TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–15

DOI: 10.1177/0361198120920627
journals.sagepub.com/home/trr
SAGE

Lane Detection and Lane-Changing Identification with High-Resolution Data from a Swarm of Drones

Emmanouil Barmpounakis[1], Guillaume M. Sauvin[1], and Nikolaos Geroliminis[1]

Abstract

In the era of big data, new transportation-related concepts and methodologies need to be proposed to understand how congestion propagates. *pNEUMA*, a unique dataset that was acquired during a first-of-its-kind experiment using a swarm of drones over a dense city center, has uncovered new opportunities for revisiting and evaluating existing concepts, and new ways to describe significant traffic-related phenomena. This dataset is part of an open science initiative shared with the research community and consists of more than half a million detailed trajectories of almost every vehicle that was present in the study area. The aim of this paper is to describe the first methodological approach to how such information can be utilized to extract lane-specific information from this new kind of data and set the benchmark for possible future approaches. Specifically, we describe the methodological framework of two related algorithms: lane detection and lane-changing maneuver identification. Azimuth was the main concept utilized in this methodological framework to overcome existing issues in the literature related to identifying lane-changing maneuvers. The combination of high-quality data, clustering techniques, and detailed spatial information in the lane-detection algorithm indicated it was an effective tool without the need for complex computational effort. Moreover, high-resolution data together with modern time-series analysis tools for lane-changing identification, showed that high-accuracy predictive algorithms can be obtained. The accuracy of both tools was over 95%. Challenging scenarios are identified for future studies and to further improve the tools.

For several years, the traffic community has been craving data to deal with traffic-related phenomena more effectively. Recent advances in data acquisition and management tools have opened up new ways for monitoring, studying, and modeling congestion propagation. In this emerging era of big data, new concepts, tools, and methodologies are expected to be developed to better understand congestion and provide fresh targeted solutions. Considering phenomena in arterials and dense city centers, is particularly challenging, as several issues may arise as a result of the limitations in current data collection methods (inadequate traffic penetration rates, privacy issues, GPS errors, etc.) (*1*).

In this context, a unique experiment was recently conducted utilizing a swarm of 10 drones over the congested central district of Athens, Greece (*2*). The aim of the experiment was to record traffic streams in the multimodal congested environment of an urban setting, using unmanned aerial systems (UAS—more commonly known as drones) that could facilitate in-depth investigation of critical traffic phenomena. One of the aims of the work was to reveal the fundamental mechanism of congestion pattern formation for large-scale networks based on the complete dataset generated by the drones. A highly detailed dataset containing more than half a million trajectories with data points every 0.04 s was created (see Data Collection section). This unique dataset—nicknamed *pNEUMA*—that was acquired with the analysis of the videos collected, includes detailed trajectory data of more than half a million vehicles, including cars, taxis, buses, motorcycles, and heavy and medium vehicles. Thus, the *pNEUMA* dataset offers new opportunities to study specialized phenomena related to multimodal interactions that take place mostly in urban environments. Primary analyses showed that significant insights could be provided on how drones' unique characteristics can overcome existing limitations in traffic monitoring and about the drones' potential for becoming a viable part of the Intelligent Transportation Systems

[1]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Corresponding Author:
Emmanouil Barmpounakis, manos.barmpounakis@epfl.ch

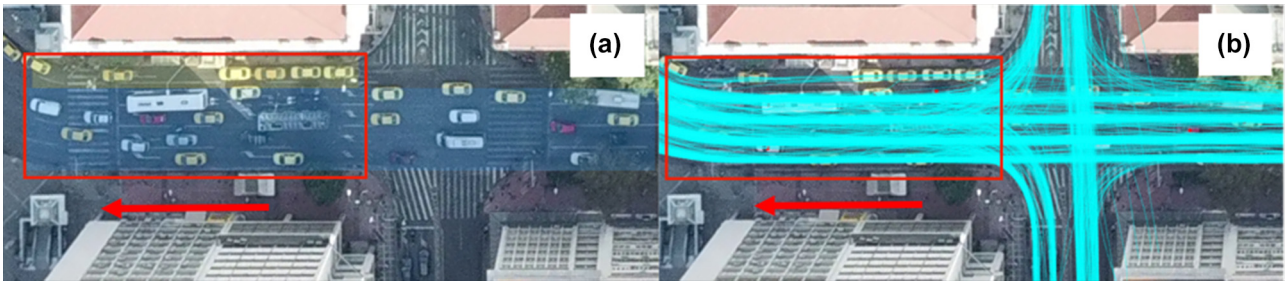


Figure 1. An example of the differentiation in existent lanes at a road section: (*a*) a taxi stop highlighted in yellow and (*b*) lanes that are actually used.

(ITS) infrastructure. The *pNEUMA* dataset includes a significant amount of information that is rarely available for arterial networks (e.g., detailed interactions for almost every pair of vehicles moving in a congested urban environment). Clearly, this will allow the revisiting of existing models, and the creation of a new era of microscopic traffic flow models, which will improve the accuracy, calibration, and validation of such microsimulation tools. More information on the dataset and the unique possibilities open to researchers can be found in research by Barmpounakis and Geroliminis (*2*).

Recording traffic streams with drones was found to be a promising monitoring technique as it does not require installment of sensors and so forth. Although vehicle-tracking algorithms can accurately extract the position of vehicles with very high frequency, these data alone are not sufficient for microscopic traffic modeling as they are missing vital lane-specific information, especially for large urban areas. Our motivation was to develop automatic algorithms that could identify different lanes and lane changes. As the *pNEUMA* dataset was openly shared with the research community, we aimed to enhance it with additional information that could prove valuable to researchers. One of the first subjects that we believed was worthy of further analysis was how this dataset could be utilized to extract useful lane-specific information. Until now, most studies on lane detection have largely focused on environments such as highways or rural streets, where lanes are primarily used to enter/exit a highway, or for advanced driver-assistance system or autonomous vehicle (AV) applications, like lane departure warnings or road marking identification for safety reasons (*3–5*). In such cases, the input is typically images that come from cameras attached to vehicles, which can therefore be quite noisy when it comes identifying a lane's width and lane edges, especially in different road environments. Therefore, researchers have turned to advanced image processing and deep learning methods to solve the numerous issues that have emerged (*6*). When it comes to cases that are more similar to the *pNEUMA* dataset, like the NGSIM freeway database (*7*) or the highD dataset (*8*), although the lane of each vehicle is provided, the methodology of extracting such information cannot be generalized as the study area was predefined with specific road characteristics. Other researchers have used GPS trajectories to extract lane-specific information but in most of them data from urban environments are not included. For example, Knoop et al. used GPS data from a vehicle that repeated the same route 100 times (*9*), whereas Tang et al. employed low-precision GPS trajectories from taxis during off-peak hours to reduce the noise (*10*). In research by Chen and Krumm, the Gaussian mixture model for identifying lane information is introduced, and limitations related to the dynamic aspects of a road network (driving direction, traffic controls, turn restrictions, etc.) are highlighted (*11*). It was therefore evident that the existing low quality, noisy GPS data could require special techniques and methodological concepts compared with the level of detail that the *pNEUMA* dataset offers.

Knowledge of the geometric design of urban roads (for example from a map or an aerial photo) is not sufficient to identify exactly how and where vehicles are moving. For example, lanes might have different usages according to the time of day, such as dedicated bus lanes, on-street parking, and contra-flow lanes. Thus, when it comes to urban environments, the way the road network is formed and the manner in which lanes are used (turning lanes, bus lanes, etc.) require a different approach when collecting and processing data for advanced modeling techniques. One significant requirement for advanced traffic monitoring and control is to provide detailed lane-specific information, which can be a time-consuming and challenging task. A typical example from *pNEUMA* can be seen in Figure 1, where at a specific road section there are five marked lanes, but not all of them can be used to accommodate traffic flow demand. Figure 1*a* shows that the rightmost lane is used as a taxi stop (yellow area),

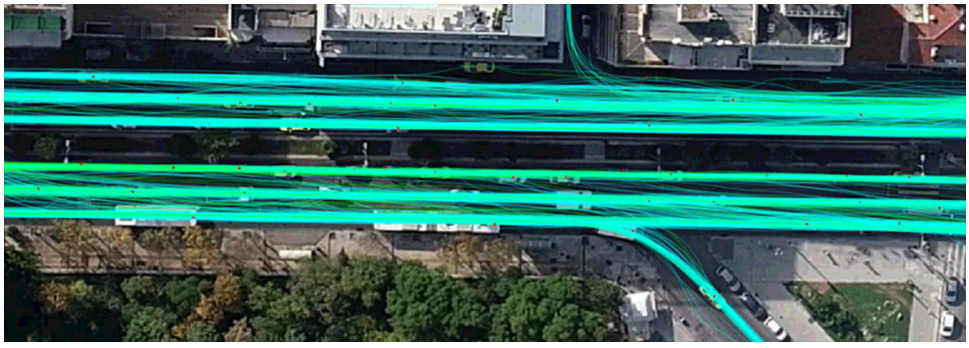


Figure 2. Lane changes are illustrated by the thin lines between the dense lines.

and only four of them are for traffic. This is also illustrated in Figure 1*b* in which all vehicle trajectories from 15 min of data are plotted.

The way lane-changing maneuvers are conducted represents another differentiation between road environments. There are few such studies in urban environments, mainly because of the lack of data. Nonetheless, lane-change modeling has attracted significant interest, especially for freeway systems (*12–16*). In principle, these maneuvers are studied from a safety perspective. When researchers focus on how they may affect traffic flow characteristics, simulation methods are mainly deployed (*17*, *18*). Lane-changing maneuvers can sometimes cause local disturbances, it is therefore crucial to understand how they affect the capacity and stability of the traffic flow (*12*). Gap-acceptance models have been used in studies of such behavior (*19*, *20*). One of the most general lane-changing models is described in work by Kesting et al. (*12*). The authors introduce the utility of a given lane and the risk of a lane change is determined in relation to longitudinal accelerations. This allows the formulation of both safety and incentive criteria for various passing rules. Although these phenomena are common in urban roads and freeways, they have mostly been examined in detail in freeways, notably after the development of the NGSIM freeway database (*7*). The main body of NGSIM freeway data contain a few hundred meters of road. While in NGSIM there are a few urban arterials (e.g., Peachtree Street), the site is small and with insufficient congestion or long queues (*21*). NGSIM data have also be found to contain significant noise, and acceleration estimations requiring smoothing when extreme values are observed (*1*, *22*, *23*).

Compared with a freeway, an urban arterial trip contains a larger number of lane changes that create a more circuitous route. For example, in Figure 2 the lane changes are visible as the trajectories (thin lines) between the different lanes (dense lines). Recent progress in video-tracking methods has allowed the collection of high-quality trajectory data from aerial observations (*24*). Recently, as seen in Barmpounakis et al. (*25*), such data have allowed the detailed study of various transportation-related phenomena and investigation of their effect on network congestion. However, realization of data-related opportunities requires going beyond the existing simulation-based approaches of modeling congestion with complex models comprising several parameters that make their validation questionable. We followed an empirical approach to understanding these mechanisms. With respect to lane-changing maneuvers for arterial streets, complete and accurate knowledge of the local environment (vehicles in the proximity) is crucial for understanding the physical properties of the interactions. Traditional sensing techniques (e.g., GPS data) are not sufficient for this purpose owing to a lack of accurate vehicle trajectories and decreased traffic penetration rates.

Cooperative systems may well benefit from the integration of predictive maneuvering models, for example, by reducing accident rates, increasing network efficiency, and improving fuel consumption (*26–28*). With the recent surge in AV-related research, several phenomena have been revisited by researchers, uncovering significant inconsistencies compared with the reality, or imposing serious limits for advanced modeling approaches (*1*, *29*, *30*). Both mandatory- and discretionary lane-changing (MLC and DLC respectively) maneuvers have been examined in various ways to identify multilane traffic dynamics (*31–33*) or to model driver decisions around whether to conduct an overtaking maneuver (*34*). Similar to lane-choice behavior, lane-changing maneuvers in urban arterials take place for various reasons, for example travel maneuver plans, current lane type, and surrounding traffic conditions (*35*), or to filter traffic when it comes to specific vehicles like powered two-wheelers (PTW) (*36*, *37*). Literature suggests that there is a strong

interaction between PTW and the surrounding vehicles, affecting how they move, which is an ongoing research topic characterized by complex traffic phenomena and trajectories (*38*).

The smartphone has also proved a useful new tool for microscopically studying driver behavior: studying driver aggressiveness and other such extreme behaviors (*39*). Given their significant advantages compared with traditional methods, lane-changing is easily identified from such source data, as is the exact circumstances the driver is experiencing in their micro-environment (*40*). However, a significant factor of driver aggressiveness that affects traffic flow, in regard to both operation and safety, relates to the quantification of lane-changing maneuvers in urban areas, as documented in previous studies (*41–43*).

Thus, little research based on real data has been conducted on how lane choice or lane-changing maneuvers affect traffic on a macroscopic scale. Challenging research questions have arisen along these lines. Can detailed lane-specific information be a significant feature to advance existing models and concepts? How negligible is the effect on traffic for different traffic conditions? What variables affect drivers' lane choice or their decision to make a lane-changing maneuver?

To answer such questions, we first need appropriate tools and methodologies to accurately detect such information from the *pNEUMA* dataset. The aim of this paper is to present a unique, complete methodological tool for extracting such information and to propose new concepts based on the new data that are available. This methodology could set a benchmark for possible future approaches from researchers wanting to approach similar topics or use lane-related information. For example, as this paper focuses on identifying whether both MLC and DLC occur, a distinction and comparative analysis of the two types could follow in future studies.

The rest of the paper is structured as follows: the next section includes a short description of the *pNEUMA* dataset. Two algorithms for lane detection and lane-changing maneuver identification are then described. Finally, we provide an evaluation of the algorithm compared with real cases from the dataset and we sum up with some conclusions and remarks.

Methodology

Data Collection

In October 2018, a swarm of 10 drones was utilized over significant parts of Athens, Greece; the city was selected because it is an urban, multimodal, busy environment that could allow different kinds of transportation phenomena to be tested (*2*). The aim of the experiment was to record traffic streams over an urban setting using drones, and to provide significant insight into how their unique features could overcome existing limitations in traffic monitoring, recording traffic streams, and to explore their potential for becoming a viable part of the ITS infrastructure. For the experiment, the morning peak (8:00 to 10:30 a.m.) was recorded for each working day of the week.

Figure 3. The study area of the *pNEUMA* experiment and the area of responsibility for each drone.

Specifically, the swarm would take-off from the two take-off/landing areas (H1 and H2 in Figure 3) at the start of the experiment and each drone would fly to its unique hovering point. Then, when all drones were in position, the recording of the traffic stream would start simultaneously; when the battery was running low, they would return to their landing point. Considering that the drones could hover up to 25 min including take-off, routing, and landing times, it was decided that each session would take place every 30 min for better coordination and standardization of the experiment. This set-up allowed for 15 to 20 min of continuous traffic monitoring; during temporal blind spots, trajectories were not recorded. The study area (1.3 km^2) analyzed included more than 100-km lanes of road network (low-, medium-, and high-volume arterials), around 100 busy intersections (signalized or not), more than 30 bus stops and close to half a million trajectories (Figure 3). It is evident that on such a scale, even for a simple traffic study, one would need more than 100 fixed sensors (or humans) to collect data, including all the measurement (or manual) errors. This aerial video footage allows the researchers involved to rewatch the videos as many times as they want, not only to eliminate errors but for other reasons and at different levels of detail, to fulfill the requirements

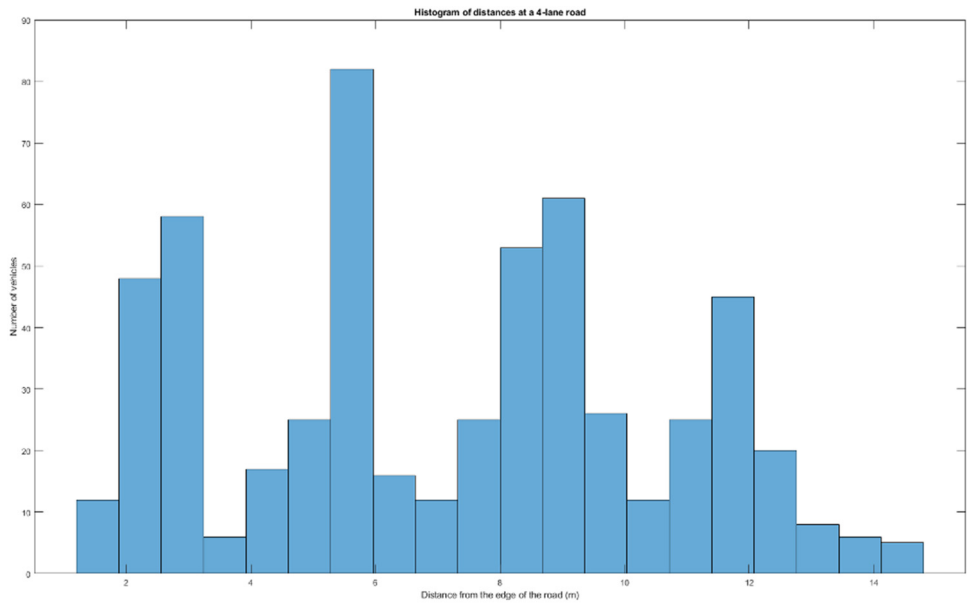


Figure 4. Histogram of the distances from the trajectories to the edge of the road.

of a variety of studies and subjects. It should be noted that the overlapped areas are of primary significance for the synchronization of the video footage in relation to space and time and the re-identification of vehicles moving from one area to the other.

A unique observatory for traffic congestion, comprising data that did not exist before at this resolution and scale has been created from processing the videos from the experiment. This massive dataset contains trajectories of every vehicle that was present in the study area, calibrated in the World Geodetic System (WGS 84), every 0.04 s, as this is the maximum frequency allowed by the video's frame rate. In addition to the features that can be produced using position information, for example speed (first derivative of position), acceleration (second derivative of position), and distance traveled, vehicle type (car, taxi, motorcycle, bus, heavy vehicle, medium vehicle) is also available. For more detail on the design of the experiment see research by Barmpounakis and Geroliminis (*2*); the data can be downloaded from https://open-traffic.epfl.ch

Lane Detection

Traditional sensing techniques (e.g., loop detector, GPS data) are not sufficiently accurate to identify the exact position of each vehicle on the road, so *pNEUMA* provides unique opportunities. Our approach was based on the idea that vehicles tend to move in the center of a lane. However, this is not the case for every type of vehicle. As there was a significant number of PTWs in the experiment location, which, depending on the traffic conditions, tend to filter traffic and move to the edges of the lanes (*44*), their trajectories were excluded in this approach as they would have caused significant noise. Therefore, when the trajectories of the remaining vehicles were plotted, different groups based on their density could be manually identified, as seen in Figure 1*b*. Though different concepts were tested, we chose to use the distance of each trajectory from the edge of the road, as this easily replicates the behavior of vehicles moving in the center of the lanes. It should be noted that all vehicles were tracked using one characteristic point on their central vertical axis. Figure 4 provides an example of how the distance from the edge of the road can illustrate the different lanes in the five-lane, 15-m wide road previously presented in Figure 1. One can easily identify four distinct peaks in the histogram, separated almost every 3 m, which is also the width of the lanes; however, as the trajectories from the slow-moving taxis at the rightmost part of the road section have been included (obviously, there are not as many as in the rest of the lanes) there is broad section displaying a low frequency at the rightmost part of the histogram.

Though one could easily manually identify the four different clusters in Figure 4 to automatize and scale up the process for the whole network, an appropriate clustering method was nonetheless necessary. We choose the Jenks optimization method, a data clustering method that seeks to minimize each class's average deviation

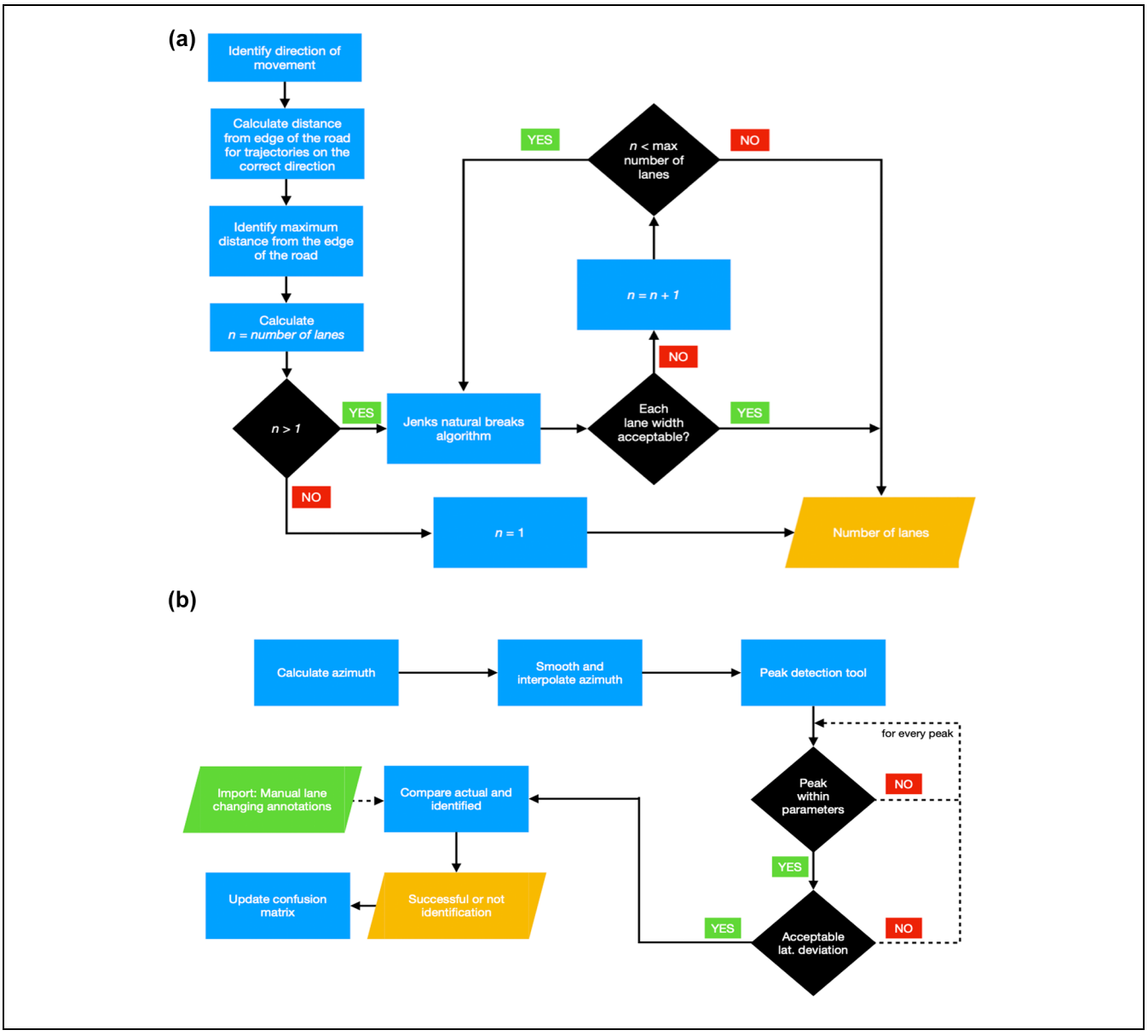


Figure 5. Flow chart of (*a*) lane-detection process and (*b*) lane-changing algorithm.

from the class average, while maximizing each class's deviation from the average of the rest of the clusters (*45*). While this method is sensitive to data containing large measurement error, given the accuracy of drone data and the physical properties of drivers' movements, the method was expected to work properly. Nevertheless, in cases involving sharp turns in the network topology, this method might have experienced difficulties, as will be discussed in Lane-Changing Maneuver.

A flow chart of the methodological framework is presented in Figure 5 and the flow of the lane-detection process at a specific road section is presented in Figure 5*a*. One of the requirements was to know the direction of the road. Therefore, the azimuth (direction of movement) of a vehicle was calculated for every available pair of coordinates (az_{car}). Azimuth is the angle that is formed between the direction of the vehicle compared with the direction of true north, which is measured as a 0° azimuth. Then, based on the mean and standard deviation of the azimuth for each vehicle, and the existence or not of two supplementary angles, we kept only the trajectories that were similar to the azimuth of the road direction (az_{road}) for which we were interested in identifying the number of lanes. The az_{road} represents the average of every vehicle's azimuth traveling that road.

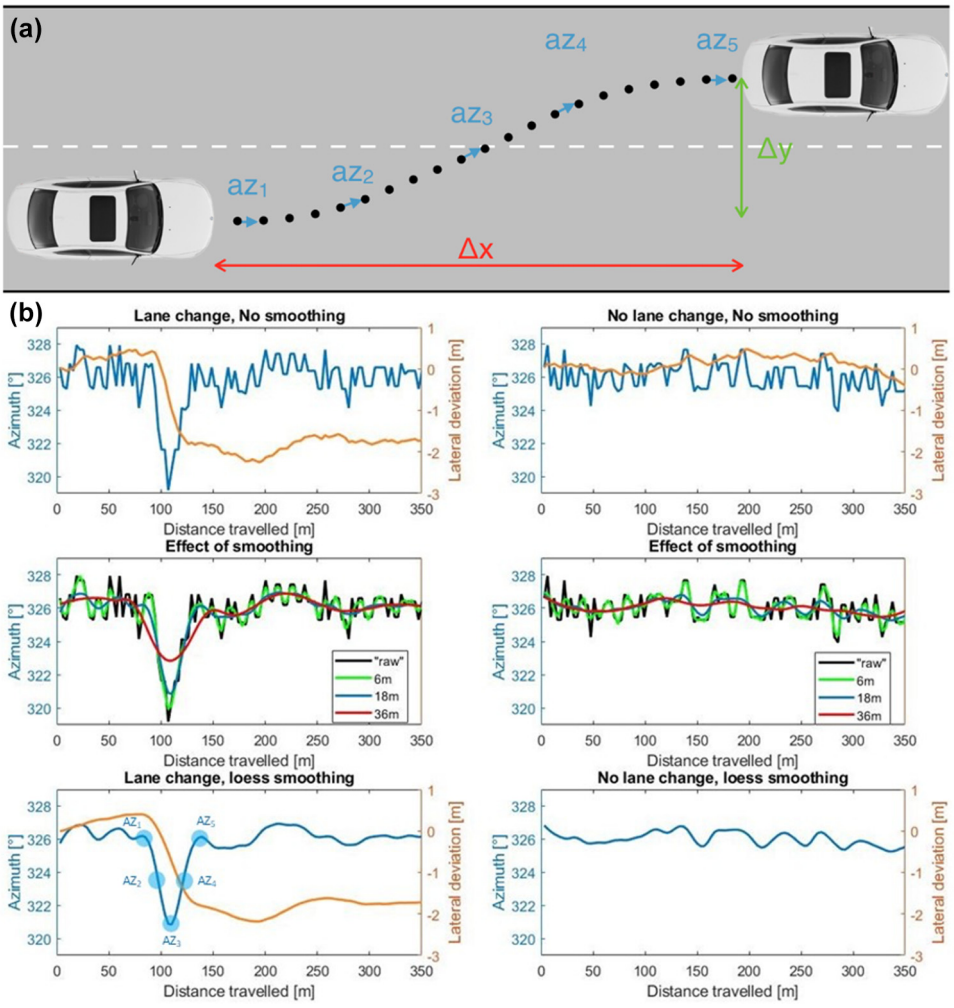


Figure 6. (*a*) Sketch of horizontal and lateral deviation and (*b*) azimuth and lateral deviation illustration for lane-changing detection.

Then, the distance from the edge was calculated. For optimization, the algorithm uses the quotient of the division between the maximum distance to the average lane width as the initial value for the number of lanes. The Jenkins algorithm subsequently clusters the data by calculating the breaks in the histogram, providing the upper and lower limit of each lane (and therefore its width) in relation to distance from the edge of the road. However, although in most cases this process is adequate for accurate results, when it comes to more complex road sections, adding a constraint on the width of the lanes can improve the quality of the results. Specifically, the width of each lane cannot be less than 2.5 m or greater than 4 m and, until this condition is met, the algorithm will keep partitioning the road section into more lanes. By complex road sections, we refer to features such as turning lanes, lanes with parked vehicles, or bus lanes. The results of the algorithm are discussed in detail in the next section.

Lane-Changing Maneuver

For the purpose of this research, a lane-changing maneuver was defined as a vehicle crossing the marker line between two adjacent lanes. The intuitive approach to detecting a lane-changing maneuver is to find the lateral deviation of a vehicle. To this end, we compared the vehicle's azimuth to the azimuth of the road; the flow chart is illustrated in Figure 5*b*. Conceptually, during the maneuver the azimuth is expected to start increasing/decreasing (depending on whether this involves a left- or right lane-change), until it reaches a local maximum/minimum and then it will gradually return to its original value, as the azimuth of the road remains constant on straight roads.

Thus, using the same concept, turning movements can be identified when the azimuth stabilizes at a different value and will therefore not be counted as a lane-changing maneuver. This is illustrated by the az_x points in Figure 6, *a* and *b*. It should be noted that as a first attempt at lane-changing identification, our approach focused solely on straight roads.

To calculate the horizontal Δx and lateral deviation Δy illustrated in Figure 6*a*, we used the cumulative sum of *dx* and *dy* (Equations 1 and 2) over distance D_x traveled by the specific vehicle, so that no noisy data appear when vehicles have stopped at a traffic signal or as a result of congestion. The minimum D_x traveled is was set to 3 m as the optimal value for dealing with both large and small D_x steps (Equation 3).

$$dx = \cos(az_{\text{car}} - az_{\text{road}})^{*} D_x \tag{1}$$

$$dy = \sin(az_{\text{car}} - az_{\text{road}})^{*} D_x \tag{2}$$

$$az_x = a\tan 2[(\sin\Delta\lambda \cdot \cos\varphi 2), (\cos\varphi 1 \cdot \sin\varphi 2 - \sin\varphi 1 \cdot \cos\varphi 2 \cdot \cos\Delta\lambda)] \tag{3}$$

where

φ_1 = the latitude of initial point (positive for N and negative for S),

φ_2 = the latitude of the final point (positive for N and negative for S),

λ_1 = the longitude of the initial point (positive for E and negative for W),

λ_2 = the longitude of the final point (positive for E and negative for W), and

$\Delta\lambda = \lambda_2 - \lambda_1$

To smooth both the azimuth and the lateral deviation, locally weighted scatterplot smoothing (LOESS) was utilized (*46*). Though various smoothing techniques were examined, as a nonparametric smoother that tries to find a curve of best fit without assuming the data must fit some distribution shape, LOESS was found to perform better. We used a second-degree polynomial model while different smoothing parameter values were tested (Figure 6*b*). The raw data of the trajectories was utilized. Therefore, the smoothing process was applied to the azimuth with the objectives of optimizing the lane-changing algorithm and identifying the peak; for other cases (such as for identifying harsh driving events and abrupt decisions) different smoothing parameters could be applied. The smoothing was then spline interpolated to avoid any sharp edges. The interpolated value at every point was based on a cubic interpolation of the values at neighboring grid points, taking place every 30 cm or per 10 raw data points. Both the results of the smoothing process and the variations between lane-changing and lane-keeping behavior can be seen in Figure 6*b*.

Peak Detection

From Figure 6*b*, one can easily understand that the local maxima/minima in the azimuth represent the time the vehicle starts to take up its position in the "new" lane. To detect the peaks in the time series of the smoothed data, a peak detection algorithm was used that returns a vector with the local maxima (peaks) of the input data. The specific algorithm used in this paper was parametrized by setting i) minimum peak prominence and ii) minimum lateral deviation. The process of identifying lane-changing maneuvers is illustrated in Figure 5*b*.

Integration of the Two Algorithms

As discussed, the lane-detection algorithm allows every vehicle to be allocated to a lane at the beginning of its trajectory. Thus, in conjunction with the information that a lane change has been executed to the right or left neighboring lane, one can deduce the lane allocation of a vehicle throughout its trajectory. The integration of the two algorithms (lane-detection and lane changing algorithm) is also a way to further evaluate and improve the predictive power of the above methodologies.

Specifically, from the lane-changing algorithm, one can calculate the precise location of the start and end of a lane-changing maneuver and the vehicle's distance from the edge of the road. Similar concepts have previously been applied when fusing GPS and lidar data (*47*, *48*). If the results do not correspond with the widths estimated from the lane-detection algorithm, this inconsistency between the two methods could be further examined to consider the results of the algorithm with the highest predictive power.

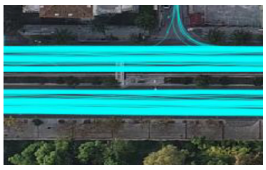
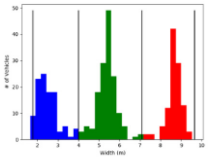
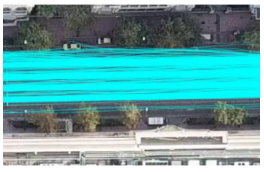
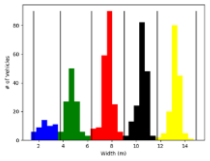
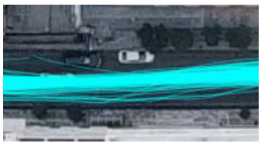
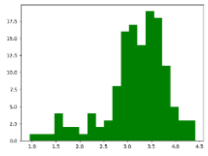
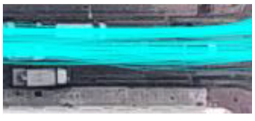
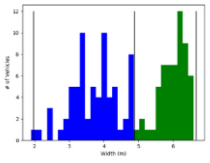
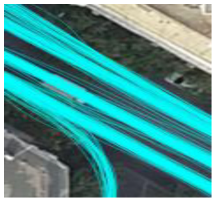
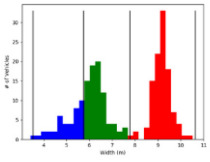
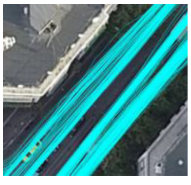
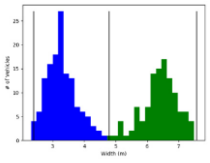
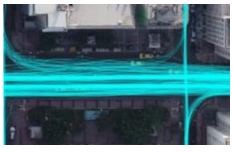
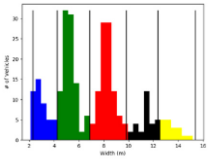
Results

Lane Detection

Eight unique cases were examined to evaluate the lane-detection algorithm described in Methodology. Diverse areas were chosen so that these cases differed from each other not only in the number of lanes, but also in relation to usage of lanes, direction of travel, traffic flow, traffic conditions, traffic signal proximity, and so forth. The results are provided in Table 1.

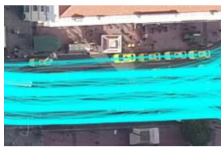
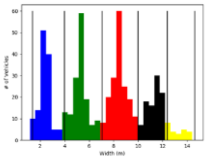
Specifically, the vertical lines that can be seen in the histograms, correspond to the lane markings for each case, therefore the number of lanes can be extracted. It can be observed that the algorithm was able to correctly identify cases with different characteristics and that the direction of travel did not affect the methodology. For example, in Case #2 there is an extra lane that is used only by buses, whereas in cases #1, #5, and #6 data from both directions are included. Parked vehicles do not seem to affect the algorithm, as can be seen in cases #3

Table 1. Lane-Detection Algorithm Results

	Screenshot	# lanes	Turning lane	Blocked lane	Resulting histogram
#1		3	na	na	
#2		5	na	na	
#3		2	na	1	
#4		2	na	na	
#5		3	1	na	
#6		2	na	na	
#7		5	na	1	

(continued)

Table 1. (continued)

	Screenshot	# lanes	Turning lane	Blocked lane	Resulting histogram
#8		5	na	1	
#9		5	na	1	na

Note: na = not applicable.

and #4. In cases where lanes are used for temporary parking (cases #7 and #8), it can be seen that the algorithm is able to identify both the differentiation in lane and the reduced capacity it provides for traffic flow. Last but not least, turning lanes can also be identified, as evident in Case #5.

One of the limitations of the developed algorithm can be identified in Case #9 involving a curve, as it was not able to provide the correct result. It was therefore surmised that to extract the number of lanes for special cases like turns or roundabouts, a different methodology should be followed.

Lane-Changing Maneuver Detection

As described in Methodology, there are numerous possible parametrization options for the lane-changing algorithm. To better examine the effect of each of the different parameters on the predictive power of the algorithm, a sensitivity analysis was conducted for the three parameters presented in Table 2.

Two different study areas were selected for the evaluation of the algorithm, although it could be applied to any part of the road network. The first was a 400-m, three-lane arterial during 8:30 to 9:00 a.m.; the second was a 400-m, five-lane arterial during 10:00 to 10:30 a.m. In total, 602 vehicles were included in the testing dataset. It should be noted that each case was reviewed manually to evaluate the results of the detection algorithm, and two vehicles conducted two lane-changes.

The different combinations of models resulting from the sensitivity analysis formed the receiver operating characteristic (ROC) curve that is seen in Figure 7. The area under the ROC (AUROC) curve illustrates the predictive ability of a binary classifier system as its parametrization changes, which for our methodology was over 98%. As the ROC curve shows the tradeoff across the parametrization of the classifier system, it can be observed that the methodology proposed has strong predictive ability for the various arithmetic values of the parameters (*49*).

Table 2. Sensitivity Analysis Information

Parameter	Sequence start	Sequence end	Step
LOESS smoothing parameter (m)	14	30	4
Peak prominence (degrees)	1	3	0.2
Lateral deviation (m)	0.8	1.2	0.1

Note: LOESS = locally weighted scatterplot smoothing.

From the models that were formed, we chose to describe the model with the best accuracy. The parametrization of this specific model was i) smoothing parameter value equal to 18 m, ii) peak prominence equal to 1.6°, and iii) lateral deviation equal to 1 m. True positive (TP) refers to the lane-changing maneuvers that were correctly identified, false negative (FN) to the ones that were not identified, false positive (FP) to lane keeping that was incorrectly identified as a lane change, and true negative (TN) to denote lane keeping that was not identified as a lane change. The confusion matrix and the classification metrics are provided in Tables 3 and 4 respectively.

The results indicate that the algorithm has high predictive power in both identifying lane- changing (TP) and lane-keeping behavior (TN). Each of the FP and FN

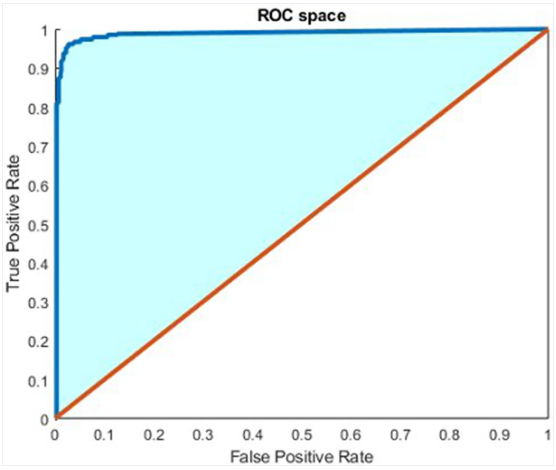


Figure 7. The Auroc that is formed for the different values of the parameters.
Note: ROC = receiver operating characteristic.

identifications were reviewed manually and it was found that a similar pattern appeared for both types of errors. Specifically, FPs may appear when aggressive drivers conduct a lateral maneuver without crossing the edges of their current lane. On the other hand, when very smooth lane-changing takes place, that is, both the azimuth and the lateral deviation change slowly over a long distance, the algorithm cannot identify any significant peaks in the time series.

The detection algorithm allowed the determination of lane-changing events alongside a multitude of additional characteristics such as i) lateral deviation, ii) speed and acceleration information, and iii) location and time of the event start-, mid-, and end point. This information could be valuable for research related to lane-changing phenomena, taking into account the significant number of lane-changing maneuvers that were monitored in the *pNEUMA* dataset. Specifically, a heat-map can be created: Figure 8 illustrates the position of lane-changing maneuvers along a stretch of road, extracted from a sample of 450 trajectories over 15 min. The color of the map represents the number of lane-changing maneuvers, e.g. darker colors indicate a greater number of lane changing maneuvers.

The results of the lane-changing maneuvers (red marks) can also be illustrated in a time–space diagram, as in Figure 9.

Integration of the Two Algorithms

The same arterial of three lanes was selected to examine the integration of the preceding results and determine precisely the lane allocation of each vehicle for its timestamp. As outlined in the Methodology section, the lane-detection algorithm is used to determine the lane allocation at the start of the road and lane-change detection allows modification of the allocation should a lane change occur. Figure 10 shows the results from the integration of the two algorithms for a particular vehicle that conducted two sequential lane-changes. Specifically, starting in the right-hand lane, the vehicle (traveling from right to left) conducted three separate lane-changes (right to middle, middle to left, and left to middle), ending its observed trajectory in the middle lane. The color scale shows that the lane allocation of a vehicle is available for every timestamp of its trajectory.

Over the sample of 450 trajectories, 164 lane changes were conducted and analyzed further to evaluate the consistency of the integrated algorithm. Using the lane-detection algorithm, three lanes were estimated, with lines separating them situated 3.1 m and 6.2 m away from the edge of the road. The lane origin/destination is provided in Table 5.

It can be seen that 18 detected lane-changes (10.9% of the total) stayed in the same estimated lane. This was partly owing to the 2.7% FP rate of the lane-changing algorithm; the vehicle is expected to stay in the same lane. Should the lane width be totally accurate and the lane-changing start- and end point be perfectly estimated,

Table 3. Confusion Matrix for the Lane-Detection Algorithm

		Actual values	
		Actual lane changes	Actual lane keeping
Identified values	Identified lane changes	TP = 291	FP = 8
	Identified lane keeping	FN = 12	TN = 293

Note: TP = true positive; FN = false negative; FP = false positive; TN = true negative.

Table 4. Classification Metrics for the Lane-Detection Algorithm

Measure	Derivation	Value
Sensitivity (true positive rate)	TP/(TP + FN)	96.0%
Specificity	TN/(FP + TN)	97.3%
Precision	TP/(TP + FP)	97.3%
Negative prediction rate	TN/(TN + FN)	96.1%
False positive rate	FP/(FP + TN)	2.7%
False discovery rate	FP/(FP + TP)	2.7%
False negative rate	FN/(FN + TP)	4.0%
Accuracy	(TP + TN)/(P + N)	96.7%
F1 score	2TP/(2TP + FP + FN)	96.7%

Note: TP = true positive; FN = false negative; FP = false positive; TN = true negative; P = positives; N = negatives.

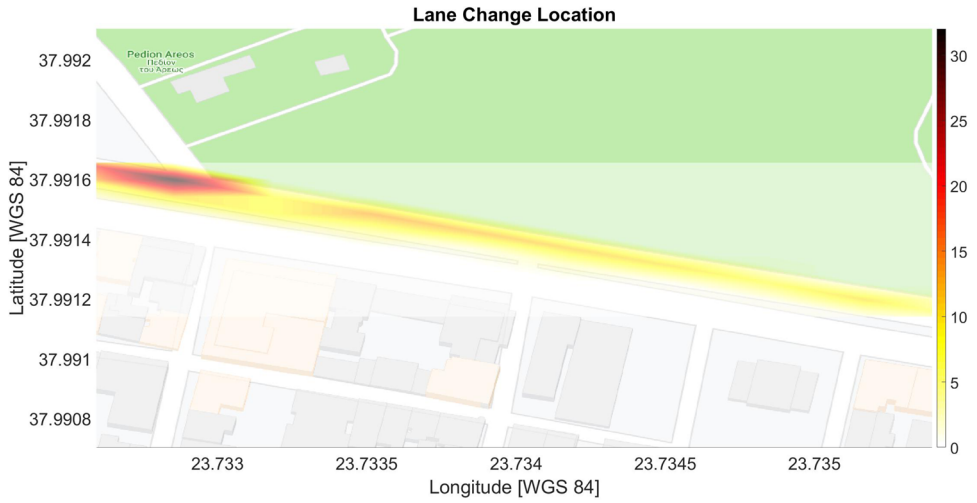


Figure 8. Lane-changing location along a road in a 15-min time frame.
Note: WGS = World Geodetic System.

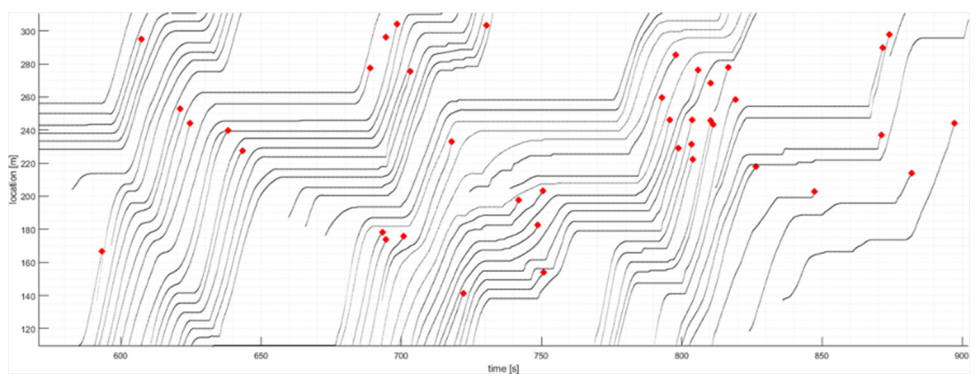


Figure 9. Lane-changing events as illustrated in a time–space diagram.

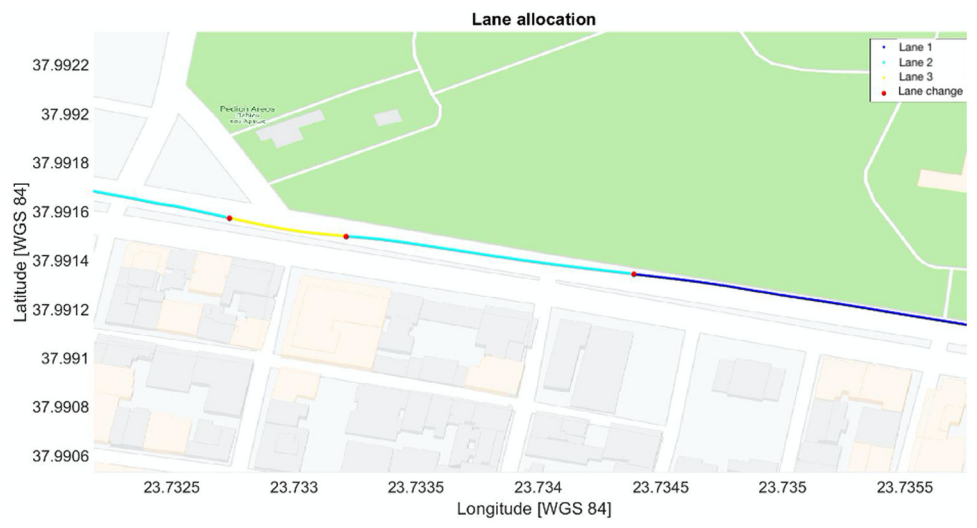


Figure 10. Lane allocation of a single vehicle over its trajectory.
Note: WGS = World Geodetic System.

Table 5. Lane Origin/Destination Table of Lane-Changing Maneuvers

	To Lane 1	To Lane 2	To Lane 3
From Lane 1	10	18	1
From Lane 2	76	2	18
From Lane 3	5	12	6

a vehicle would not be in the same lane after the maneuver. Therefore, the 8.2% of actual lane-changes without allocation, highlight the inconsistencies between the two approaches. It should be noted that that the six double lane-change events (1 to 3 or 3 to 1) reported, did not represent a vehicle going from Lane 1 to Lane 2 and later to Lane 3, but a vehicle going in a single detected event from the right-hand lane to the leftmost one.

Conclusion

In this paper, the methodology for acquiring lane-specific information from a large-scale urban dataset from aerial video footage is described. These approaches can be considered a benchmark for future attempts and concepts. It has been seen that when a high-quality dataset like *pNEUMA* is available, high predictive power can be achieved without the use of powerful or complex computational tools.

Specifically, the lane-detection algorithm demonstrated that challenging cases, such as the actual use of lanes, could be correctly identified. This methodology could be significant as, sometimes, visual information (satellite images, images, or low-quality video) cannot provide the full picture for researchers and practitioners. In addition, local or unexpected phenomena could be identified, such as the bottlenecks that are created by stopped vehicles at the edge of an arterial, a typical cause of congestion in dense urban environments. It should also be noted that the proposed methodology depended solely on real vehicle trajectories, not on road designs that might have included outdated information or been affected by traffic flow conditions or fixed information.

A precise and flexible lane-changing detection tool has been developed using computationally light techniques and concepts that require only spatial information (WGS 84 coordinates). Typical lane-changing and lane-keeping cases were successfully identified in different road environments, without being affected by the number of lanes or traffic flow conditions.

Finally, false identification patterns indicate that modifications in the algorithm could further improve its power of detectability. In future, we aim to improve detectability for the challenging cases that the algorithms failed to identify correctly. Moreover, this information will be utilized to determine the strength of the relationship between the number of lane changes and the already identified indicators of driving aggressiveness, especially on congested urban arterials.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: EB, NG; analysis and interpretation of results: EB, GS; manuscript preparation: EB, GS, NG. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially funded by Swiss National Science Foundation (grant no. 200021_188590) "pNEUMA: On the new era of urban traffic models with massive empirical data from aerial footage."

Data Accessibility Statement

Data source: pNEUMA—https://open-traffic.epfl.ch

References

1. Coifman, B., and L. Li. A Critical Evaluation of the Next Generation Simulation (NGSIM) Vehicle Trajectory Dataset. *Transportation Research Part B: Methodological*, Vol. 105, 2017, pp. 362–377. https://doi.org/10.1016/j.trb.2017.09.018.
2. Barmpounakis, E. N., and N. Geroliminis. Utilizing a Swarm of Drones for Large-Scale Traffic Measurements. *Proc.,19th Swiss Transport Research Conference*, Ascona, Switzerland, 2019.
3. Jung, S., J. Youn, and S. Sull. Efficient Lane Detection Based on Spatiotemporal Images. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 1, 2016, pp. 289–295.
4. Ozgunalp, U., R. Fan, X. Ai, and N. Dahnoun. Multiple Lane Detection Algorithm Based on Novel Dense Vanishing Point Estimation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 3, 2017, pp. 621–632.
5. Sivaraman, S., and M. M. Trivedi. Integrated Lane and Vehicle Detection, Localization, and Tracking: A Synergistic Approach. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 2, 2013, pp. 906–917.
6. Li, J., X. Mei, D. Prokhorov, and D. Tao. Deep Neural Network for Structural Prediction and Lane Detection in

Traffic Scene. *IEEE Transactions on Neural Networks and Learning Systems*. Vol. 28, No. 3, 2017, pp. 690–703.
7. NGSIM. *Next Generation Simulation*. 2006. http://ngsim.fhwa.dot.gov/.
8. Krajewski, R., J. Bock, L. Kloeker, and L. Eckstein. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. *Proc., IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, 2018, pp. 2118–2125.
9. Knoop, V. L., P. F. De Bakker, C. C. Tiberius, and B. Van Arem. Lane Determination with GPS Precise Point Positioning. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 9, 2017, pp. 2503–2513.
10. Tang, L., X. Yang, Z. Kan, and Q. Li. Lane-Level Road Information Mining from Vehicle GPS Trajectories Based on Naïve Bayesian Classification. *ISPRS International Journal of Geo-Information*, Vol. 4, No. 4, 2015, pp. 2660–2680.
11. Chen, Y., and J. Krumm. Probabilistic Modeling of Traffic Lanes From GPS Traces. *Proc., 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, San Jose, CA, 2010, pp. 81–88.
12. Kesting, A., M. Treiber, and D. Helbing. General Lane-Changing Model MOBIL for Car-Following Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2007. 1999: 86–94.
13. Laval, J. A., and C. F. Daganzo. Lane-Changing in Traffic Streams. *Transportation Research Part B: Methodological*, Vol. 40, No. 3, 2006, pp. 251–264. https://doi.org/10.1016/j.trb.2005.04.003.
14. Coifman, B., S. Krishnamurthy, and X. Wang. Lane-Change Maneuvers Consuming Freeway Capacity. In: *Traffic and Granular Flow '03* (S. P., Hoogendoorn, S. Luding, P. H. L. Bovy, M. Schreckenberg, and D. E. Wolf, eds.), Springer, Berlin, Heidelberg, 2007.
15. Wei, H., J. Lee, Q. Li, and C. J. Li. Observation-Based Lane-Vehicle Assignment Hierarchy: Microscopic Simulation on Urban Street Network. *Transportation Research Record: Journal of the Transportation Research Board*, 2007. 1710: 96–103.
16. Nagel, K., D. E. Wolf, P. Wagner, and P. Simon. Two-Lane Traffic Rules for Cellular Automata: A Systematic Approach. *Physical Review E*, Vol. 58, No. 2, 1998, p. 1425.
17. Hidas, P. Modelling Lane Changing and Merging in Microscopic Traffic Simulation. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 5–6, 2002, pp. 351–371. https://doi.org/10.1016/S0968-090X(02)00026-8.
18. Li, X., and J. Q. Sun. Studies of Vehicle Lane-Changing Dynamics and Its Effect on Traffic Efficiency, Safety and Environmental Impact. *Physica A: Statistical Mechanics and its Applications*. Vol. 467, 2017, pp. 41–58.
19. Gipps, P. G. A Model for the Structure of Lane-Changing Decisions. *Transportation Research Part B: Methodological*,Vol. 20, No. 5, 1986, pp. 403–414. https://doi.org/10.1016/0191-2615(86)90012-3.
20. Toledo, T, H. N. Koutsopoulos, and M. E. Ben-Akiva. Modeling Integrated Lane-Changing Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2003. 1857: 30–38.
21. Ramezani, M., and N. Geroliminis. Queue Profile Estimation in Congested Urban Networks with Probe Data. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 30, No. 6, 2015, pp. 414–32. https://doi.org/10.1111/mice.12095.
22. Thiemann, C., M. Treiber, and A. Kesting. Estimating Acceleration and Lane-Changing Dynamics from Next Generation Simulation Trajectory Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2008. 2088: 90–101.
23. Punzo, V., M. T. Borzacchiello, and B. Ciuffo. On the Assessment of Vehicle Trajectory Data Accuracy And Application to the Next Generation Simulation (NGSIM) Program Data. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 6, 2011, pp. 1243–1262. https://doi.org/10.1016/j.trc.2010.12.007.
24. Hoogendoorn, S. P, W. Daamen, and P. H. Bovy. Extracting Microscopic Pedestrian Characteristics from Video Data. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.
25. Barmpounakis, E. N., E. I. Vlahogianni, and J. C. Golias. Identifying Predictable Patterns in the Unconventional Overtaking Decisions of PTW for Cooperative ITS. *IEEE Transactions on Intelligent Vehicles*, Vol. 3, No. 1, 2018, pp. 102–111.
26. European Comission. *C - ITS Platform: Final Report*. 2016. https://ec.europa.eu/transport/sites/transport/files/themes/its/doc/c-its-platform-final-report-january-2016.pdf.
27. Lee, J., and B. Park. Development and Evaluation of a Cooperative Vehicle Intersection Control Algorithm under the Connected Vehicles Environment. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 1, 2012, pp. 81–90.
28. Bauza, R., J. Gozalvez, and J. Sanchez-Soriano. *Road Traffic Congestion Detection through Cooperative Vehicle-to-Vehicle Communications*. *Proc., IEEE Local Computer Network Conference*, Denver, CO, 2010, pp. 606–612.
29. Kaufmann, S., B. S. Kerner, H. Rehborn, M. Koller, and S. Klenov. Aerial Observation of Inner City Traffic and Analysis of Microscopic Data at Traffic Signals. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
30. Laval, J. A., and L. Leclercq. A Mechanism to Describe the Formation and Propagation of Stop-and-Go Waves in Congested Freeway Traffic. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, Vol. 368, No. 1928, 2010, pp. 4519–4541. https://doi.org/10.1098/rsta.2010.0138
31. Pan, T. L., W. H. Lam, A. Sumalee, and R. X. Zhong. Modeling the Impacts of Mandatory and Discretionary Lane-Changing Maneuvers. *Transportation Research Part C: Emerging Technologies*, Vol. 68, 2016, pp. 403–424. https://doi.org/10.1016/j.trc.2016.05.002.
32. Xiaorui, W., and Y. Hongxu. A Lane Change Model with the Consideration of Car Following Behavior. *Procedia - Social and Behavioral Sciences*, Vol. 96, 2013,

pp. 2354–2361. https://doi.org/10.1016/j.sbspro.2013.08.264.

33. Talebpour, A., H. S. Mahmassani, and S. H. Hamdar. Modeling Lane-Changing Behavior in a Connected Environment: A Game Theory Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 7, 2015, pp. 216–232. https://doi.org/10.1016/j.trpro.2015.06.022.
34. Balal, E., R. L. Cheu, and T. Sarkodie-Gyan. A Binary Decision Model for Discretionary Lane Changing Move Based on Fuzzy Inference System. *Transportation Research Part C: Emerging Technologies*, Vol. 67, 2016, pp. 47–61. https://doi.org/10.1016/j.trc.2016.02.009.
35. Wei, H., E. Meyer, J. Lee, and C. Feng. Characterizing and Modeling Observed Lane-Changing Behavior: Lane-Vehicle-Based Microscopic Simulation on Urban Street Network. *Transportation Research Record: Journal of the Transportation Research Board*, 2007. 1710: 104–113.
36. Vlahogianni, E. I. Powered-Two-Wheelers Kinematic Characteristics and Interactions during Filtering and Overtaking in Urban Arterials. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 24, 2014, pp. 133–145. https://doi.org/10.1016/j.trf.2014.04.004.
37. Barmpounakis, E. N., E. I. Vlahogianni, and J. C. Golias. Modeling Cooperation and Powered-Two Wheelers Short-Term Strategic Decisions during Overtaking in Urban Arterials. *International Journal of Transportation Science and Technology*, Vol. 5, No. 4, 2016, pp. 227–238.
38. Barmpounakis, E. N., E. I. Vlahogianni, and J. C. Golias. Intelligent Transportation Systems and Powered Two Wheelers Traffic. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 4, 2016, pp. 908–916.
39. Vlahogianni, E. I., and E. N. Barmpounakis. Driving Analytics Using Smartphones: Algorithms, Comparisons and Challenges. *Transportation Research Part C: Emerging Technologies*, Vol. 79, 2017, pp. 196–206. https://doi.org/10.1016/j.trc.2017.03.014.
40. Aly, H., A. Basalamah, and M. Youssef. Robust and Ubiquitous Smartphone-Based Lane Detection. *Pervasive and Mobile Computing*, Vol. 26, 2016, pp. 35–56. https://doi.org/10.1016/j.pmcj.2015.10.019.
41. Hamdar, S. H., H. S. Mahmassani, and R. B. Chen. Aggressiveness Propensity Index for Driving Behavior at Signalized Intersections. *Accident Analysis and Prevention*, Vol. 40, No. 1, 2008, 315–326. https://doi.org/10.1016/j.aap.2007.06.013.
42. Kumtepe, O., G. B. Akar, and E. Yuncu. Driver Aggressiveness Detection via Multisensory Data Fusion. *EURASIP Journal on Image and Video Processing*, Vol. 2016, No. 1, 2016, pp. 1–16. https://doi.org/10.1186/s13640-016-0106-9.
43. Goswami, V., and G. H. Bham. Gap Acceptance Behavior in Mandatory Lane Changes under Congested and Uncongested Traffic on a Multilane Freeway. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
44. Nikias, V., E. I. Vlahogianni, T. C. Lee, and J. C. Golias. Determinants of Powered Two-Wheelers Virtual Lane Width in Urban Arterials. *Proc., 15th International IEEE Conference on Intelligent Transportation Systems*, Anchorage, AK, 2012, pp. 1205–1210.
45. Jenks, G.F. The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, Vol. 7, 1967, pp. 186–190.
46. Cleveland, W. S., and S. J. Devlin. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, Vol. 83, No. 403, 1988, pp. 596–610.
47. Gao, B., and B. Coifman. Vehicle Identification and GPS Error Detection from a LIDAR Equipped Probe Vehicle. *Proc., IEEE Intelligent Transportation Systems Conference*, Toronto, Ontario, 2006, pp. 1537–1542.
48. Xuan, Y., and B. Coifman. Identifying Lane-Change Maneuvers with Probe Vehicle Data and an Observed Asymmetry in Driver Accommodation. *Journal of Transportation Engineering*, Vol. 138, No. 8, 2012, pp. 1051–1061.
49. Bradley, A. P. The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, Vol. 30, No. 7, 1997, pp. 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2.

The Standing Committee on Traffic Flow Theory and Characteristics (AHB45) peer-reviewed this paper (20-01415), which was selected to receive the 2020 Greenshields Prize awarded by the committee honoring the contributions of Bruce Greenshields to the field of traffic flow theory.