algorithms

MDPI

Article

From Stationary to Nonstationary UAVs: Deep-Learning-Based Method for Vehicle Speed Estimation

Muhammad Waqas Ahmed [1], Muhammad Adnan [1,*], Muhammad Ahmed [2], Davy Janssens [1], Geert Wets [1], Afzal Ahmed [3] and Wim Ectors [1]

[1] UHasselt, Transportation Research Institute (IMOB), Martelarenlaan 42, 3500 Hasselt, Belgium; muhammadwaqas.ahmed@uhasselt.be (M.W.A.); davy.janssens@uhasselt.be (D.J.); geert.wets@uhasselt.be (G.W.); wim.ectors@uhasselt.be (W.E.)
[2] Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan; muhammadahmed@neduet.edu.pk
[3] Institute of Transportation Studies, University of Leeds, Leeds LS2 9JT, UK; a.ahmed@leeds.ac.uk
* Correspondence: muhammad.adnan@uhasselt.be

Abstract: The development of smart cities relies on the implementation of cutting-edge technologies. Unmanned aerial vehicles (UAVs) and deep learning (DL) models are examples of such disruptive technologies with diverse industrial applications that are gaining traction. When it comes to road traffic monitoring systems (RTMs), the combination of UAVs and vision-based methods has shown great potential. Currently, most solutions focus on analyzing traffic footage captured by hovering UAVs due to the inherent georeferencing challenges in video footage from nonstationary drones. We propose an innovative method capable of estimating traffic speed using footage from both stationary and nonstationary UAVs. The process involves matching each pixel of the input frame with a georeferenced orthomosaic using a feature-matching algorithm. Subsequently, a tracking-enabled YOLOv8 object detection model is applied to the frame to detect vehicles and their trajectories. The geographic positions of these moving vehicles over time are logged in JSON format. The accuracy of this method was validated with reference measurements recorded from a laser speed gun. The results indicate that the proposed method can estimate vehicle speeds with an absolute error as low as 0.53 km/h. The study also discusses the associated problems and constraints with nonstationary drone footage as input and proposes strategies for minimizing noise and inaccuracies. Despite these challenges, the proposed framework demonstrates considerable potential and signifies another step towards automated road traffic monitoring systems. This system enables transportation modelers to realistically capture traffic behavior over a wider area, unlike existing roadside camera systems prone to blind spots and limited spatial coverage.

Keywords: UAV; drone; traffic monitoring; computer vision; YOLO

check for updates

Citation: Ahmed, M.W.; Adnan, M.; Ahmed, M.; Janssens, D.; Wets, G.; Ahmed, A.; Ectors, W. From Stationary to Nonstationary UAVs: Deep-Learning-Based Method for Vehicle Speed Estimation. *Algorithms* **2024**, *17*, 558. https://doi.org/10.3390/a17120558

Academic Editor: Massimiliano Caramia

Received: 23 October 2024
Revised: 29 November 2024
Accepted: 4 December 2024
Published: 6 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Recently, unmanned aerial vehicles (UAVs) or drones have gained substantial attention due to the offered level of automation, cost-effectiveness, and mobility [1]. These characteristics have led to the widespread use of drones in various fields such as agriculture, earth observation, geology, and climatology [2]. With continuous technological advancements, UAVs are now being used for applications beyond just reconnaissance and remote sensing. Modern drones are being utilized for activities such as agrochemical applications, maritime rescue, and firefighting [3,4]. Recently, drones have also been employed in package delivery, logistics, and humanitarian aid, often in remote locations where human access is restricted or posed by severe risk [4]. UAVs have been widely used by the military for land mine detection and reconnaissance missions [5]. This proves the utility and versatility of UAVs in military and civil applications alike. There are several types of UAVs offering utility to different use cases. These types include fixed-wing UAVs, single-rotor drones, multirotor

(quadcopters, hexacopters, and octocopters), and fixed-wing hybrid VTOL UAV systems [6]. Fixed wing drones offer great utility for package deliveries and remote inspections, while multirotor drones are greatly considered for search and rescue operations due to their hovering capabilities [7]. For recreational purposes, such as photography and high-resolution aerial imaging, smaller drones equipped with professional-grade cameras are used [8]. The applications, pros, and cons of each drone type are discussed in Table 1.

Table 1. Drone types, some of their civil applications (non-exhaustive), advantages, and disadvantages.

Drone Type	Advantages	Disadvantages	Uses
Multirotor UAVs	- Vertical take-off and landing (VTOL) - Hovering enabled - User-friendliness	- Shorter flight durations - Smaller payload capacity	- Aerial inspection, thermal reports, and 3D scans.
Fixed-Wing UAVs	- Increased coverage area - Extended flight time - Enhanced speed	- No hovering capability - Difficult for novice pilots - Higher costs	- Aerial mapping, precision agriculture, surveillance, and construction.
Single-Rotor UAVs	- Hovering enabled - Greater endurance - VTOL - Greater payload capabilities	- Difficult for novice pilots - Higher costs	- Aerial LIDAR laser scan and drone surveying.
Fixed-Wing Hybrid UAVs	- Vertical take-off and landing - Long-endurance flight	- Best of both worlds: with a little trade-off in hovering and forward flight.	- Deliveries/logistics.

Among the popular civil applications of drones, road traffic monitoring (RTM) systems have witnessed significant development. An RTM system primarily focuses on two tasks: detecting road accidents and identifying traffic congestion [9]. However, traditional surveillance methods lack the aerial perspective of UAVs, limiting a comprehensive analysis [10]. With the integration of global navigation satellite systems (GNSS), UAVs offer researchers a geospatial viewpoint, enabling them to conduct meaningful research in the field [11].

In recent times, there have been significant advancements in vehicle detection methods through the use of computer vision and deep learning techniques [12]. These technologies have greatly enhanced the capabilities of object detection and tracking methods, which are vital for tasks such as estimating vehicle trajectories and analyzing traffic flow [13]. Without accurate speed measurement, it is unviable to implement an accurate RTM system. Despite these advancements in computer vision, there are still technical limitations that need to be addressed [14]. Much of the existing literature focuses on RTM systems that rely on fixed camera systems with limited spatial coverage. In contrast, a moving drone can provide increased mobility, better spatial coverage, and reduced blind spots. Additionally, the current speed estimation techniques used by law enforcement only capture a single point speed (using LiDAR-based systems), which may not be sufficient for comprehensive analysis and could hinder the decisionmaker's ability to implement appropriate traffic control measures.

This study aims to enhance the existing systems by providing more accurate speed measurements and trajectory estimations by utilizing AI and UAVs. This method offers a practical solution that effectively works with both stationary and nonstationary aerial footage, demonstrating remarkable flexibility. This method is capable of accurately mapping vehicle trajectories in real geographical space. Furthermore, it shows high precision in measuring velocity, with an error margin as low as 0.53 km/h. Implementing this solution can provide significant value for intelligent road traffic monitoring systems.

2. Related Works

Recently, a multitude of research literature has been published tackling a similar problem but focusing on fixed cameras. Computer vision has wider applicability within road traffic monitoring systems and road safety. The biggest challenge of using these solutions is the real-world practical implementation [15]. In their study, the authors of [16] present a real-time vehicle velocity estimation method using a vehicle-mounted monocular camera. The authors' approach involves a computer-vision-based object tracker combined with a small neural network to estimate the velocity of the vehicles from the tracked bounding boxes. To calculate the distance traveled by the vehicles, the authors use the focal length of the monocular camera, a fixed height, and the bottom edge midpoint of the bounding box. The method yields promising results with a vehicle estimation error of 1.28 km/h, but the major limitation is the practicality of the experimental setup itself, which is extremely inflexible, acting as a barrier to real-world implementation. A related study [17] showcased a system that uses a stereo camera and a high-speed, high-precision semantic segmentation model. With the proposed system, authors could estimate relative speeds by measuring changes in distance and time differences between frames. The proposed approach adds value due to its segmentation methodology, which captures more information than the one-stage object detectors. In similar research [18], the authors developed an experimental setup with small vehicles to test the accuracy of a preexisting model that estimates vehicle speeds. The speed calculations are validated by comparing the measurements obtained with the reference measurements recorded from an infrared sensor. The experimental results also provided insights to the frame-skipping threshold to reduce the processing time of the overall footage—a direction toward real-time implementation. The authors planned to test this system on real vehicular traffic.

Optical occlusions are a barrier to the real-world implementation of vision-based systems. Hernández Martínez and Fernandez Llorca [19] tried to address this problem by creating an experimental setup that utilizes multiple cameras positioned at different angles, coupled with a complex 3D convolutional neural network (CNN) architecture. The study yielded promising results, paving the way for view-invariant vehicle speed measurement systems. You Only Look Once (YOLO) is a single-stage object detection algorithm that has received widespread attention in various fields and holds tremendous potential for traffic monitoring tasks. In a study by Peruničić and Djukanović [20], the authors used the YOLOv5 for vehicle detection and tracking, while employing an RNN for speed estimation. The proposed system achieved an error rate of 4.08 km/h, significantly lower than the acoustic, sensor-based measurements. The authors further discussed the prospects of a multimodal system—combining audio and video data to improve accuracy.

Similar to fixed camera systems, researchers have also explored the prospects of combining UAVs with intelligent systems to estimate vehicle tracks and speeds. In a study by Chen and Zhao [21], the potential of UAVs in RTM systems was explored. The experiment was conducted by collecting and analyzing traffic footage taken from varying altitudes and resolutions and implementing a YOLO architecture for detection and tracking. The study also discusses the limitations faced by nonstationary camera systems, which includes camera calibration resulting in inconsistencies in speed estimations. The proposed method achieved an accuracy of 1.1 km/h, which is remarkable but, like other research works, is implemented only on a stationary camera or UAV. To develop a practically viable, vision-based RTM system, scientists have been exploring the right balance between accuracy and computational efficiency. The available edge computing systems can offer scalability, but the major challenge is developing a system that is accurate and fast enough to enable real-time or at least near-real-time processing. This challenge is discussed in detail by Tran and Pham [15], utilizing 20 single camera views and a lightweight deep learning architecture coupled with edge computing devices. The authors utilized a fixed camera setup coupled with different edge devices including Nvidia Jetson TX2, Nvidia Xavier NX, and Nvidia AGX Xavier. The proposed method is effective despite some limitations, e.g., detection accuracy, optical occlusions caused by nearby reflective surfaces, and inherent

real-world implementation challenges. Similarly, the modularity of UAVs has enabled scientists to mount edge computing systems on drone systems. One of these experiments includes a DJI Phantom 3 retrofitted with an NVIDIA Xavier NX system under the moniker of a "MultEYE" system designed especially for real-time vehicle detection, tracking, and speed estimation. The system consists of a YOLOv4 detector coupled with a minimum output sum of squared error (MOSSE) tracker. To estimate the vehicle speed, the onboard system calculated ground sampling distances by taking the camera's UAV altitude, pixel width, and focal length. The mean average error observed is remarkably low, with a figure of 1.13 km/h.

UAVs have the potential to not only reduce optical occlusions encountered but also provide an aerial perspective, greater mobility, and freedom for enhanced spatial coverage. Previous research works exhibit great potential for fixed cameras and hovering UAVs but lack the practical implementation for moving camera platforms. Our research aims to devise a solution that works with both stationary and nonstationary aerial footage with notable accuracy. The developed solution can also log the vehicle trajectories geospatially, enabling superior analytical capabilities and offering value for intelligent RTM systems.

3. Data and Methods

The proposed methodology enables robust traffic analysis using computer vision and geospatial data analysis techniques to detect, track, and map objects. It begins with preparing a video and an orthomosaic reference image. The orthomosaic was developed using the WebODM (version 2.5.0) application of OpenDroneMap™. The experiment footage was collected in the daytime with clear sky conditions using a DJI mini 3 pro (Shenzhen, China) at 4K (3840 $\times$ 2160) resolution with a frame rate of 30 FPS. For the template-matching algorithm to function effectively, conducting the experiment in daylight is essential. For validation, Pro Laser III speed gun manufactured by Kustom Signals Inc. (Owensboro, KY, USA) was used. To synchronize the video with the speed gun, measurements were taken five times (as discussed in Table 3), and parts with speed gun measurements served as a benchmark for validation. Subsequently, SIFT (scale-invariant feature transform) is used for feature recognition on the reference image and georeferencing. The key difference between stationary and moving drone footage lies in the frequency of the pixel calibration. For stationary drone footage, georeferencing is performed on the first frame and remains constant throughout the entire video. In contrast, for moving drone footage, pixel calibration must occur for each frame, because the homography of the input frames changes with the drone's movement and variations in altitude. By replicating the homography of the template image, it is possible to dynamically adjust the calibration of the cells despite the drone's movement and shifts in altitude. YOLOv8 model is then applied to detect objects within each video frame. The next crucial step involves using a transformation matrix to translate pixel-based coordinates into real-world geographical locations, connecting visual data with physical geography. The method tracks object movements across frames, logging their geographical positions and other key parameters, such as ID, class, and frame of appearance, which enables the real-time velocity measurement. Figure 1 provides a step-by-step illustration of the workflow.

3.1. Automated Georeferencing and Pixel Coordinate Conversion

Reducing the manual intervention of pixel recalibration is important to develop an automated system. Since the experiment allows for the free movement of UAVs, it is vital to utilize tools that can enable an automated workflow. This was achieved by creating an automated georeferencing system based on the scale-invariant feature transform (SIFT) algorithm developed by Lowe [22]. SIFT works by detecting the most consistent features between two images that are resilient to rotation, scaling, and lighting variations. After identifying the most stable key points, a dominant orientation is assigned. Then, it creates a 128-dimensional descriptor for each key point, capturing detailed information about local image gradient magnitudes and orientations. A descriptor is formed by divid-

ing the region around the key point into smaller subregions and creating histograms of gradient orientations. Using these descriptors, SIFT can identify key points across various images, facilitating functionalities, such as object detection, merging images, and creating 3D models. In contrast to the other feature-matching algorithms, like SURF and ORB, SIFT is slower but has superior resilience to variation in pixel intensities, making it ideal for applications with temporal variations, like georeferencing [23].

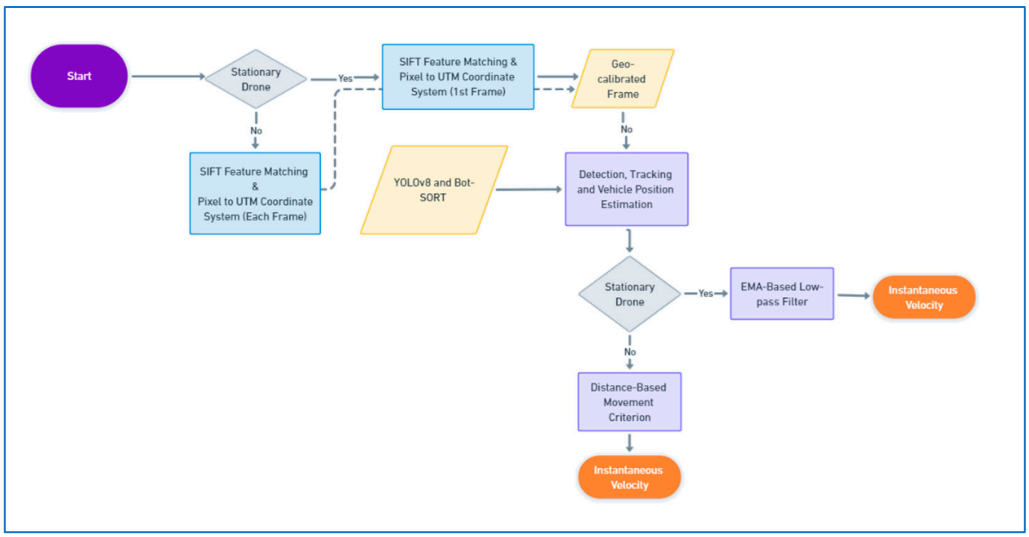


Figure 1. Showcases the methodological framework of the study.

An orthomosaic template image was created to project geographical space onto the input footage using a DJI mini 3 pro drone flying at 90 m, equipped with a 48-megapixel camera set at an angle of 90 degrees. A visual line of sight was maintained throughout the flight, and several images were taken to generate orthomosaics over a larger area. The image was then georeferenced using the WebODM application of OpenDroneMap, and an orthomosaic was constructed with a UTM projection system. This georeferenced image acted as a template to automatically transfer geographical coordinates onto each frame. The process involved identifying the matching features between a template image and the input frame and then georeferencing the input image based on the matching key points (as illustrated in Figure 2). For the quality assurance of descriptors, Lowe's ratio test was employed, and only the matches passing the criteria were used for the homography calculation [24]. The root mean square error (RMSE) threshold was also enforced as an additional check, and frames with a higher RMSE were discarded [25].

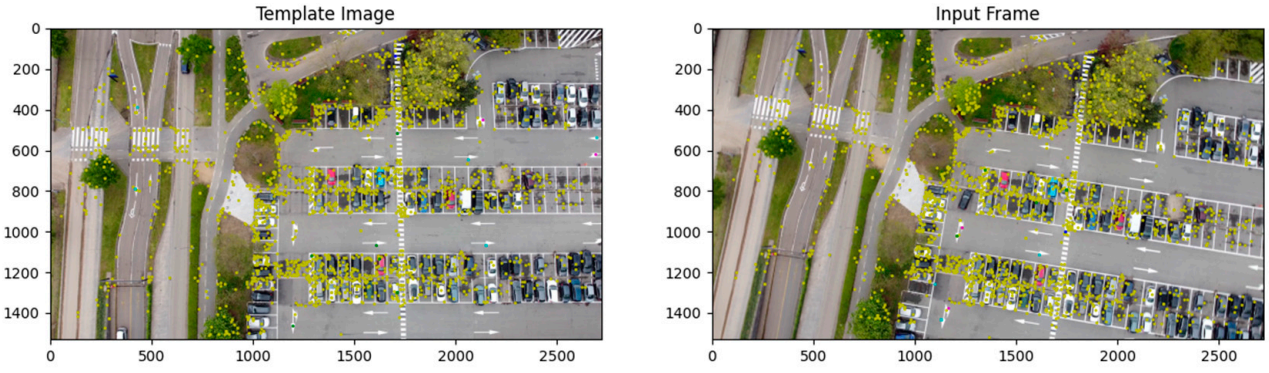


Figure 2. Feature-matching algorithm SIFT applied to input and template image. The highlighted markers depict the key points matched between the two images.

3.2. Vehicle Detection and Tracking

The YOLO algorithm is a significant advancement in real-time object detection, introducing the concept of single-stage detection. It works by dividing the input image into a grid and predicting both bounding boxes and class probabilities from each cell. This unique grid-based approach enables YOLO to perform predictions swiftly, making it an ideal choice for real-time applications, especially on UAV platforms with limited resources [26]. Over time, YOLO has undergone several improvements (discussed in Table 2), resulting in various versions, with each iteration delivering noteworthy improvements in both accuracy and speed. The 8th generation YOLO architecture has gained widespread attention due to its improved identification capabilities and has been widely tested in several scenarios by academia and industry. The decision to use YOLOv8 over its subsequent iterations, like the 10th generation YOLO architecture, was also motivated by the superior performance it exhibited in previous research works in detecting larger vehicle classes, such as cars, vans, and trucks. This contrasts with the 10th generation YOLO, which exhibits improved detection capabilities for smaller objects [27].

YOLOv8, by default, utilizes Bot-SORT for object tracking, which possesses the ability to reidentify objects even if they temporarily disappear, ensuring continuous and accurate object tracking, which is crucial for applications requiring uninterrupted tracking of objects over time [28,29]. The multiobject tracking algorithm enables the proposed workflow to record the speeds of multiple vehicles simultaneously.

Table 2. The evolution of the You Only Look Once (YOLO) algorithm over the years [30,31].

Version	Year of Release	Strengths and New Features
YOLOv1	2015	- Real-time object detection. - Regression-based approach for bounding box and class probability prediction.
YOLOv2	2016	- Batch normalization and anchor boxes.
YOLOv3	2018	- More efficient backbone network and spatial pyramid pooling.
YOLOv4	2020	- Enhanced with mosaic data augmentation and other upgrades.
YOLOv5	2020	- Hyperparameter optimization and improved performance.
YOLOv6	2022	- Popular for Meituan's delivery robots.
YOLOv7	2022	- Introduced pose estimation capabilities.
YOLOv8	2023	- Quick feature fusion. - Improved object identification
YOLOv9	2024	- Introduces innovative methods like programmable gradient information and generalized efficient layer aggregation network (GLEAN).
YOLOv10	2024	- IOU-free inference. - Enhanced inference speed.

In the proposed study, the YOLOv8 model was trained on the VisDrone2019-DET dataset with average precision AP@0.5 of 64% for the class of interest; cars, which is sufficient for the experiment as the vehicle used for speed measurement remained consistently detected and tracked throughout the input footage. The VisDrone dataset is specifically

designed for object detection in aerial images [21]. The standard Bot-SORT tracker was utilized for object tracking. To prevent overestimations of bounding boxes and the double detection of vehicles close to each other, the intersection over union (IoU) threshold was set to 0.3. This threshold was determined to be sufficient, considering the controlled traffic in the experimental footage. Detection and tracking accuracy are crucial for accurate trajectory extraction. Tracking inaccuracies can introduce noise into vehicle trajectories and consequently impact the speed measurements. However, this noise can be filtered out by implementing a low-pass filter, such as the exponential moving average (EMA) (further discussed in the Section 4). For geospatial trajectory mapping, the vehicle tracks were identified, and the pixel coordinates were converted into corresponding geographical coordinates using a transformation matrix obtained by implementing the automatic georeferencing workflow using SIFT. The geographical coordinates of tracks across each frame were stored along with other relevant information, including the frame number, distance traveled, track ID, and class ID, in JSON format. The vehicle velocities were calculated and compared with observations taken from a speed gun using the logged information from the object tracking.

4. Results and Discussion

The experimental results were evaluated based on the speed and positional accuracy of the vehicle tracks by drawing comparisons against the LiDAR-based speed gun measurements and manually drawn vector maps. Speed measurements were conducted under three conditions: a stationary drone, a drone moving at 5 m/s while following the vehicle track, and a drone moving at 10 m/s. The findings indicated that vehicle velocities estimated from the stationary drone had the highest accuracy, exhibiting a minimal absolute error of 0.53 m/s (as discussed in Table 3). However, this error increased as the drone's speed increased. Notably, at higher drone speeds, some discrepancies within the footage were observed. These discrepancies can be attributed to factors such as the increased drone speed, varying wind conditions, and the inherent limitations of gimbal stabilization systems. To address these challenges and improve data quality, we propose adjustments, such as modifying the drone's altitude to cover a broader area and optimizing the UAV's speed. These adjustments are expected to enhance the accuracy of vehicle speed estimations under varying operational conditions.

Table 3. The velocity measurements obtained from this workflow with measurements taken from speed guns at various UAV altitudes and speeds.

#	Speed Gun (km/h)	Proposed Method (km/h)	Absolute Error	UAV Altitude (m)	UAV Speed (m/s)
1	26	25.47	0.53	65	-
2	26	25.18	0.82	65	5
3	30	29.44	0.55	65	5
4	34	33.33	0.67	65	5
5	35	37.5	2.5	50	10

The second challenge of this proposed method is the presence of jumpy vehicle tracks, mainly due to changing inference confidence and the proximity of the detected vehicles. In the case of stationary drone footage, a low-pass filter using exponential moving averages (EMA) was implemented on the centroidal coordinates of the vehicle tracks to stabilize the recorded velocities. EMA stabilizes the abrupt changes in the initial positions reducing fluctuations in velocity estimates. EMA is applied to vehicle positions to ensure the vehicle tracks are smooth and representative of the real-world situation. The smoothing factor α was set to 0.1, significantly dampening fluctuations within the vehicle tracks. A higher α becomes more sensitive to fluctuation and increases variability, as shown in Figure 3.

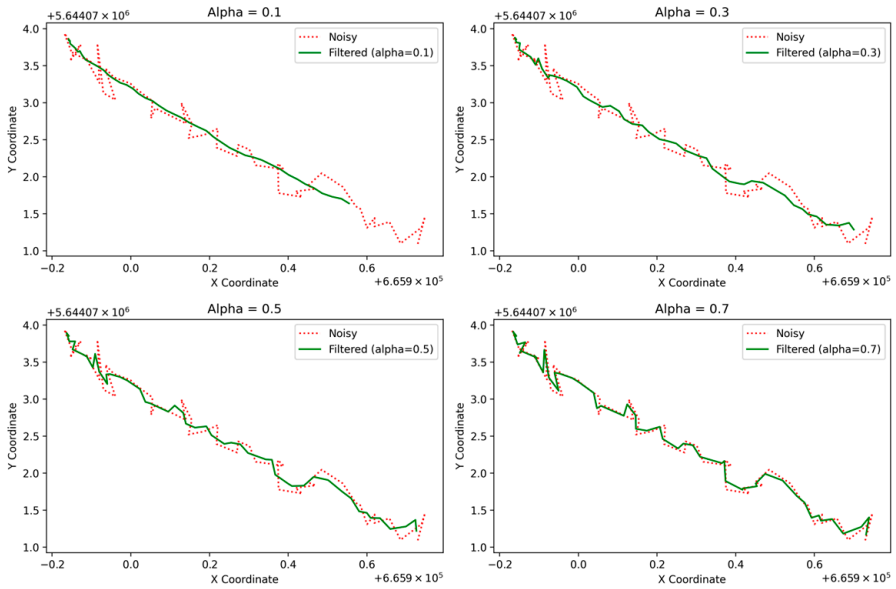


Figure 3. Comparison of noisy and EMA-filtered trajectories with different alpha values.

After applying the exponential moving average (EMA), the positioning of the vehicle tracks aligns consistently with the actual vehicle movements, resulting in more stabilized velocity measurements. Initial velocity measurements are closer to zero, because the filter is applied directly to the positions instead of velocity measurements; therefore, it requires the first few positions to identify movement. This delay can be reduced by increasing the α but could lead to increased variability in the velocity measurements. Figure 4 shows the original tracks in yellow and the corrected tracks in red after EMA application with $\alpha = 0.1$. Resultingly, the vehicle velocity measurements were also stabilized (illustrated in Figure 5).

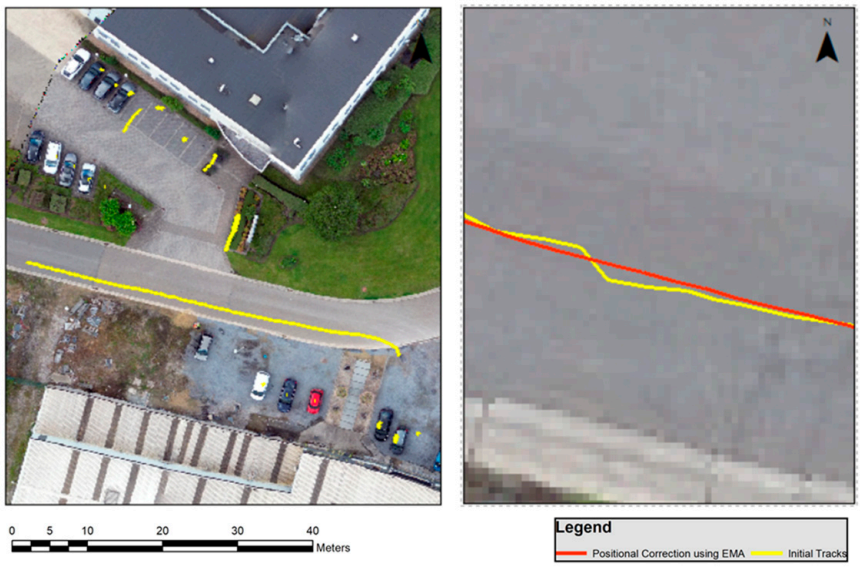


Figure 4. The mapped vehicle trajectories before and after EMA application.

Correcting the vehicle tracks in the stationary drone footage is straightforward. However, tracking vehicles in moving drone footage presents greater challenges, as the drone's movement introduces additional motion, affecting both stationary and moving objects (as shown in Figure 6). In these instances, a more rigorous approach is necessary for effectively removing noise from vehicle tracks.

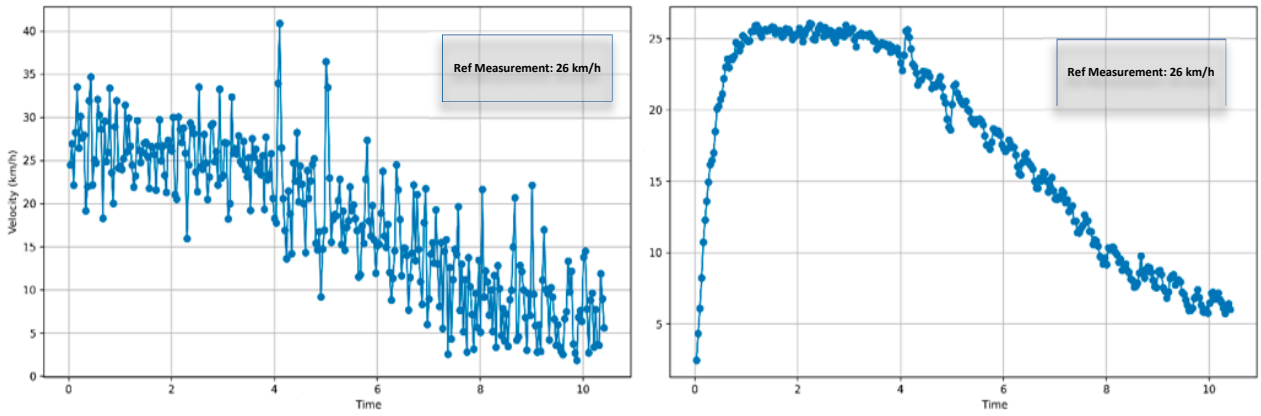


Figure 5. The fluctuations in velocity (in km/h) over time (in seconds) and the removal of errors using an EMA-based low-pass filter ($\alpha = 0.1$). The single-point reference speed measured by the speed gun was 26 km/h.

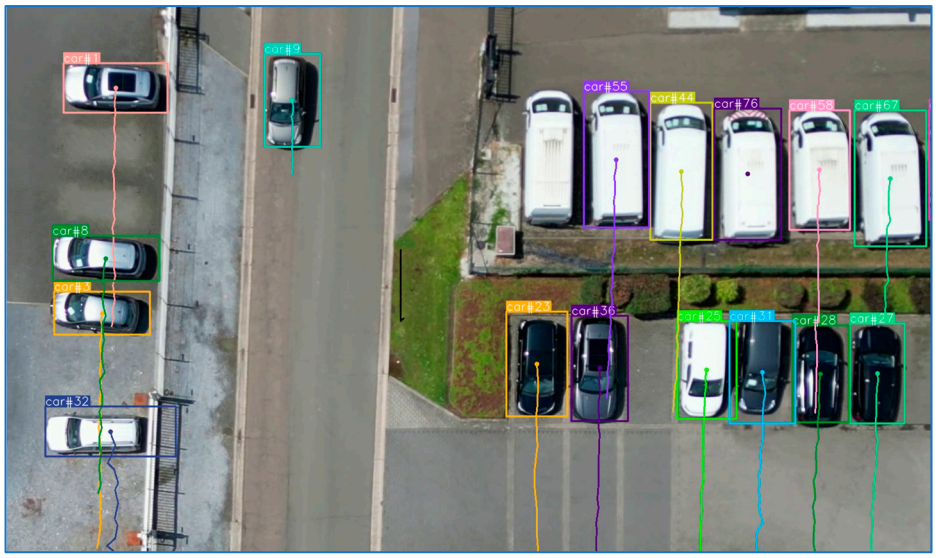


Figure 6. The pseudo tracks generated by the object tracking algorithm due to UAV movement.

The added movements, along with georeferencing errors, can notably influence the precision of the vehicle tracks, potentially leading to exaggerated velocity measurements (see Figure 7). This problem is not resolved with a low-pass filter. Instead, a distance-based movement threshold was implemented to decrease positional inaccuracies and, consequently, refine the velocity measurements. While this approach does introduce a certain level of discretization in the output, it is a solution aimed at enhancing the overall accuracy of vehicle tracking in nonstationary UAV footage. This limitation, however, also opens up a valuable opportunity for further research. It highlights the need for innovative solutions that can improve the positional accuracies in nonstationary UAV footage without the discretization of valuable information.

In the absence of a reference vehicle position, the buffer overlay method can be used to measure the accuracy of mapped vehicle trajectories. A vector path of the actual vehicle path was drawn manually, considering the target vehicle's position with respect to time, and a buffer of 1 m was constructed (illustrated in Figure 8). Then, the tracks generated by the proposed method were compared with the ground-truthing buffer, calculating the total length inside the buffer.

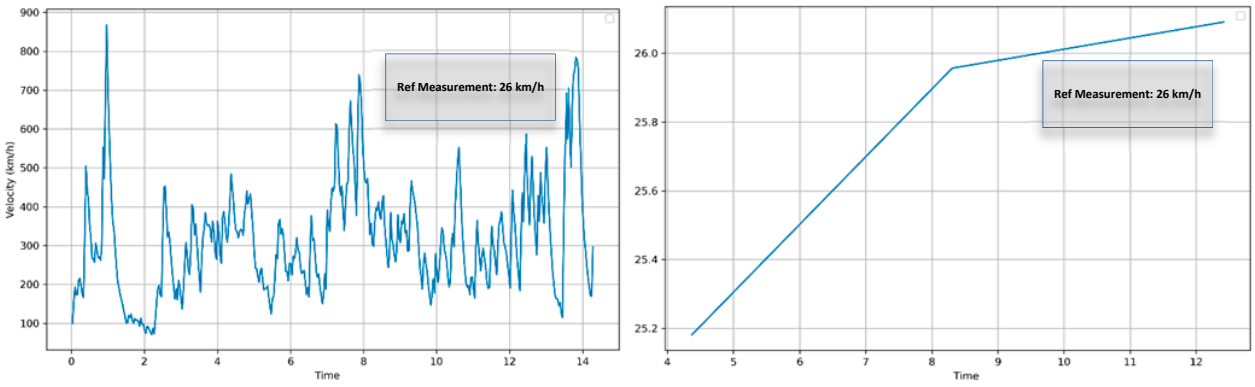


Figure 7. Extreme velocity (km/h) over time (s) with fluctuation resulting from pseudo tracks and their removal from the distance-based movement threshold (after introducing the distance threshold, the first measurement starts at 4.3 s). The single-point reference speed measured by the speed gun was 26 km/h.



Figure 8. The method used for determining the positional accuracies of vehicle tracks on (**a**) tracks from stationary drone footage and (**b**) tracks from moving drone footage.

In the comparative analysis, it was observed that positional accuracies depend on the speeds of the UAVs. The tracks extracted from nonstationary drone footage were notably accurate and consistent. However, tracks obtained from nonstationary footage displayed minor positional inaccuracies, which tended to increase with the optical destabilization caused by higher UAV speeds. For example, track 09 was 81% inside the 1 m buffer, and track 13 was 61% inside the buffer, as detailed in Table 4.

Table 4. Comparative analysis of the positional accuracy of the vehicle tracks in three different drone settings.

Track	Setting	UAV Speed (m/s)	Track Length (m)	Track Length (m) (Inside Buffer) *
02	Stationary	-	52	52
09	Nonstationary	05	93	76
13	Nonstationary	10	128	78

* Ground truthing buffer has a width of 1 m. Higher buffer widths are attributed to relaxation in the evaluation criterion.

The computational expense of processing nonstationary drone footage is a significant limitation. On a system with an Intel® Core™ i9-9900 CPU @ 3.60 GHz and 64 GB RAM, the average processing time for nonstationary drone footage was 63 s per frame, compared with just 0.42 s for stationary drone footage. The longer processing time for nonstationary footage is due to the computationally expensive SIFT application for each frame. However, this can be reduced by 8% through the use of precalculated key points and descriptors. Georeferencing also causes errors, resulting in jumpy locations across frames, requiring aggressive noise removal methods, such as movement threshold. Nonetheless, the proposed system can accurately estimate vehicle speed and position. Future research will be focused on computational optimization techniques and innovative data-denoising methods for improving output quality in nonstationary drone footage.

5. Conclusions

This study demonstrates the potential of combining artificial intelligence (AI) and unmanned aerial vehicles (UAVs) to improve road traffic monitoring systems, specifically in estimating vehicle speed and trajectory—a novel method using advanced feature matching and deep learning techniques alongside UAV technology. The experimental findings confirm that UAV-based systems equipped with AI can overcome many limitations of existing RTM systems and provide more accurate speed measurements compared with point-based estimations. The proposed system offers near-real-time speed when applied to stationary drone footage; although, there is a trade-off in processing speed with dynamic drone footage. Improving the processing speed could make the system more scalable in all cases. Drones' ability to provide a mobile aerial perspective adds a valuable dimension to traffic analysis, offering more comprehensive coverage and detail. Moreover, the use of AI for automating vehicle detection and tracking has been shown to reduce the need for manual intervention, making the process more efficient and accurate. This advancement is crucial for practically feasible RTM systems, where swift and accurate data analysis and insights are essential. Despite the promising results, the study acknowledges the inherent challenges of developing a system that is both efficient and fully adaptable to real-world conditions. UAV-based operations are only feasible during clear daylight hours and cannot be conducted at night or in extreme weather situations. Additionally, the range limitations of drones and battery life restrict perpetual flight, meaning this system should be viewed as a supplementary solution to ground-based, fixed camera systems. The aerial perspective provided by drones offers significant advantages, such as covering larger areas and enhanced maneuverability. Integrating a UAV-based RTM system can yield substantial benefits. Given the accuracy of the measurements that the system can provide, it adds considerable value. This method is particularly useful for short-term traffic monitoring in potential conflict zones and helping to understand road user behavior. By analyzing this behavior from an aerial perspective, life-saving safety interventions can be implemented; however, environmental factors, such as bird migration routes, must be considered during aerial surveillance. Future research will focus on refining this system by incorporating multisource data, including ground and aerial surveillance footage, for more comprehensive analysis. Furthermore, efforts will be made to enhance processing speeds

and to implement methods that prevent data loss caused by the current error removal techniques used in nonstationary drone experiments.

In conclusion, this work demonstrates the significant advantages of using UAVs and AI in road traffic monitoring, representing a step forward in the pursuit of safe and efficient transportation systems. As technology advances, integrating these smart systems holds the promise of revolutionizing how we understand and manage road traffic, ultimately contributing to better, more responsive urban environments.

Author Contributions: Conceptualization, M.W.A. and W.E.; methodology, M.W.A. and W.E.; software, M.W.A.; validation, W.E.; formal analysis, M.W.A.; investigation, M.W.A. and W.E.; resources, W.E., M.A. (Muhammad Adnan), D.J. and G.W.; data curation, M.W.A. and W.E.; writing—original draft preparation, M.W.A.; writing—review and editing, M.W.A. and W.E.; visualization, M.W.A.; supervision, W.E. and M.A. (Muhammad Ahmed); project administration, W.E., D.J., M.A. (Muhammad Adnan), M.A. (Muhammad Ahmed), G.W. and A.A.; funding acquisition, W.E., D.J. and G.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by BOF-BILA program of UHasselt, grant number 14406 (BOF24BL02).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy concerns.

Acknowledgments: The authors would like to express their sincerest gratitude to the BOF/BILA program of UHasselt for funding this research. Additionally, we would like to thank our colleague, Farhan Jamil, for his assistance during the experiment with the speed gun.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Boukoberine, M.N.; Zhou, Z.; Benbouzid, M. A critical review on unmanned aerial vehicles power supply and energy management: Solutions, strategies, and prospects. *Appl. Energy* **2019**, *255*, 113823. [CrossRef]
2. Tewes, A. Investigating the Potential of UAV-Based Low-Cost Camera Imagery for Measuring Biophysical Variables in Maize. Ph.D. Thesis, Universitäts-und Landesbibliothek Bonn, Bonn, Germany, 2018.
3. Karbowski, J. Using a drone to detect plant disease pathogens. *Int. Multidiscip. Sci. Geoconf. SGEM* **2022**, *22*, 455–462.
4. Bogue, R. Beyond imaging: Drones for physical applications. *Ind. Robot. Int. J. Robot. Res. Appl.* **2023**, *50*, 557–561. [CrossRef]
5. Anil Kumar Reddy, C.; Venkatesh, B. Unmanned Aerial Vehicle for Land Mine Detection and Illegal Migration Surveillance Support in Military Applications. In *Drone Technology: Future Trends and Practical Applications*; Scrivener Publishing LLC: Beverly, MA, USA, 2023; pp. 325–349. [CrossRef]
6. Garg, P. Characterisation of Fixed-Wing Versus Multirotors UAVs/Drones. *J. Geomat.* **2022**, *16*, 152–159. [CrossRef]
7. Sönmez, M.; Pelin, C.-E.; Georgescu, M.; Pelin, G.; Stelescu, M.D.; Nituica, M.; Stoian, G.; Alexandrescu, L.; Gurau, D. Unmanned aerial vehicles—Classification, types of composite materials used in their structure and applications. In Proceedings of the 9th International Conference on Advanced Materials and Systems, Bucharest, Romania, 26–28 October 2022.
8. Heiets, I.; Kuo, Y.-W.; La, J.; Yeun, R.C.K.; Verhagen, W. Future Trends in UAV Applications in the Australian Market. *Aerospace* **2023**, *10*, 555. [CrossRef]
9. Elloumi, M.; Dhaou, R.; Escrig, B.; Idoudi, H.; Saidane, L.A. Monitoring road traffic with a UAV-based system. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018; pp. 1–6.
10. Butilă, E.V.; Boboc, R.G. Urban traffic monitoring and analysis using unmanned aerial vehicles (UAVs): A systematic literature review. *Remote Sens.* **2022**, *14*, 620. [CrossRef]
11. Nonami, K. Prospect and recent research & development for civil use autonomous unmanned aircraft as UAV and MAV. *J. Syst. Des. Dyn.* **2007**, *1*, 120–128.
12. Zhou, S.; Xu, H.; Zhang, G.; Ma, T.; Yang, Y. Leveraging Deep Convolutional Neural Networks Pre-Trained on Autonomous Driving Data for Vehicle Detection from Roadside LiDAR Data. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22367–22377. [CrossRef]
13. Duan, Z.; Yang, Y.; Zhang, K.; Ni, Y.; Bajgain, S. Improved deep hybrid networks for urban traffic flow prediction using trajectory data. *IEEE Access* **2018**, *6*, 31820–31827. [CrossRef]
14. Janai, J.; Güney, F.; Behl, A.; Geiger, A. Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Found. Trends® Comput. Graph. Vis.* **2020**, *12*, 1–308. [CrossRef]
15. Tran, D.N.-N.; Pham, L.H.; Nguyen, H.-H.; Jeon, J.W. A Vision-Based method for real-time traffic flow estimation on edge devices. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8038–8052. [CrossRef]
16. McCraith, R.; Neumann, L.; Vedaldi, A. Real Time Monocular Vehicle Velocity Estimation using Synthetic Data. In Proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan, 11–17 July 2021; pp. 1406–1412.

17. Kang, H.; Lee, J. A Vision-based Forward Driving Vehicle Velocity Estimation Algorithm for Autonomous Vehicles. In Proceedings of the 2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Delft, The Netherlands, 12–16 July 2021; pp. 492–497.
18. Timofejevs, J.; Potapovs, A.; Gorobetz, M. Algorithms for Computer Vision Based Vehicle Speed Estimation Sensor. In Proceedings of the 2022 IEEE 63th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, Latvia, 10–12 October 2022; pp. 1–6. [CrossRef]
19. Hernández Martínez, A.; Fernandez Llorca, D.; García Daza, I. Towards view-invariant vehicle speed detection from driving simulator images. *arXiv* **2022**, arXiv:2206.00343.
20. Peruničić, A.; Djukanović, S.; Cvijetić, A. Vision-based Vehicle Speed Estimation Using the YOLO Detector and RNN. In Proceedings of the 2023 27th International Conference on Information Technology (IT), Zabljak, Montenegro, 15–18 February 2023. [CrossRef]
21. Chen, Y.; Zhao, D.; Er, M.J.; Zhuang, Y.; Hu, H. A novel vehicle tracking and speed estimation with varying UAV altitude and video resolution. *Int. J. Remote Sens.* **2021**, *42*, 4441–4466. [CrossRef]
22. Lowe, G. Sift-the scale invariant feature transform. *Int. J.* **2004**, *2*, 2.
23. Karami, E.; Prasad, S.; Shehata, M. Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images. *arXiv* **2017**, arXiv:1710.02726.
24. Kaplan, A.; Avraham, T.; Lindenbaum, M. Interpreting the ratio criterion for matching SIFT descriptors. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part V; Springer: Cham, Switzerland, 2016.
25. Long, T.; Jiao, W.; He, G.; Zhang, Z. A fast and reliable matching method for automated georeferencing of remotely-sensed imagery. *Remote Sens.* **2016**, *8*, 56. [CrossRef]
26. Boudjit, K.; Ramzan, N. Human detection based on deep learning YOLO-v2 for real-time UAV applications. *J. Exp. Theor. Artif. Intell.* **2022**, *34*, 527–544. [CrossRef]
27. Sundaresan Geetha, A.; Alif, M.A.R.; Hussain, M.; Allen, P. Comparative Analysis of YOLOv8 and YOLOv10 in Vehicle Detection: Performance Metrics and Model Efficacy. *Vehicles* **2024**, *6*, 1364–1382. [CrossRef]
28. Kalake, L.; Wan, W.; Hou, L. Analysis Based on Recent Deep Learning Approaches Applied in Real-Time Multi-Object Tracking: A Review. *IEEE Access* **2021**, *9*, 32650–32671. [CrossRef]
29. Yang, Y.; Pi, D.; Wang, L.; Bao, M.; Ge, J.; Yuan, T.; Yu, H.; Zhou, Q. Based on improved YOLOv8 and Bot SORT surveillance video traffic statistics. *J. Supercomput.* **2024**. [CrossRef]
30. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **2023**, *11*, 677. [CrossRef]
31. Alif, M.A.R.; Hussain, M. YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain. *arXiv* **2024**, arXiv:2406.10139.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.