

## Article

# Image Data Extraction and Driving Behavior Analysis Based on Geographic Information and Driving Data

Huei-Yung Lin <sup>1</sup>, Jun-Zhi Zhang <sup>2</sup> and Chin-Chen Chang <sup>3,\*</sup>

<sup>1</sup> Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 10608, Taiwan; lin@ntut.edu.tw

<sup>2</sup> Department of Electrical Engineering, National Chung Cheng University, Chiayi 621, Taiwan; [jffc850629@gmail.com](mailto:jffc850629@gmail.com)

<sup>3</sup> Department of Computer Science and Information Engineering, National United University, Miaoli 36003, Taiwan; ccchang@nuu.edu.tw

\* Correspondence: ccchang@nuu.edu.tw

**Abstract:** This paper proposes an approach for image data extraction and driving behavior analysis using geographic information and driving data. Information derived from geographic and global positioning systems was used for image data extraction. In addition, we used an onboard diagnostic II and a controller area network bus logger to record driving data for driving behavior analysis. Driving behavior was analyzed using sparse automatic encoders and data exploration to detect abnormal and aggressive behavior. A regression analysis was used to derive the relationship between aggressive driving behavior and road facilities. Several traffic improvements were proposed for specific intersections and roads. The results indicate that lane ratios without lane markings and with straight lane markings are important features that affect aggressive driving behaviors.

**Keywords:** image data extraction; driving behavior analysis; geographic information system; global position system

## 1. Introduction

Considerable research has been conducted on autonomous vehicle systems for self-driving cars. One key component of self-driving cars is the understanding of human driving behavior to avoid human-machine conflict [1-3]. With recent advances in machine learning techniques, data-driven approaches for complicated human behavior modeling have become more increasingly feasible. In particular, several studies have made significant progress using deep-learning techniques [4].

In recent learning approaches, in addition to the design of network structures, another critical issue is the collection of large amounts of training data. Autonomous vehicle systems typically collect data from onboard sensors and extract information for specific analyses. These may encompass images captured by in-car cameras and proprioceptive driving data recorded by onboard diagnostic systems. The extraction of adequate data segments for neural network training and testing is critical. For instance, learning road sign recognition uses certain traffic scene images or modeling a driver's acceleration behavior using selected gas pedal information. Large training datasets are typically used in deep neural networks to achieve better performance.

In early related research, data annotation or labeling was mainly completed manually and sometimes through crowdsourcing. For driving data, the dataset collected according to different tasks contained various scenes and features. The selection and filtering of adequate data require significant time and human labor. The development of technologies for searching specific traffic scenes within a large number of image sequences has become important.

In this paper, we propose an image data extraction system based on geographic information and driving behavior analysis using various types of driving data. We used information derived from a geographic information system (GIS) and a global positioning system (GPS) with recorded driving videos to identify road scenes with static objects such as traffic lights, traffic signs, bridges, and tunnels. Additionally, we utilized an onboard diagnostic II (OBD-II) and a controller area network bus (CANbus) logger to collect driving data [5]. By measuring several parameters at a high sampling rate, we can completely observe the driving behaviors and understand the influences of traffic and road infrastructure.

To analyze the relationship between driving behavior and transportation infrastructure, visualization on a map provides a way for better observation and investigation. We used machine learning methods to extract unique features from the driving data and then mapped these features to the RGB color space to visualize the driving behavior. Data mining algorithms were utilized for data analysis and to classify driving behavior into four categories, from normal to aggressive. A regression analysis was then conducted on the relationship between aggressive driving behavior and road features at intersections.

The contributions of this paper are as follows. a) We present a road scene extraction approach for specific landmarks and transportation infrastructure indicators. b) To perform a more comprehensive analysis, we present visual driving behavior, aggressive driving behavior, and traffic light information on a map to associate driving behavior with roads. c) Based on the results of the driving behavior analysis, suggestions and verifications for traffic improvement at intersections are presented to prevent accidents.

## 2. Related Works

As learning-based algorithms become popular, the acquisition of training and testing data has become crucial. Current data extraction approaches are primarily divided into two categories: information and image-based approaches. Hornauer et al. and Wu et al. [6-7] proposed unsupervised image classification methods for extracting images similar to those provided by general users [8]. In supervised classification, the data were manually labeled. People need a similar understanding to annotate the same scene. The network of this approach is based on the feature similarity for a first-person driving image query. However, this may not fully meet the needs of the users.

In addition to image extraction and classification, Naito et al. [9] proposed a browsing and retrieval approach for analyzing driving data. Their approach provides a multidata browser, a retrieval function based on queries and similarities, and a quick browsing function to skip extra scenes. The top- $N$  images that were highly similar to the current driving scenario were retrieved from a database for scene retrieval. While an image sequence is processed, this approach calculates the similarity between the input scene and the scenes stored in the database. A predefined threshold was used to identify similarities between the images. Because the approach mainly searched for the driving video itself, it could not be determined whether the images contained objects or information that was interesting to the users for precise extraction.

The key technologies for automotive driving-assistance systems have matured [10]. However, autonomous vehicles can still not be employed without human drivers. Owing to the current limitations of driving assistance systems, researchers and developers are seeking solutions to improve human driving capabilities. Because changing driving habits is challenging, developing a human-centered driving environment to avoid dangerous situations is crucial. By understanding the relationship between traffic lights, road infrastructure, and driving behavior, suggestions for transportation improvement can be provided. In addition, knowing human reactions is also a crucial issue for mixed human drivers and self-driving cars.

For driving behavior analysis, Liu et al. [11-12] proposed a method that uses various types of sensors connected to a control area network. A deep sparse autoencoder (SAE)

was then used to extract hidden features from the driving data to visualize driving behavior. Constantinescu et al. [13] used both the principal component analysis (PCA) and hierarchical cluster analysis (HCA) techniques to analyze driving data. The performance of their techniques was verified by classifying the driving behavior into six categories based on aggressiveness. In the approach proposed by Kharrazi et al. [14], driving behavior was classified into three categories, calm, normal, and aggressive, using quartiles and *K*-means clustering. Their results demonstrated that *K*-means clustering can provide good classification results for driving behaviors.

Tay et al. [15] used a regression model to associate driving accidents with environmental factors. Wong et al. [16] utilized a negative binomial regression to analyze the number of driving accidents and road features at intersections. Road intersections can also be improved by simulations based on the analysis results. Schorr et al. [17] presented an approach for recording the driving data in one- and two-way lanes. Based on analysis of variance (ANOVA), a conclusion regarding the impact of lane width on driving behavior was drawn. Abojaradeh et al. [18] proposed a method to identify driving behaviors and driver mistakes based on questionnaires and highlighted their effects on traffic safety. They used regression analysis to derive the correlation between the number of accidents and types of dangerous driving behaviors.

Regarding the improvement of transportation infrastructure, various suggestions have been proposed for different road and intersection designs. Chunhui et al. [19] proposed an approach to optimize signal lights at intersections to make pedestrian crossing easier [20]. The efficiency of intersections was improved by reducing the conflicts between turning vehicles and pedestrians. Ma et al. [21] proposed a technique for adding a dedicated left-turn lane and waiting area based on the average daily traffic volume at an intersection. The proposed method can supply more vehicles waiting for a left-turn. They also analyzed three common left-turn operation scenarios at intersections and compared their differences. In addition to suggestions for road infrastructure, several traffic improvements have been suggested based on traffic-light analysis. Anjana et al. [22] presented a method based on different traffic volumes at intersections to evaluate the safety of green traffic lights.

### 3. Dataset Extraction

For image data extraction, we first collected information regarding traffic lights, traffic signs, and road information from OpenStreetMap (OSM) [12] and the government's GIS-T transportation geographic information storage platform. Traffic and road information were used to identify locations of interest using GPS coordinates. We compared the GPS information obtained from the driving data to the locations of interest. The associated images were then extracted and stored in video sequences for specific applications, such as training and testing data for traffic light detection. Figure 1 illustrates a flowchart of image data extraction.

The driving recorder contained images with resolutions of  $1280 \times 720$  and a  $110^\circ$  field-of-view (FOV) in the horizontal direction. To extract suitable image data, users must consider the geographic range of the target. For example, in road scene extraction with traffic lights, the size of the traffic signal in an image may be larger than  $25 \times 25$  pixels for specific tasks. This corresponds to approximately 50 m away from the vehicle; thus, the video should be pushed back for 5 s to start image data extraction.

We established an interface for user operations. The interface structure is divided into two parts: image extraction and video filtering. The items included in the image data extraction were traffic lights and traffic signs, whereas the video filtering included highways and tunnels. A program interface was created for users to easily operate the data and assign parameters. It consists of a folder for selection, an item menu for extraction, an OSM map display, and a driving image screen. The user first selects the folder where the driving record video and driving GPS information are located, and the folder where the

extracted image will be stored. The user then selects traffic infrastructure or road information for extraction. At the interface, the vehicle’s GPS trajectory and user-selected traffic infrastructure simultaneously overlay the OSM window, and a synchronized driving video is displayed on the right for inspection.

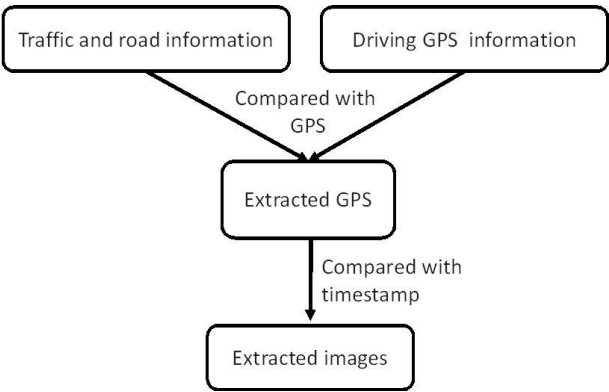


Figure 1. A flowchart of image data extraction.

4. Driving Behavior Analysis

In the proposed approach, we analyzed driving behavior and its correlation with traffic and road features. Driving behavior was classified as normal or aggressive and analyzed through data visualization and a regression model of the number of aggressive driving behaviors and road features.

A flowchart of the driving behavior analysis is presented in Figure 2. First, we used the ODB-II and CANbus loggers for collection of driver data. We then utilized a SAE to extract features from the driving data and compress the high-dimensional features into three dimensions. Subsequently, we mapped the three-dimensional features to the RGB space for display on the OSM. Moreover, we used the K-means clustering algorithm to further classify the driving behavior based on aggressiveness. Finally, we used a negative binomial regression model to analyze the road features at intersections and interchanges.

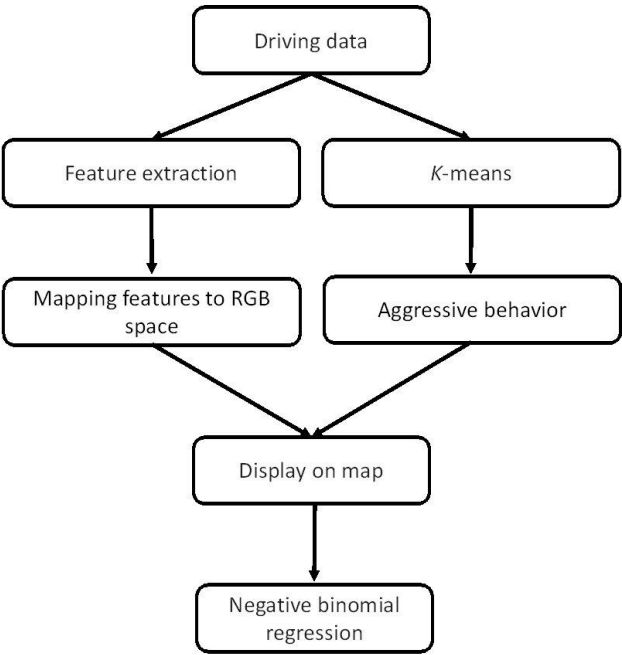


Figure 2. A flowchart of driving behavior analysis.

#### 4.1 Data Collection

In addition to exteroceptive sensors (such as LiDAR, GPS, and cameras), the information collected from the proprioceptive sensors of the vehicle can also be used to analyze driving behavior [23]. Sensor data derived directly from vehicle operations can provide more comprehensive driving information.

We used OBD-II and CANbus loggers for driver data collection. Unlike most previous studies that only used the information obtained from GPS receivers (with GPS messages, vehicle speed, and acceleration), OBD-II and CANbus loggers can collect various types of driving data for analysis. The specific data types used for the proposed driving behavior analysis are as follows.

- OBD-II: engine rotation speed, engine load, throttle pedal position, acceleration XYZ, and vehicle speed.
- CANbus logger: engine rotation speed, throttle pedal position, braking pedal position, steering angle, wheel speed, and vehicle speed.
- GPS receiver: GPS and coordinated universal time (UTC).

In addition, we used two public datasets: the DDD17 dataset [24] and UAH DriveSet [25]. These two datasets are acquired by the driving monitoring application DriveSafe and are mainly used to verify the classification and analysis methods [26].

#### 4.2 Visualization of Driving Behavior

The relationship between driving behavior and traffic infrastructure can be observed through the data visualization on a map. We used the SAE to extract features from the driving data, compress the high-dimensional features into three dimensions, and map the 3D features to RGB space for display on the OSM. The loss function with sparse constraints is given as follows:

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho || \hat{\rho}_j) \quad (1)$$

where  $J(W, b)$  is the cost function with parameters  $W$  and  $b$  and  $\beta$  controls the weight of the sparsity penalty term. The term  $\rho$  is a sparsity parameter, and  $\hat{\rho}_j$  is the average activation of hidden unit  $j$ . Moreover,  $s_2$  is the number of neurons in the hidden layer, and  $KL(\rho || \hat{\rho}_j)$  is Kullback–Leibler divergence between  $\rho$  and  $\hat{\rho}_j$ .

The difference between the SAE and autoencoder (AE) is that a penalty term is added to the loss function. Thus, the activation of hidden nodes decreases to the required value. Using this property, the relative entropy is added to the loss function to penalize the value of the average activation degree far away from the level  $\rho$ . The parameters maintain the average degree of activation of the hidden nodes at this level. Thus, the loss function only requires the addition of a penalty term for relative entropy without sparse constraints.

Figure 3 illustrates the flowchart and network structure used to visualize driving behavior. The network contained nine hidden layers, and the dimensionality reduction of each layer was half the number of nodes in the previous layer. The data collected by OBD-II contained seven types and 70 dimensions after the windowing process. Thus, the dimension reduction in the network is  $70 \rightarrow 35 \rightarrow 17 \rightarrow 8 \rightarrow 3 \rightarrow 8 \rightarrow 17 \rightarrow 35 \rightarrow 70$ , and the features are extracted by the last five layers. The data collected by the CANbus logger contained six types and 60 dimensions after the windowing process. Likewise, the input to the network comprises 60 nodes, and the dimension reduction is given as follows:  $60 \rightarrow 30 \rightarrow 15 \rightarrow 7 \rightarrow 3 \rightarrow 7 \rightarrow 15 \rightarrow 30 \rightarrow 60$ . Finally, driving behavior was visualized using the OSM. Figure 4 shows an example of the driving behavior visualized on the OSM for the network structure  $70 \rightarrow 35 \rightarrow 17 \rightarrow 8 \rightarrow 3 \rightarrow 8 \rightarrow 17 \rightarrow 35 \rightarrow 70$ .



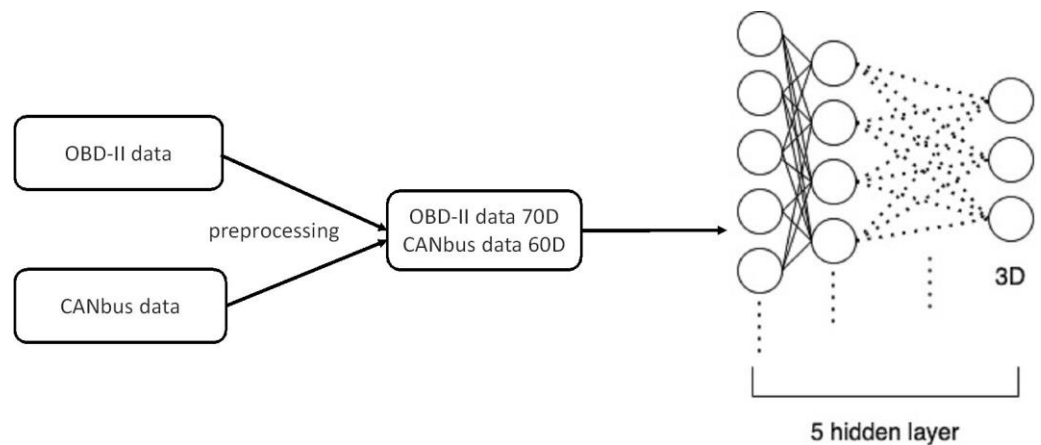


Figure 3. A flowchart and network structure for driving behavior analysis.



Figure 4. An example of driving behavior visualized on the OSM for the network structure  $70 \rightarrow 35 \rightarrow 17 \rightarrow 8 \rightarrow 3 \rightarrow 8 \rightarrow 17 \rightarrow 35 \rightarrow 70$ .

In addition, we used the *K*-means clustering algorithm to further classify the driving behavior. The elbow method was used to determine the most appropriate *k* value to classify driving behavior according to aggressiveness [27]. Driving behavior is classified into four levels, from normal to aggressive, and the most aggressive driving behavior is marked on the OSM.

#### 4.3 Negative Binomial Regression

Referring to [16], we used a negative binomial regression model to analyze the road features at intersections and interchanges. It is an extended version of the Poisson regression to process the data overdispersion problem. The negative binomial regression model is used to predict the number of aggressive driving behavior  $\mu_i$ , defined by

$$\mu_i = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i) \quad (2)$$

where  $\beta_i$  is the correlation term associated with each road feature parameter and  $\varepsilon_i$  is an error term. Pearson's chi-squared test was performed [28] to verify whether the data were overdispersed. When the ratio was greater than 1, the data were considered overdispersed.

To evaluate whether Poisson or negative binomial regression could better fit our data, the Akaike information criterion (AIC) was computed for these two models [29]. The AIC is an effective measure of data fitting in regression models and is defined as

$$AIC = 2k - 2\ln(L) \quad (3)$$

where  $k$  is the number of features, and  $\ln(L)$  is the maximum likelihood. A smaller  $AIC$  value implies a better-fitting model.

After classifying driving behavior using  $K$ -means clustering, aggressive driving behavior was found to occur more frequently at interchanges and intersections. Negative binomial regression analysis was performed for these two specific driving scenarios. We adopted the road features proposed by Wong [16] and those commonly appearing in Taiwanese road scenes.

- Interchanges: (1) section length, (2) lane width, (3) speed limit, and (4) traffic flow.
- Four-arm intersection: (1) without lane markings, (2) straight-lane markings, (3) left-lane markings, (4) right-lane markings, (5) shared-lane markings, (6) shared-lane markings at the roadside, (7) motorcycle priority, and (8) branch road.
- Three-arm intersection: (1) without lane markings, (2) straight-lane markings, (3) shared-lane markings at the roadside, (4) lane ratio, (5) motorcycle priority, and (6) branch road.

## 5. Experimental Results

We divided the experiments into two parts: image data extraction of training and testing datasets and the driving behavior analysis based on the driving and road features.

### 5.1 Extraction of Training and Testing Data

We demonstrate the image data extraction for road scenes with traffic lights. Figure 5(a) shows the driving trajectory (marked by the red curve) and traffic light positions (marked by blue circles) on the OSM. The driving videos were filtered through an extraction system to contain traffic lights from far to near. The extracted images in Figures 5(b) and 5(c) correspond to orange dots (a) and (b) in Figure 5(a), respectively.

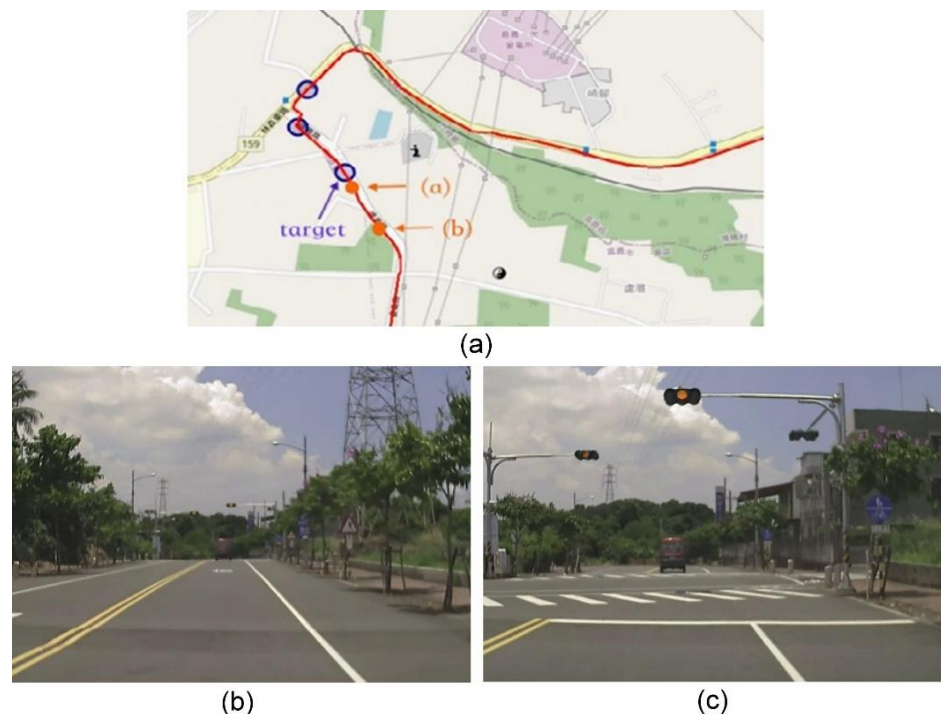


Figure 5. Image data extraction for road scenes with traffic lights. (a) Driving trajectory (marked in red curve) and traffic-light positions (marked in blue circles) on the OSM. (b) Extracted image containing traffic lights (long-range image). (c) Extracted image containing traffic lights (short-range image).

5.2 Driving Behavior Analysis

For the driving behavior analysis, we first presented the visualization and *K*-means classification and then performed an analysis of the driving behavior and road features.

5.2.1 Visualization and *K*-means Classification

We used five segments of driving data in the UAH DriveSet, and the drivers showed normal and aggressive behaviors separately. In each data segment, 50 samples were used for classification. The results are presented in Table 1 with the percentage of correct classifications, where *D1–D5* represent the five drivers. *N* and *A* denote normal and aggressive driving, respectively. The table shows that *K*-means classification can provide satisfactory classification results for normal and aggressive driving behaviors.

Table 1. *K*-means classification performance on the UAH DriveSet.

<i>D1</i>		<i>D2</i>		<i>D3</i>		<i>D4</i>		<i>D5</i>	
<i>N</i>	<i>A</i>	<i>N</i>	<i>A</i>	<i>N</i>	<i>A</i>	<i>N</i>	<i>A</i>	<i>N</i>	<i>A</i>
100%	80%	100%	100%	100%	96%	98%	100%	98%	98%

Figure 6 shows the visualized driving behavior and the corresponding driving data chart. Figure 6(a) shows the visualized driving behavior (including aggressive driving) using the driving data in the UAH DriveSet visualized on the OSM with traffic light location information. Red circles A and B in Figure 6(a) correspond to the driving data chart enclosed by red rectangles A and B, respectively, in Figure 6(b). The driving images corresponding to the red circles A and B in Figure 6(a) are shown in Figures 7(a) and 7(b), respectively. In this example, the aggressive driving behavior at the location indicated by red circle A in Figure 6(a) was due to the sudden braking caused by the car in front of the intersection (Figure 7(a)), and that at the location indicated by red circle B in Figure 6(b) was due to the change in lanes (Figure 7(b)).

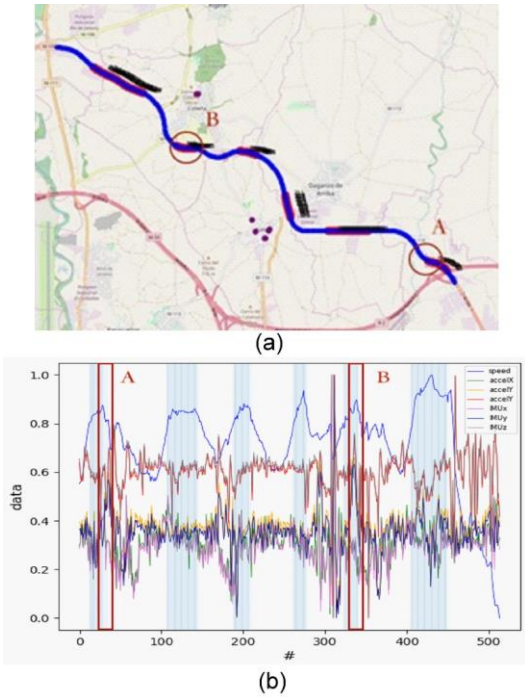


Figure 6. The visualized driving behavior and driving data chart. (a) The aggressive driving data and traffic lights are marked on the OSM. The locations A and B (red circles) in Figure 6(a) corresponded to the data enclosed by red rectangles A and B in Figure 6(b), respectively. (b) The red



rectangles A and B in Figure 6(b) correspond to the locations indicated by the red circles A and B in Figure 6(a), respectively.



(a)



(b)

Figure 7. Driving images. (a) The image acquired at the location corresponding to red circle A in Figure 6(a). (b) The image acquired at the location corresponding to red circle B in Figure 6(a).

By visualizing the driving behavior and displaying aggressive driving behaviors on OSM with reference to the driving video, we can observe a correlation between driving behavior and traffic infrastructure. The three situations were analyzed as follows:

- a. Influence of two-way lanes on driving behavior: The vehicle speed in a two-way lane was higher than that in a one-way lane. Thus, aggressive driving behaviors with fast driving and emergency braking are more likely to occur in two-way lanes.
- b. Influence of traffic lights on driving behavior: The most aggressive driving behavior occurs at intersections. There may be many reasons for this, such as fast-changing signals and poor road design. This generally causes more conflicts between drivers and other vehicles.
- c. Influence of interchanges on driving behavior: In highway traffic, the most aggressive driving behaviors occur at interchanges. A vehicle entering an interchange entrance tends to drive in the inner lane. This generally causes the other drivers to change lanes or slow down.

### 5.2.2 Negative binomial regression

Because aggressive driving behaviors frequently occur near intersections and interchanges, we further investigated these driving scenarios using negative binomial regression analysis of the correlation between the number of aggressive behaviors and road features.

The  $P$ -value can be used to evaluate the statistical significance of the features of aggressive driving behaviors [30]. The following two driving scenarios were examined.

**Four-Arm Intersection:** Eight different road features were defined at the intersections. The regression analysis is shown in Figure 8(a), where Intercept is the error term of the regression model, LEFT is the left-turn lane mark, STRA is the straight lane mark, RIGHT is the right-turn lane mark, TWO is the shared lane mark, SHARE is the shared lane mark on the side of the road, NO is no lane mark, MOTOR is the number of priority locomotive lanes, CROSS is the number of branch roads, and the coefficient term is the parameter of the regression model. The features that have considerable impacts on aggressive driving behaviors included “straight lane marking,” “shared lane marking at roadside,” and “without lane marking.” The influences of these features on the driving behavior were positive, negative, and positive correlations, respectively. When  $P > |z| < 0.05$  held, the feature significantly affected aggressive behavior.

**Highway Interchange:** Four different road features were defined for highways. The regression analysis results are shown in Figure 8(b), where LONG is the length of the interchange, LANE is the lane width, LIMIT is the ramp speed limit, and FLOW is the average daily traffic volume. The features that had a considerable impact on aggressive driving behaviors were “speed limit” and “average daily traffic volume.” The influences of these features on driving behavior showed positive and negative correlations, respectively.

Generalized Linear Model Regression Results						
=====						
Dep. Variable:	AGGRE	No. Observations:	21			
Model:	GLM	Df Residuals:	13			
Model Family:	NegativeBinomial	Df Model:	7			
Link Function:	log	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-43.927			
Date:	Sat, 02 May 2020	Deviance:	1.1128			
Time:	07:55:51	Pearson chi2:	1.12			
No. Iterations:	5					
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	1.9725	0.367	5.379	0.000	1.254	2.691
LEFT	-0.0342	0.099	-0.344	0.731	-0.229	0.160
STRA	0.0521	0.020	2.632	0.008	0.013	0.091
RIGHT	-0.0156	0.279	-0.056	0.955	-0.562	0.531
TWO	0.0114	0.102	0.111	0.912	-0.189	0.212
SHARE	-0.1328	0.062	-2.130	0.033	-0.255	-0.011
NO	0.0521	0.020	2.632	0.008	0.013	0.091
MOTOR	0.0826	0.197	0.419	0.675	-0.304	0.469
CROSS	-0.1153	0.144	-0.802	0.423	-0.397	0.167
=====						

(a)

Generalized Linear Model Regression Results						
=====						
Dep. Variable:	AGGRE	No. Observations:	8			
Model:	GLM	Df Residuals:	3			
Model Family:	NegativeBinomial	Df Model:	4			
Link Function:	log	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-23.381			
Date:	Mon, 20 Apr 2020	Deviance:	9.0832			
Time:	06:59:31	Pearson chi2:	8.88			
No. Iterations:	5					
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	18.9674	9.660	1.963	0.050	0.033	37.902
LONG	0.0010	0.000	2.487	0.013	0.000	0.002
LANE	-5.8833	2.390	-2.462	0.014	-10.567	-1.200
LIMIT	0.0967	0.034	2.818	0.005	0.029	0.164
FLOW	-0.0002	6.06e-05	-2.567	0.010	-0.000	-3.68e-05

(b)

Figure 8. Negative binomial regression analysis results. (a) at a four-arm intersection. (b) at a highway interchange.

## 6. Conclusions

We present an image data extraction system based on geographic information, and a driving behavior analysis approach that uses various types of driving data. The experimental results show that lane ratios without lane markings and with straight lane markings are important features that affect aggressive driving behaviors. Finally, traffic improvements are proposed based on the analysis of a case study at an intersection. In the future, we will add more driving data for a more accurate analysis of driving behavior.

**Author Contributions:** Methodology, J.-Z.Z. and H.-Y.L.; Supervision, H.-Y.L. and C.-C.C.; Writing—original draft, J.-Z.Z.; Writing—review & editing, H.-Y.L. and C.-C.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially/partially supported by Create Electronic Optical Co., LTD, Taiwan.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dong, B.T.; Lin, H.Y. An on-board monitoring system for driving fatigue and distraction detection. Proceedings of 2021 IEEE International Conference on Industrial Technology (ICIT 2021), Valencia, Spain, 10–12 March 2021.
2. Wawage, P.; Deshpande, Y. Smartphone sensor dataset for driver behavior analysis, *Data in Brief* **2022**, *41*, 107992.
3. Zhang, J.Z.; Lin, H.Y. Driving behavior analysis and traffic improvement using onboard sensor data and geographic information. The 7th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2021), Prague, Czech, April 28–30, 2021; pp. 284–291.
4. Hartford, J.S.; Wright, J.R.; Leyton-Brown, K. Deep learning for predicting human strategic behavior. Proceedings of 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 5–10 December 2016.
5. Malekian, R.; Moloisane, N.R.; Nair, L.; Maharaj, B.; Chude-Okonkwo, U.A. Design and implementation of a wireless OBD II fleet management system. *IEEE Sensors Journal* **2014**, *13*, 1154–1164.
6. Hornauer, S.; Yellapragada, B.; Ranjbar, A.; Yu, S. Driving scene retrieval by example from large-scale data. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019; pp. 25–28.
7. Wu, Z.; Xiong, Y.; Yu, S.X.; Lin, D. Unsupervised feature learning via non-parametric instance discrimination. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3733–3742.
8. Wu, Z.; Efros, A.A.; Yu, S.X. Improving generalization via scalable neighborhood component analysis. Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 685–701.
9. Naito, M.; Miyajima, C.; Nishino, T.; Kitaoka, N.; Takeda, K. A browsing and retrieval system for driving data. Proceedings of 2010 IEEE Intelligent Vehicles Symposium (IVS), La Jolla, CA, USA, 21–24 June 2010; pp. 1159–1165.
10. Lin, H.Y.; Dai, J.M.; Wu, L.T.; Chen, L.Q. A vision based driver assistance system with forward collision and overtaking detection. *Sensors* **2020**, *20*(18), 100–109.
11. Liu, H.; Taniguchi, T.; Takano, T.; Tanaka, Y.; Takenaka, K.; Bando, T. Visualization of driving behavior using deep sparse autoencoder. Proceedings of 2014 IEEE Intelligent Vehicles Symposium (IVS), Dearborn, MI, USA, 08–11 June 2014; pp. 1427–1434.
12. Liu, H.; Taniguchi, T.; Tanaka, Y.; Takenaka, K.; Bando, T. Visualization of driving behavior based on hidden feature extraction by using deep learning. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *18*(9), 2477–2489.
13. Constantinescu, Z.; Marinoiu, C.; Vladiu, M. Driving style analysis using data mining techniques. *International Journal of Computers and Communications and Control* **2010**, *5*(5), 654–663.
14. Kharrazi, S.; Frisk, E.; Nielsen, L. Driving behavior categorization and models for generation of mission-based driving cycles. Proceedings of IEEE Intelligent Transportation Systems Conference 2019, Auckland, New Zealand, 27–30 October 2019; pp. 1349–1354.
15. Tay, R.; Rifaat, S.M.; Chin, H.C. A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes. *Accident; analysis and prevention* **2008**, *40*(4):1330–6.
16. Wong, C.K. Designs for safer signal-controlled intersections by statistical analysis of accident data at accident blacksites. *IEEE Access* **2019**, *7*, 111302–111314.

17. Schorr, J.; Hamdar, S.H.; Silverstein, C. Measuring the safety impact of road infrastructure systems on driver behavior: Vehicle instrumentation and exploratory analysis. *Journal of Intelligent Transportation Systems* **2016**, *21*(5), 364–374.
18. Abojaradeh, M.; Jew, B.; Al-Ababsah, H. The effect of driver behavior mistakes on traffic safety. *Civil and Environment Research* **2014**, *6*(1), 39–54.
19. Chunhui, Y.; Wanjiang, M.; Ke, H.; Xiaoguang, Y. Optimization of vehicle and pedestrian signals at isolated intersections. *Transportation Research Part B: Methodological* **2017**, *98*, 135–153.
20. Wang, Y.; Qian, C.; Liu, D.; Hua, J. Research on pedestrian traffic safety improvement methods at typical intersection. Proceedings of 2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT), Guilin, China, 26–28 April 2019; pp. 190–193.
21. Ma, W.; Liu, Y.; Zhao, J.; Wu, N. Increasing the capacity of signalized intersections with left-turn waiting areas. *Transportation Research Part A: Policy and Practice* **2017**, *105*, 181–196.
22. Anjana, S.; Anjaneyulu, M. Safety analysis of urban signalized intersections under mixed traffic. *Journal of Safety Research* **2015**, *52*, 9–14.
23. Yeh, T.W.; Lin, S.Y.; Lin, H.Y.; Chan, S.W.; Lin, C.T.; Lin, Y.Y. Traffic light detection using convolutional neural networks and lidar data. Proceedings of 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Taipei, Taiwan, 03–06 December 2019; pp. 1–2.
24. Binas, J.; Neil, D.; Liu, S.C.; Delbruck, T. Ddd17: End-to-end davis driving dataset. Proceedings of 2017 Workshop on Machine Learning for Autonomous Vehicles, Sydney, Australia, 6–11 August 2017; pp. 1–9.
25. Romera, E.; Bergasa, L.M.; Arroyo, R. Need data for driver behaviour analysis? Presenting the public uah-driveset. Proceedings of 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 01–04 November 2016; pp. 387–392.
26. Bergasa, L.M.; Almería, D.; Almazán, J.; Yebes, J.J.; Arroyo, R. Drivesafe: an app for alerting inattentive drivers and scoring driving behaviors. Proceedings of IEEE Intelligent Vehicles Symposium (IVS), Dearborn, MI, USA, 08–11 June 2014; pp. 240–245.
27. Thorndike, R.L. Who belongs in the family? *Psychometrika* **1953**, *18*(4), 267–276.
28. Pearson, K. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **1900**, *50*(302), 157–175.
29. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **1974**, *19*(6), 716–723.
30. Dahiru, T. *P*-value, a true test of statistical significance? A cautionary note. *Annals of Ibadan Postgraduate Medicine* **2008**, *6*, 21–26.