condition in order to make the method effective. Furthermore, the scene is assumed to be mostly static, as in (Seidaliyeva et al., 2020). Indeed, in this working dynamic objects, such as UAVs, can be easily detected via background subtraction: moving objects can be determined by applying a minimum threshold to the absolute difference between the current frame and a past one.

Some outlier measurements can be rejected by imposing some regularity restrictions on the detected object trajectory and on the object area. To be more specific, first, minimum and maximum thresholds are imposed on the detected object area in order to be potentially considered a UAV. Then, a Kalman filter is used to track the trajectories of the detected objects on the image plane. In particular, trajectory regularity restrictions are imposed, limiting the trajectories performed by UAVs as those that at each frame are not too far from the predicted locations on the image plane. When multiple objects are detected in the same frame, such Kalman filter is used for data association as well: each track is continued with the 2D location closed to its Kalman prediction.

It is worth to notice that the above mentioned method for object tracking on the image plane may not be able to distinguish between UAVs and other dynamic objects/birds. To such aim, many studies already considered the usage of deep learning-based methods, in particular those exploiting YOLO-based approaches (Jiang et al., 2022, Aydin and Singha, 2023). Since several works already faced this problem, such aspect is not investigated here. Instead, the reader is referred to (Jiang et al., 2022, Aydin and Singha, 2023) and the references therein.

Once that a UAV is detected on the image frame, a machine learning-based regression is exploited in order to estimate the UAV 3D location in the desired reference frame. To such aim, a set of features is extracted from the image frame and fed into the used machine learning tool. In particular, the centroid of the UAV region on the image plane is extracted, and, intuitively, used to determine the direction of the UAV with respect to the camera reference frame. Then, the area of the UAV region on the image frame is extracted as well, being such area inversely proportional to the square of the UAV-camera distance.

A supervised approach has been used to implement the above mentioned procedure: a learning dataset, where the UAV ground truth position is assumed to be known as well, is used to properly train the machine learning-based regressor. To be more specific, a different machine learning regressor is independently trained on each direction, x, y, z. Gaussian kernel regression, Support Vector Machine (SVM) based regression and tree ensemble regression (100 regression trees aggregated by means of the Least Squares Boosting algorithm) are considered, for comparison, as machine learning-based regressions. Without loss of generality, ground truth UAV locations are assumed to be expressed as UTM (Universal Transverse Mercator) projected coordinates, however, the adaptation of the proposed approach to different choices is trivial. Then, the trained model(s) can be used to predict the UAV position in new camera frames.

3. Case study

The method presented in the previous section is tested on a dataset collected at the Agripolis campus of the University of Padua. Such dataset has been collected during a data acquisition campaign by a joint team of researchers, participating to the Italian PRIN 2022 project PAIN AND GAIN, and within the IAG WG 4.1.5 “Wireless positioning with terrestrial instruments”.

A DJI Matrice 210 UAV was used during the test, being filmed by a Sony alpha ILCE-5100 camera (video resolution 1440×1080 pix, at 25 Hz), mounted on tripod, positioned on the ground, at a few tens of meters from the fight area. The DJI Matrice 210 flew for approximately 15 minutes, along the trajectory shown in Fig. 1, at different altitudes, but being still visible by the camera, when in its field of view. Fig. 2 shows the distribution of UAV-camera distances during the flight. Fig. 3 shows a frame captured by the Sony alpha ILCE-5100 during the flight of the DJI Matrice 210 UAV.

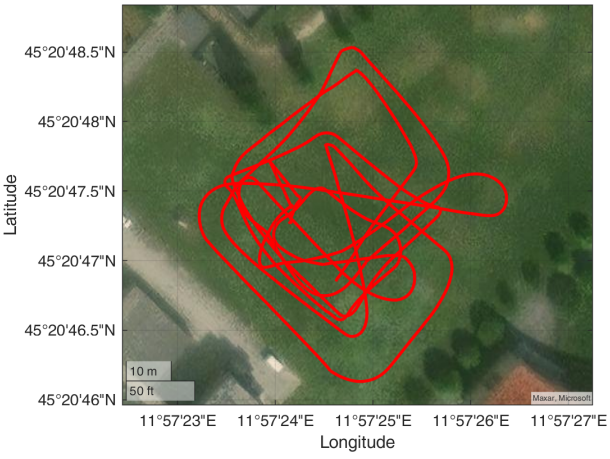

Figure 1. Track of DJI Matrice 210 during the test (red line).

Reference positions of the UAV were collected by a GNSS receiver mounted on the UAV, working in NRTK mode at 5 Hz, with the closest permanent station of the used network at less than 100 m from the flight area. Only fixed solutions were used (either in the training or validation of the method) in this work. Before being fed into the machine learning regression models, UTM projected coordinates were computed (ETRS89-UTM32), and locally shifted to make them more easily readable.

The camera collected ≈22500 frames during the flight. The UAV was visible on approximately 70% of them, hence 2/3 of

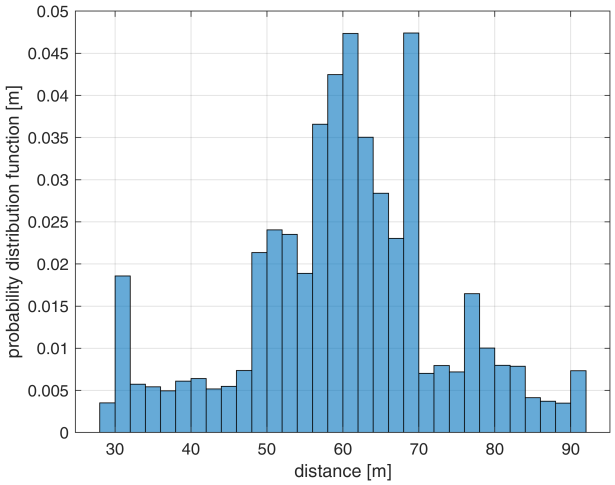

Figure 2. Distribution of DJI Matrice 210-camera distances during the test.

Figure 3. DJI Matrice 210 flying during the test, shown in a Sony alpha ILCE-5100 frame.

such frames were randomly selected to be used in the training phase, whereas the remaining ones were used for validation.

4. Results

4.1 Experimental results

The three considered machine learning methods (Gaussian kernel regression, SVM-based regression, tree ensemble regression) were compared on the collected dataset, leading to the positioning results shown in Table 1, where mean and Median Absolute Deviation (MAD) error values are shown along the x, y, z directions.

The tree ensemble regression model performed apparently better than the other two considered methods. The mean errors along the x, y, z directions were very close to zero, whereas the error variability, assessed with MAD, in of some tens of centimeters.

The sources of such errors can be both the vision-based UAV detection methods and the implemented machine learning-based regression. Since the experimental results shown in this subsection does not allow to separate such two contributions, the following subsection is dedicated to determining the method performance on synthetic data, where the only factor influencing the error is the implemented machine learning-based regression method.

Fig. 5 shows the estimation error temporal correlation, on a sequence 100 s long.

	Gaussian kernel regression		
	x [m]	**y [m]**	**z [m]**
mean	1.12	0.31	0.84
MAD	9.63	10.67	4.28
	SVM-based regression		
	x [m]	**y [m]**	**z [m]**
mean	2.02	-3.54	2.07
MAD	4.43	8.83	3.90
	Tree ensemble regression		
	x [m]	**y [m]**	**z [m]**
mean	-0.04	0.02	0.00
MAD	0.56	0.64	0.13

Table 1. Comparison of positioning errors for different machine learning regressors on experimental data.

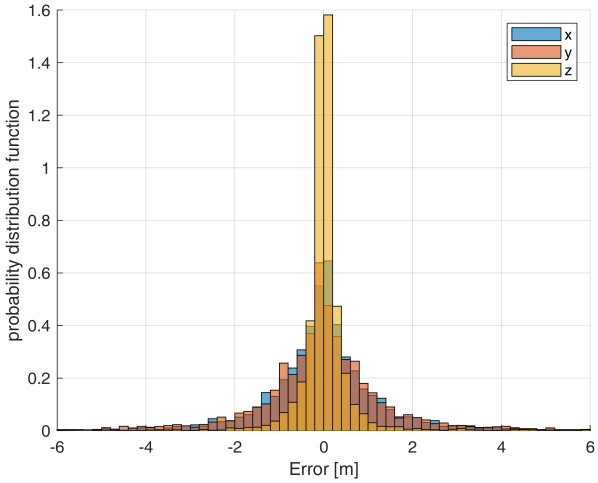

Figure 4. Positioning error distribution along the x, y and z directions, obtained with one camera tracking a drone flying on an area of approximately 60 m $\times$60 m, using a tree ensemble regression.

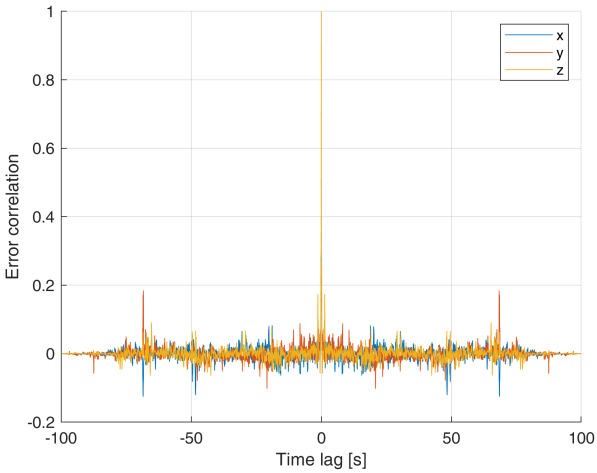

Figure 5. Positioning error temporal correlation computed on a 100 s interval.

4.2 Results on a synthetic dataset

This subsection aims at evaluating the performance of the proposed machine learning-based regression method on synthetic

data, in order to bypass the vision-based error unavoidable in the experimental data. Since SVM and Gaussian kernel regressions proved to perform much worse than the tree ensemble one on experimental data, only the latter is considered in this subsection.

The UAV positions were randomly uniformly generated on a 100 m $\times$ 100 m $\times$ 20 m region, with minimum height above the ground of 10 m and minimum camera-UAV distance of $\approx$43.6 m. Camera model was assumed to be the pinhole one, without any distortion. Pixel size was set sufficiently small to have a negligible impact on the results.

The training dataset size was of 100 k random samples, whereas the regression model was tested on a different dataset of 100 k random samples. The distance probability distribution function for the training samples is shown in Fig. 6.

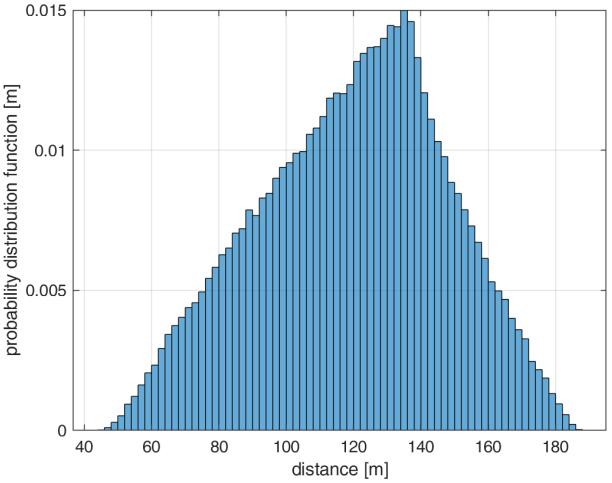

Figure 6. Distribution of UAV-camera distances in the synthetic training dataset.

The results obtained on the test dataset are shown in Table 2, whereas the probability density distribution of the error, along the x, y and z directions, is shown in Fig. 7.

	Tree ensemble regression		
	x [m]	**y [m]**	**z [m]**
mean	0.00	0.00	0.00
MAD	0.68	0.73	0.26

Table 2. Positioning errors for the tree ensemble regression model on synthetic data.

5. Discussion and conclusions

The proposed positioning method employs a single, uncalibrated ground camera to track a UAV, when visible by the camera, using machine learning techniques. Among the three compared machine learning regressors, the tree ensemble one performed significantly better, reaching close to zero average positioning error along all the three directions, both in experimental and synthetic data, and sub-meter error variability (MAD), i.e. around 0.6 meters along x and y, while smaller on the z direction.

Since the experimental results shown in the paper do not allow to separate the error due to machine learning regression from the one caused by the vision-based UAV detection, the proposed method has been tested on synthetic data as well, showing similar results to those obtained on experimental data. Such results show that the current performance level is primarily limited by the machine learning regression method, rather than the vision factors. This observation suggests that there is some potential for further improvements, if a more effective machine learning method is implemented.

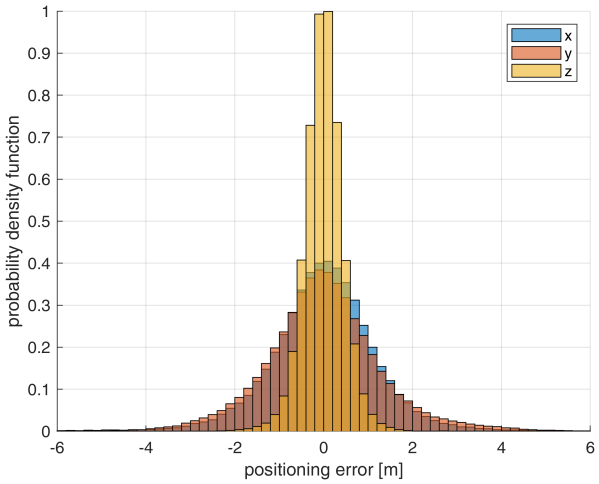

Figure 7. Positioning error distribution along the x, y and z directions, for the tree ensemble regression model, on synthetic data.

In addition to the above observation, it is also worth to notice that temporal regularity of the UAV trajectory has not been exploited in the method implemented for estimating the UAV position. The close-to-zero error along all the three directions and the fact that the errors appear to be temporally low-correlated (Fig. 5) suggest that some smoothing, for instance via Kalman filtering, shall help improving the obtained results.

Additional input features or alternative regression models can also be viable ways to further improve the model regression capability. In particular, including deep learning approaches may potentially show a superior regression performance.

Increasing the dataset size may also help improving the training results, however, training on a larger synthetic dataset showed a performance similar to the one obtained on the experimental data.

While this work only investigated the use of the proposed method on just one type of UAV, extending its usability to different UAVs can clearly be valuable in real applications and it will be considered in our future works. Such generalization will probably explicitly or implicitly involve a UAV classification step, such as in (Rahman et al., 2024), which however is expected to be challenging for large UAV-camera distances.

Using multiple cameras, high-resolution cameras and/or calibrated cameras will also be investigated in our future works: such options are expected to ensure some further improvements on the positioning performance, however implying the usage of more complex/expensive systems.

Finally, UAV detection based on neural networks (e.g. YOLO-like networks) will also be implemented in order to make the proposed procedure more robust to the presence of dynamic objects/birds in the scene.

Acknowledgements

This work is supported by the Italian PRIN 2022, project PAIN AND GAIN-Positioning And INtelligent Alarms supported by a New Dense GNSS Affordable Infrastructure, MUR code 2022P8C7ZA (PRIN 2022-DD 104, 02/02/2022)-CUP B53D23007380006.

References

Aydin, B., Singha, S., 2023. Drone detection using YOLOv5. *Eng*, 4(1), 416–433.

Barra, J., Creuzet, T., Lesecq, S., Scorletti, G., Blanco, E., Zarudniev, M., 2022. Micro-drone ego-velocity and height estimation in gps-denied environments using an FMCW MIMO radar. *IEEE Sensors Journal*, 23(3), 2684–2692.

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

Hussain, M., 2023. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines*, 11(7), 677.

Islam, A., Asikuzzaman, M., Khyam, M. O., Noor-A-Rahim, M., Pickering, M. R., 2020. Stereo vision-based 3D positioning and tracking. *IEEE Access*, 8, 138771–138787.

Jiang, Y., Jingliang, G., Yanqing, Z., Min, W., Jianwei, W., 2022. Detection and tracking method of small-sized uav based on yolov5. *2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 1–5.

Kazerouni, I. A., Fitzgerald, L., Dooly, G., Toal, D., 2022. A survey of state-of-the-art on visual SLAM. *Expert Systems with Applications*, 205, 117734.

Kraus, K., 2011. *Photogrammetry: geometry from images and laser scans*. Walter de Gruyter.

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.

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., Fissore, F., Antonello, R., Cenedese, A., Vettore, A., 2019. A comparison of UWB and motion capture UAV indoor positioning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 1695–1699.

Masiero, A., Fissore, F., Vettore, A., 2017. A low cost UWB based solution for direct georeferencing UAV photogrammetry. *Remote Sensing*, 9(5), 414.

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., Remondino, F., 2023b. Vision and uwb-based collaborative positioning between ground and uas platforms. *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 748–754.

Mateos-Ramirez, P., Gomez-Avila, J., Villaseñor, C., Arana-Daniel, N., 2024. Visual odometry in gps-denied zones for fixed-wing unmanned aerial vehicle with reduced accumulative error based on satellite imagery. *Applied Sciences*, 14(16), 7420.

Mustafah, Y. M., Azman, A. W., Akbar, F., 2012. Indoor UAV positioning using stereo vision sensor. *Procedia Engineering*, 41, 575–579.

Nabavi-Chashmi, S.-Y., Asadi, D., Ahmadi, K., 2023. Image-based UAV position and velocity estimation using a monocular camera. *Control Engineering Practice*, 134, 105460.

Opromolla, R., Fasano, G., Rufino, G., Grassi, M., Savvaris, A., 2016. Lidar-inertial integration for uav localization and mapping in complex environments. *2016 international conference on unmanned aircraft systems (ICUAS)*, IEEE, 649–656.

Paredes, J. A., Álvarez, F. J., Aguilera, T., Villadangos, J. M., 2017. 3D indoor positioning of UAVs with spread spectrum ultrasound and time-of-flight cameras. *Sensors*, 18(1), 89.

Rahman, M. H., Sejan, M. A. S., Aziz, M. A., Tabassum, R., Baik, J.-I., Song, H.-K., 2024. A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions. *Remote Sensing*, 16(5), 879.

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.

Redmon, J., Farhadi, A., 2017. Yolo9000: better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271.

Seidaliyeva, U., Akhmetov, D., Ilipbayeva, L., Matson, E. T., 2020. Real-time and accurate drone detection in a video with a static background. *Sensors*, 20(14), 3856.

Shanliang, L., Yunlong, L., Jingyi, Q., Renbiao, W., 2022. Airport uav and birds detection based on deformable detr. *Journal of Physics: Conference Series*, 2253number 1, IOP Publishing, 012024.

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

Wang, L., Ai, J., Zhang, L., Xing, Z., 2020. Design of airport obstacle-free zone monitoring UAV system based on computer vision. *Sensors*, 20(9), 2475.

Yan, X., Fu, T., Lin, H., Xuan, F., Huang, Y., Cao, Y., Hu, H., Liu, P., 2023. UAV Detection and Tracking in Urban Environments Using Passive Sensors: A Survey. *Applied Sciences*, 13(20). https://www.mdpi.com/2076-3417/13/20/11320.

Zahran, S., Masiero, A., Mostafa, M., Moussa, A., Vettore, A., El-Sheimy, N., 2019. UAVs enhanced navigation in outdoor GNSS denied environment using UWB and monocular

camera systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 665–672.

Zahran, S., Mostafa, M., Masiero, A., Moussa, A., Vettore, A., El-Sheimy, N., 2018. Micro-RADAR and UWB aided UAV navigation in GNSS denied environment. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1, 469–476.