Yagnik M Bhavsar[1], Mazad S Zaveri[2], Mehul S Raval[2], and Shaheriar B Zaveri[1]

[1]Affiliation not available
[2]Ahmedabad University

November 19, 2024

U-UTM: A Cyber-Physical System for Road Traffic Monitoring using UAVs

Yagnik M Bhavsar[1], Mazad S Zaveri[2], Mehul S Raval[2], Shaheriar B Zaveri[3]

*Abstract***— In low- or middle-income countries, road traffic safety becomes challenging because of the increasing number of vehicles, limited road infrastructure, and inability to capture a variety of traffic violations. Traffic Management and Information Control Centres (TMICCs) collect and analyse traffic data to improve road traffic safety. State-of-the-art technologies, such as the Internet of drones and computer vision, could help to automatically extract traffic data and provide a hands-free solution for road traffic monitoring. Hence, we propose a "UAV-based Urban Traffic Monitoring (U-UTM)" system to automatically extract traffic violations based on the movement of vehicles and traffic flow parameters, which helps to improve road traffic safety. Along with proposing a framework, this paper implemented three U-UTM operations - estimating Ground Sample Distance (GSD), detecting lanes using vehicles' tracks, and post-processing to reduce broken tracks. This paper derived GSD mapping from empirical data, improved it by 25.09% over theoretical GSD mapping, accurately detected lanes/zones, and reduced broken tracks of vehicles by 48.19%. Videos, data, and codes are available at https://github.com/ERYAGNIK003/U-UTM.**

I. INTRODUCTION

Traffic Management and Information Control Centres (TMICCs) [1] or Transportation Management Centres (TMCs) are established by the government to support several activities such as a collection of real-time traffic data (speed, traffic flow/volume, traffic violations) and analysis of these data to accomplish the objectives (*obj*) [2] such as traffic management (*Obj1*) and road traffic safety (*Obj2-6*), using state-of-the-art technologies. As shown in Fig. 1, the collected data are valuable to stakeholders for making transportation and urban policies, initiating traffic law enforcement drives, identifying possible causes of road accidents, and inferring problems or limitations of road infrastructure. In Fig. 1, though stakeholders are shown for India, the other countries in the Indian subcontinent have similar beneficiaries.

"Lane indiscipline" and "Driving against the authorised flow of the traffic" related traffic violations are the significant causes of road accidents [3] in India. Moreover, such violations are based on the movement of the vehicles and are specific to the road infrastructure (type of road junction, road markings). Further, these kinds of violations help to study "Human error" (traffic violations) correlated to "road condition", which contributes 41% in road accidents [2]; hence, such traffic data help to improve road traffic safety.

Fixed-wing Unmanned Aerial Vehicles (UAVs) give a stable view of the scene and require less maintenance than rotary-winged ones, and a group of such UAVs, also known as the Internet of Drones (IoD) [2], [4] help monitor urban city road traffic. UAV-based imagery (bird's eye view (90° downwards)) provides a complete (entry to exit) view of road junction, a non-occluded view of the traffic scene, and information about road infrastructure, which is not possible with other modalities such as GPS and CCTVs [2]. Recent advancements in Computer Vision (CV) [5], [6] allow accurate detection and tracking of vehicles in aerial traffic videos, enabling the extraction of road traffic data (count of automobiles, traffic volume, flow, speed of vehicles, traffic rule violations).

A UAV-based Cyber-Physical System (CPS) for traffic management [7] was developed to demonstrate real-time traffic measurement. Authors [7] established video streaming from UAV to computer and used theoretical Ground Sample Distance (GSD) mapping and UAV camera imaging for traffic estimation. Still, the system [7] lacks traffic violation detection and analysis methods. Also, they focused on traffic modelling rather than road traffic safety.

Given the above discussion, this paper presents a UAV-based cyber-physical system named "UAV-based Urban Traffic Monitoring" (U-UTM) for automatic road traffic monitoring. The U-UTM system focuses only on traffic violations based on the movement of the vehicles and hence employs methodology described in [2] for violation detection and probing traffic data. Moreover, in aerial imagery, an accurate pixel-to-actual ground distance mapping known as GSD is required to estimate the actual ground distance for traffic data measurement, such as vehicle speed. Therefore, this paper implements methods for accurate GSD mapping, automatic zone detection, and reducing broken vehicles' tracks. It also verifies the effect of traffic density (vehicles per video frame) on computation time. The rest of the paper is organised as follows: The architecture and methods of the U-UTM system are described in Section II. Results are discussed in Section III. Section IV concludes the paper.

[1]Yagnik M Bhavsar is a doctoral candidate at the School of Engineering and Applied Science, Ahmedabad University, 380009, Gujarat, India yagnik.b@ahduni.edu.in

[2]Mazad S Zaveri and Mehul S Raval are faculties at the School of Engineering and Applied Science, Ahmedabad University, 380009, Gujarat, India mazad.zaveri@ahduni.edu.in, mehul.raval@ahduni.edu.in

[3]Shaheriar B Zaveri is a road safety consultant at Road Safety Automotive Management, Gujarat, India roadsafetyam@gmail.com

979-8-3503-8727-8/24/$31.00 ©2024 IEEE

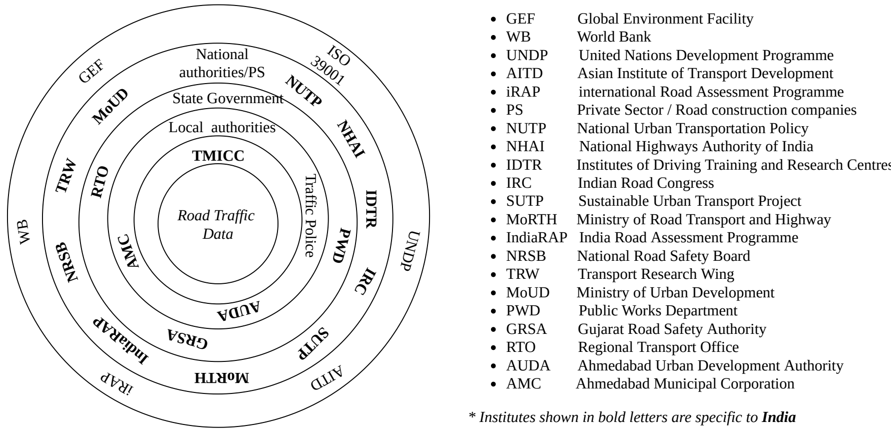


Fig. 1: Stakeholders of road traffic data in India. (Similar bodies under different names exist in other countries.)

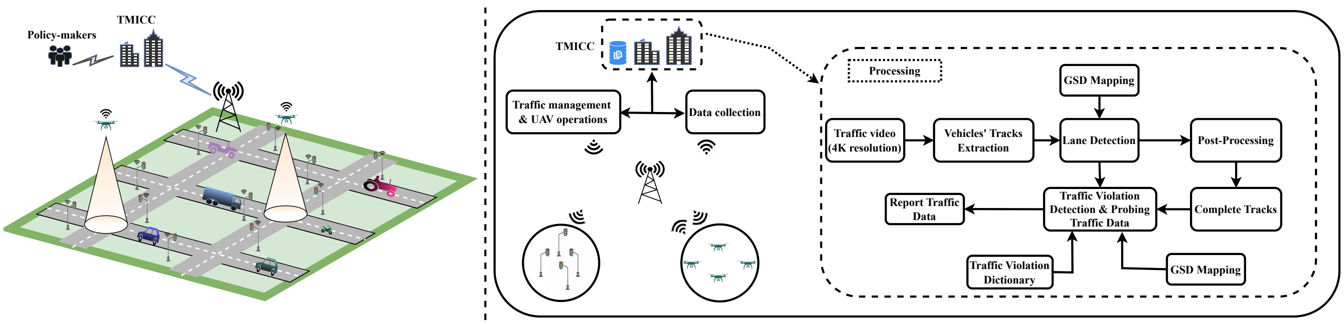


Fig. 2: Overview and architecture of U-UTM system. Note various methods in the architecture.

II. Proposed system

U-UTM enables automatic extraction of traffic insights from UAV traffic videos, which helps improve road traffic safety. In Fig. 2, the outline and architecture of the U-UTM are given. Each method is discussed in detail in the following sections.

A. Road Traffic Videos Collection

Fixed-wing UAVs (gimbal 90° downwards) record traffic videos at a 4K resolution (3840×2160 pixels). UAVs fly at a particular flight height to cover all entry-exit points/lanes of a road junction. Along with videos, UAVs also store subtitle files (SRT), which contain valuable information on UAV location (GPS parameters) and camera parameters (aperture, ISO). This information helps automate UAV operations.

B. Vehicle Detection and Tracking

U-UTM focuses on traffic violations based on the movement of the vehicles. Hence, vehicles' tracks are required to detect traffic violations. YOLO [5] models are widely used in vehicle detection [2] because of their various features, such as auto-learning anchor boxes, concatenating (object) features from different convolution layers, support for larger input image size, and real-time inference. Such vehicle detection models require a dataset for training, and Visdrone [8] is one of the publicly available UAV-based vehicle datasets. It contains "small objects" [2], allowing the vehicle detector to detect two-wheelers and pedestrians efficiently. A total of 60% samples have resolutions smaller than 50×50 pixels [2]. After detection, a Multi-Object Tracking algorithm (MOT) is required to track the vehicles. A BoT-SORT [6] is a widely used MOT based on a tracking-by-detection approach, which uses a Simple Online Real-Time (SORT) tracking algorithm and SBS-S50 deep neural network to extract the appearance features of the object. Introducing an appearance extractor improves tracking accuracy and reduces broken tracks against the SORT, as discussed in [2]. Considering the algorithmic efficiency, this paper uses "YOLOv8x" [5] as a vehicle detection model and BoT-SORT for tracking.

C. Traffic Violations Detection and Probing Traffic Data

Authors in [2] have discussed in detail "How to detect traffic violations based on the movement of vehicles using UAV videos." They used YOLOv7 and SORT for vehicle detection and tracking, respectively, to extract vehicles' tracks. Then, they divided the traffic scene into several zones (entry-exit points or lanes for road junction) (refer to Fig. 4) and re-defined vehicle track in terms of zone traversal sequence. Further, they formulated a "Violation Dictionary"-mapping between zone traversal sequence (vehicle movement) and traffic violations (sections of the Government of

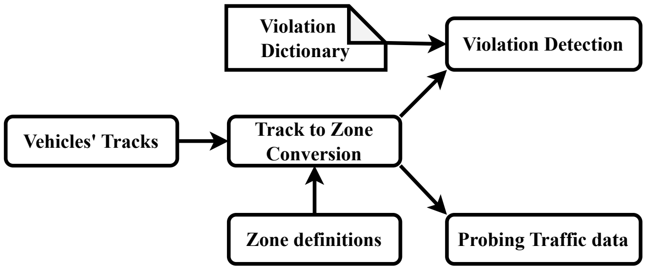


Fig. 3: Modules of “traffic violation detection and probing data” system [2].

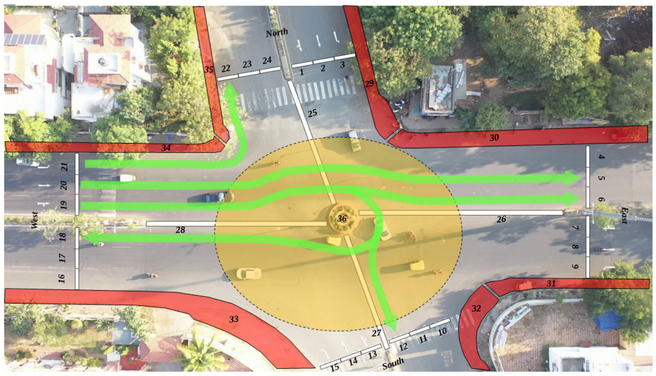


Fig. 4: Road scene with defined zones at every entry-exit point and near a multi-lane urban roundabout [2]. (Best-viewed as colour image)

India rules/regulations/acts (MVDR/MVA) [2]). Each vehicle's track (in terms of zone traversal sequence) is matched against the violation dictionary to detect traffic violations (refer to Fig. 3). Also, authors [2] show a method to probe traffic flow parameters such as the number of vehicles, speed of vehicles, lane-wise traffic distribution, and traffic flow rate. These data are helpful in traffic flow and congestion analysis. They used theoretical GSD_T based on UAV camera parameters and flight height, and it is reproduced as Eq. 1. These parameters can be easily found in the SRT file.

$$GSD_T = 0.03337 \times \textit{Flight height} \tag{1}$$

Even though the work presented in [2] can extract traffic data, it can not be used for automatic traffic monitoring as the zones (lanes/entry-exit points) were manually defined, and the problem of broken tracks was not thoroughly addressed. Also, GSD_T used by them is error prone because of lens distortion [9]; therefore, the proposed paper addresses the above-mentioned issues.

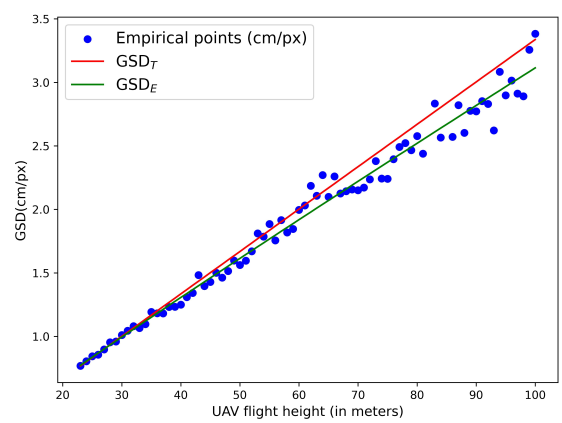
D. Empirical GSD mapping

Accurate estimation of GSD is vital in road traffic analysis, such as extracting traffic flow parameters (speed of vehicles) and measuring road attributes [10] (road/lane width, size of roundabout). This paper derives a rotation-invariant estimation of GSD mapping from empirical data, an empirical *GSD* (GSD_E).

To do so, UAV video of railway tracks was collected for an experiment at a flight height of 23 to 100 meters. We chose railway tracks because they are perfect parallel lines, and the UAV was aligned diagonally with railway tracks to generalise

(a) Lines detected using line segment detector.

(b) Longest lines (in black colour) and best pair (in green colour).



(c) GSD mapping estimation.

Fig. 5: Derivation of GSD_E using railway tracks' UAV video.

mapping in both dimensions. The distance (centre-to-centre) between two parallel rails measured 174 centimetres, and the width of the rail was 6.35 centimetres.

Straight lines in an image are detected using line segment detector [11], as shown in Fig. 5a. Then, we select only the longest lines to avoid detection of false rails and calculate their slopes and minimum distance between each pair of lines, refer to Fig. 5b. Next, we find parallel line pairs with slope-difference less than a threshold value, and the best pair of lines is selected based on the distance nearest to the calculated distance using Eq. 1, as shown in Fig. 5b. Ultimately, we use distances (in pixels and centimetres) between the lines in the best pair to calculate *GSD* value for each frame in a video. We chose the centre-to-centre distance between two parallel rails; the maximum measurement error observed is $\pm$6.35 centimetres. Using curve fitting on empirical points (shown in Fig. 5c), we found power function [12] (refer to Eq. 2) as best fit with R^2 value of 0.9898.

$$GSD_{\mathrm{E}} = 0.0396 \times \textit{Flight height}^{0.9478} \tag{2}$$

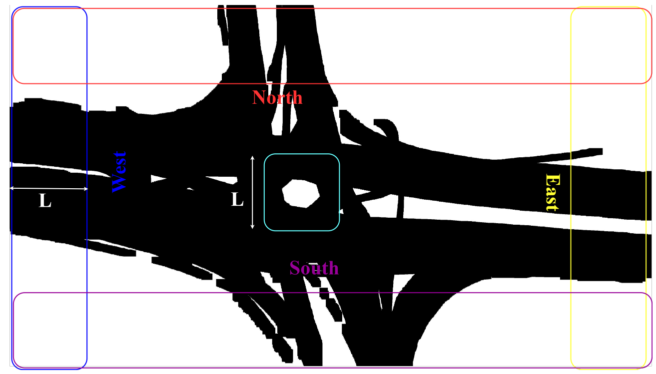
E. Lanes/Zones Detection

Segmentation-based deep learning models are used for the road (Lanes/Zones) detection in aerial imagery but face problems of unavailability of the dataset, shadows on the road, and weather conditions [13]. This paper presents lanes/zone detection using image processing and vehicle tracks for such cases. This paper assumes "road-dividers" are present at the road junctions. Our method includes several steps to detect lanes/zones as follows:

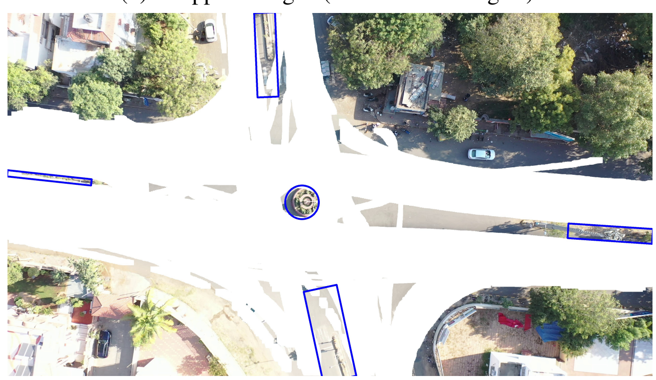
- Impose moving vehicles' tracks on a black background (the same size as the original image), which results in a grey-scale image (white and black pixels represent vehicles' tracks and backgrounds, respectively). Then, blur and erode (a morphological image processing operation) an image to remove noise and thin edges; refer to Fig. 6a.
- Invert the resultant image so the unused road scene (road dividers, roundabout, parking area, and shops/buildings) turns to white pixels. Crop the image into five parts (for each direction and around a centre) to a certain length (*L* meters); refer to Fig. 6b.
- Each cropped image is then processed for connected component analysis to find road dividers and roundabouts based on the area of contours. The minimum area contour is identified as a divider, whereas the maximum area contour is identified as a roundabout; contours are shown in Fig. 6c.
- After detecting the dividers and the centre of the roundabout, lanes/zones are derived using the *number of lanes*, *width of the road*, GSD_E, and appearance (size and orientation) of the dividers. Ultimately, numbering for entry-exit zones (1-24) is defined based on directions and proximity to the dividers. Similarly, zones (25-28) around a roundabout are defined based on directions and centroids of dividers and roundabouts. Zones (29-35) for parking violations are defined based on the polygon formed by leftmost zone lines and image boundaries. Zone (36) for over-speeding violation is defined as a circular region (with a specified *radius*) around the roundabout's centre. All zones with predefined numbering schemes [2] are shown in Fig. 6d.

This paper used the OpenCV library for image processing-related tasks. Values of various parameters used in this method are as follows: *L*=2 meters, *radius* (of zone 36)= 2 meters, *number of lanes*=3 and *road width*=10.5 meters (as it is multi-lane roundabout [2]).

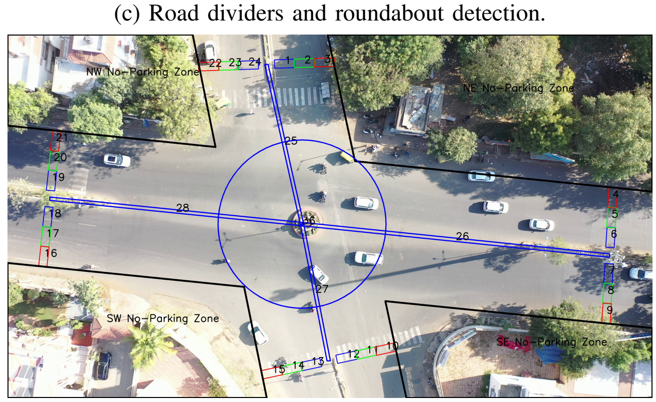
(a) Vehicles' tracks on black background.



(b) Cropped images (shown as rectangles).



(c) Road dividers and roundabout detection.



(d) Road scene with detected zones.

Fig. 6: Lanes/Zones detection using vehicles' tracks and connected component analysis.

F. Post-processing of Tracks

We have used the state-of-the-art tracking model to extract vehicles' tracks, but in several scenarios, such as occlusion (while the vehicle passes under the tree, refer to the north-west corner in Fig. 4) and vehicle turning, it can fail to track it and this lead to multiple broken tracks of that vehicle. A complete track is necessary to extract road traffic data. Hence, we employ a post-processing module to stitch broken tracks of the same vehicle and generate a complete track.

The steps of the post-processing of tracks are as follows: First, we segregate tracks (from YOLO) into complete and broken tracks, considering their starting and ending positions. If the track starts/ends before/after any of the entry/exit zones, it is regarded as a complete track; otherwise, it is a broken track. Then, we recursively iterate over each broken track and stitch them based on their spatiotemporal proximity and vehicle class. If two broken tracks (of the same class) have a temporal difference less than a *timeThresh* and a spatial difference (euclidean distance) less than a *distThresh*, we stitch them. The *timeThresh* is a hyper-parameter. The *distThresh* is calculated using the vehicle's average speed and the temporal difference.

G. UAV & Traffic Management Operations

Internet of Drones (IoD) allows TMICCs to automatically initiate UAV operations and control the traffic, as shown in Fig. 2.

UAV operations: Deploy the UAVs at the required height to capture a complete road junction; Record traffic video along with auxiliary data such as camera parameters, battery power, UAV serial number; Share the data with TMICCs; Return when the battery is drained.

Traffic management operations: Collect traffic videos and other data sent by UAVs; Detect traffic violations and probe traffic data; Store traffic data in the local database; Control the traffic signal timing to manage the traffic based on estimated traffic flow distribution.

Wireless interfaces: To accomplish the operations mentioned above, real-time wireless protocols such as Real-Time Messaging Protocol (RTMP) and Message Queuing Telemetry Transport (MQTT) are required [7], [14]. RTMP can stream traffic videos from UAVs to the TMICCs using 4G, as shown in [7]. On the other hand, TMICCs need to control UAVs and traffic signals so that MQTT can help transfer small chunks of data.

H. Computation time

A complete processing time for road traffic data extraction from the UAV video can calculated using Eq. 3. We have considered only computationally intensive processes. We use the tracking-by-detection approach, so the total time (denoted as *Veh_Det_Trk* in Eq. 3) required to extract tracks from a video depends on the number of frames. The post-processing method stitches broken tracks by iterating recursively over all of them, and at the end, each complete track will proceed with the road traffic data extraction, denoted as *Post-Processing* and *Vio_Det_Probe_Data*, respectively in Eq. 3.

$$
\begin{aligned}
T = \sum_{frames} \textit{Veh\_Det\_Trk} + \prod_{tracks_{broken}} \textit{Post-Processing} \\
+ \sum_{tracks_{complete}} \textit{Vio\_Det\_Probe\_Data}
\end{aligned} \tag{3}
$$

III. Results and Discussion

We collected road traffic videos of various junctions in Ahmedabad, India, using DJI Mavic 2 Pro fixed-wing UAV. We used NVIDIA Quadro RTX 6000/8000 GPU for computing. YOLO model is trained for 300 epochs and achieved 60% mAP@0.5.

We tested GSD mapping on UAV videos (recorded at various heights and with full rotation) of railway tracks. We computed RMSE error for both GSD_T and GSD_E, and achieved 25.09% improvement over GSD_T; refer to Table I.

TABLE I: Comparison of GSD_T and GSD_E using RMSE.

Video No.	Height(meters)	Full rotation	$RMSE_T$	$RMSE_E$
Video 1	23-100	No	21.01	15.86
Video 2	85	Yes	24.05	17.62
Video 3	63	Yes	19.25	14.62

The accuracy of the Lane/Zone detection algorithm depends on the road/lane/zone usage. As shown in Fig. 6d, the "North" and "South" direction zones are slightly deviated from their ideal location compared to "West" and "East" because of less road usage [2]; refer to Fig. 6c. Values of the various parameters need be set according to the type of road junction (*number of lanes* and *road width*), placement of zones and UAV flight height (*L* and *radius*).

For road traffic data extraction (including violation detection), we only use complete tracks because broken tracks fail to provide exact traffic data, such as the exit lane of the track was missing before post-processing, refer to Fig. 7. We tested and visually verified the post-processing of tracks (for *timeThresh*=5 seconds) on different videos. We extracted 533 vehicles' tracks, out of which 450 were complete and 83 were broken tracks. After post-processing, we reduced 48.19% broken tracks. Value of *timeThresh* needs to be set according to the performance of the tracking algorithm.

In Eq. 3, *Veh_Det_Trk* time is the dominant component over others as it depends on the number of frames. We used traffic videos of different traffic densities to calculate the *Veh_Det_Trk* time. In Table II, calculated *Veh_Det_Trk* time for our detection and tracking model with average vehicles per frame are shown. *Veh_Det_Trk* time helps to estimate the total time *T* and allows to schedule such operations at TMICCs. Table II shows that *T* does not depend on traffic density when computing is performed on GPU.

Collected road traffic data over time have various applications, such as suggesting policy-related changes as discussed in [2] and controlling traffic signals to reduce congestion. Lane-wise traffic distribution extracted from UAV videos [2] (as shown in Fig. 8) allows the control of traffic signalling based on traffic movement across the road junction [15] as in this case, green light for "West-to-East" and "East-to-West"

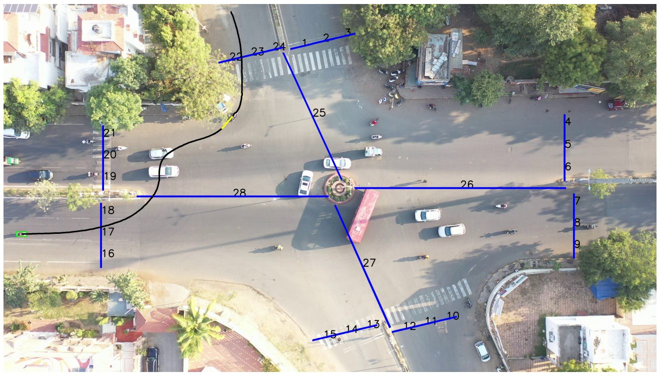


Fig. 7: Example of a post-processed complete track (Broken tracks are shown in black colour and interpolated track shown in yellow colour).

TABLE II: Traffic density and computing time.

Video No.	Traffic density	$T \approx Veh\_Det\_Trk$ time (in ms)
Video 4	21	55.02
Video 5	35	54.97
Video 6	71	54.41
Video 7	99	54.19
Video 8	145	54.26

directions should be on for approximately 3 to 5 times more than the other directions.

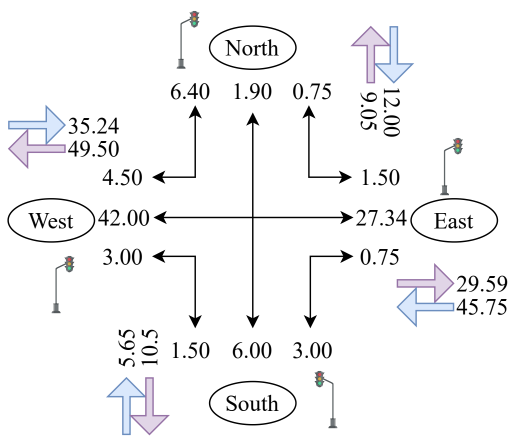


Fig. 8: Lane-wise traffic distribution (in %) [2].

The U-UTM system primarily focuses only on traffic violations based on the movement of the vehicles (methodology given in [2]), whereas existing CPS [7] focuses on traffic modelling. Table 7 [2] provides a further comparison of traffic violation detection methodology.

IV. Conclusion and Future Work

This paper presents a UAV-based cyber-physical traffic monitoring system, U-UTM, to support TMICC operations by automatically extracting traffic data (traffic violations based on the movement of vehicles and traffic flow parameters). We derived a GSD_E mapping from empirical data and achieved an improvement of 25.09% over GSD_T. We employed an image processing-based method to automatically detect lanes/zones of a road junction using vehicles' tracks. Also, we introduced post-processing on tracks to get complete tracks from broken tracks and achieved a 48.19% reduction in broken tracks, thereby increasing the accuracy of road traffic data extraction. Collected traffic data could be helpful for the government (policy-makers and other stakeholders) in improving road traffic safety. Various hyper-parameters related to our methodology need to be set according to the performance of the tracking algorithm, type of road junction, and UAV flight height. This paper can extract road traffic data, but it is possible to extract road condition data with the help of computer vision, which will be very useful to TMICCs.

Acknowledgement

We thank the Traffic Police, Ahmedabad City (Gujarat, India), for drone permission (app no. G/725/Traffic/3186/2021). This work was funded by seed grant URBSEASI21A3, Ahmedabad University and the grant GUJCOST/STI/2021-22/3858 by Govt. of Gujarat.

References

[1] Ministry of Urban Development, "Traffic management and information control centre (tmicc)," Govt. of India, New Delhi, Tech. Rep., Nov. 2016. [Online]. Available: https://smartnet.niua.org/content/ddcc994d-1091-48d7-b614-a7684cc6499f

[2] Y. M. Bhavsar, M. S. Zaveri, M. S. Raval, and S. B. Zaveri, "Vision-based investigation of road traffic and violations at urban roundabout in india using uav video: A case study," *Transportation Engineering*, vol. 14, 12 2023.

[3] Ministry of Road Transport and Highways, "Road accidents in india 2022," Govt. of India, New Delhi, Tech. Rep., Oct. 2023. [Online]. Available: https://morth.nic.in/road-accident-in-india

[4] F. Outay, H. A. Mengash, and M. Adnan, "Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges," *Transportation Research Part A: Policy and Practice*, vol. 141, pp. 116–129, Nov. 2020. [Online]. Available: https://doi.org/10.1016/j.tra.2020.09.018

[5] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," 1 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[6] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," *arXiv preprint arXiv:2206.14651*, 2022.

[7] X. Ma, X. Liang, M. Ning, and A. Radu, "Metric: Toward a drone-based cyber-physical traffic management system," in *IEEE Intl. Conf. on Systems, Man, and Cybernetics*, 2022, pp. 3324–3329.

[8] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[9] F. Liebold, D. Mader, H. Sardemann, A. Eltner, and H.-G. Maas, "A bi-radial model for lens distortion correction of low-cost uav cameras," *Remote Sensing*, vol. 15, no. 22, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/22/5283

[10] iRAP, "About International Road Assessment Program." Accessed June. 20, 2024. [Online]. Available: https://irap.org/

[11] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: a Line Segment Detector," *Image Processing On Line*, vol. 2, pp. 35–55, 2012, https://doi.org/10.5201/ipol.2012.gjmr-lsd.

[12] D. Zhao, Y. Chen, and S. Yu, "Tracking and speed estimation of ground vehicles using aerial-view videos," in *5th Intl. Conf. on Automation, Control and Robotics Engineering*, 2020, pp. 597–601.

[13] J. Pruthi and S. Dhingra, "A review of research on road feature extraction through remote sensing images based on deep learning algorithms," in *3rd Intl. Conf. on Innovative Sustainable Computational Technologies*, 2023, pp. 1–5.

[14] J. O. Netto, R. C. Brito, F. Favarim, L. F. Priester, and E. Todt, "Implementing a communication network between bases station applied for group of drones," in *IEEE 29th Annual Software Technology Conf.*, 2022, pp. 135–140.

[15] B. Pratama, J. Christanto, M. T. Hadyantama, and A. Muis, "Adaptive traffic lights through traffic density calculation on road pattern," in *Intl. Conf. on Applied Science and Technology*, 2018, pp. 82–86.