JST: Smart Systems and Devices
Volume 33, Issue 1, January 2023, 043-053

Deep Learning Based Vehicle Speed Estimation on Highways

Nguyen Thi Thu Hien[1], Tran Thi Hien[2], Le Dinh Chung[3], Tien Dzung Nguyen[1*]
[1]School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Ha Noi, Vietnam
[2]Nam Dinh University of Technology Education, Nam Dinh, Vietnam
[3]Department of Military Medicine, General Department of Logistics,
Ministry of National Defence, Ha Noi, Vietnam
[*]Corresponding author email: dzung.nguyentien@hust.edu.vn

Abstract

Traffic management always is a matter requiring the attention of highway system managers in terms of vehicle monitoring and speed estimation. This paper proposes an efficient deep learning-based vehicle speed estimation on highway lanes in the Vietnam transport system. The input videos are recorded by fixed surveillance cameras. An optimized single shot multibox detector network, called SSD, is utilized for vehicle license plate detection (LPD). The deep SORT (simple online and real-time tracking) model is first applied to video vehicle tracking and performed in the detected license plate area. This tracking process investigates the traveling distance of a vehicle to estimate its speed. In this study, the dataset has been normalized to improve the efficiency of vehicle localization and tracking to improve the time elapsing in the estimation of the distance traveled by vehicles on highways. The results showed that the proposed system has achieved better accuracy in terms of the determined speeds with errors ranging between [-1.5, +1.1] km/h, equivalent to 98% of the error limit by the regulation in Viet Nam.

Keywords: Vehicle detection, speed estimation, feature extraction, LPD-SSD network, SORT model

1. Introduction

The rapid development of intelligent transportation systems (ITS) has proven an important role in intelligent traffic monitoring, management, dynamic information, and vehicle control services. Various methods have been introduced for monitoring vehicles on the road through domestic and international studies. The proposed methods of measuring vehicle speed using improved image processing technologies have significant speed errors for vehicles moving at high speeds.

Video-based vehicle speed measurement has been recently receiving more and more research interest because it is suitable for stealth speed measurement and has a low cost. It can also provide vehicle speed determination and identification information in the same video at the same time. While the methods mentioned are video-based vehicle speed estimation, the method falls into feature-based ones for identification of vehicles from video frames by their visual characteristics. There is a motion-based one to locate vehicles by an optical flow. The common point of these methods is to calculate traveling distance of a vehicle detected from the surveillance videos, and thus estimate the average speed of the vehicle.

The speed is estimated in 3D reconstruction and visualization for transportation detection. There is a method that can solve the speed detection of simultaneously moving vehicles with the lowest root mean square error (RMSE). However, it uses perspective projection and requires vehicles to move in a straight line, at a constant speed, and parallel to the lane under testing. Vehicle speed determination investigates the captured stereo video frames, where feature extraction is extracted to identify the media for estimation of the depth map and then measures vehicle speed. The superior advantage of this method is to handle speed estimation problems when vehicles move in curvilinear structured environment. However, stereo image tracking is accomplished through particle filtering, which relies on manual setting of the vehicle's features. One-time manual setup is required for a feature, meanwhile, various vehicles may have different features leading to more complex cases. Therefore, multiple manual settings are adopted if some vehicles appear in given image frames for speed measure. Several methods are focusing on motion-based speed detection that subtracts a static background from a video frame for media detection, depth calculation, and speed measurement. However, these approaches face to difficulty in distinguishing vehicles and other moving objects detected in a video sequence. The study in [1] proposes an intelligent system with integrated vehicle tracking based on stereo motion estimation to achieve accurate speed determination. However, this tracking uses speedup robust features (SURF) in [2] to perform license plate tracking in consecutive frames. In fact, up to this point the deep SORT [3] tracking model demonstrates higher efficiency than that of SURF in terms of speed detection.

ISSN: 2734-9373
https://doi.org/10.51316/jst.163.ssad.2023.33.1.6
Received: June 29, 2022; accepted: September 28, 2022

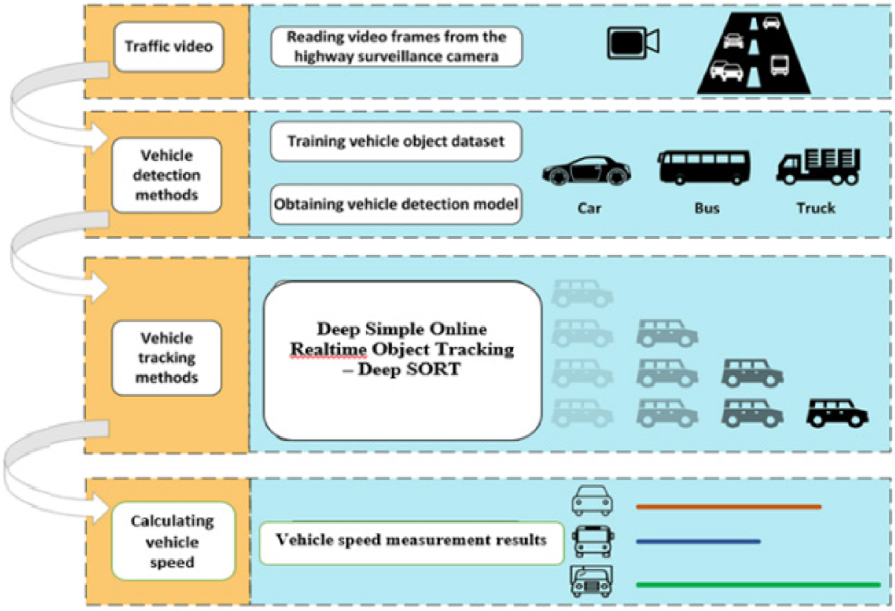

Fig. 1. Overall system of the proposed deep learning-based vehicle speed estimation

The study in this paper proposes a speed estimation approach utilizing a deep learning algorithm as demonstrated in Fig. 1. The input video is captured by a surveillance camera. The LPD - SSD license plate detection network is recommended for synchronous extraction of typical features from the localized license plate, and the vehicle tracking in the video using the SORT network is performed only in the license plate area.

2. Related Works

Most recently, A. S. Gunawan *et al.* measured vehicles' speed using a single camera [4]. In their proposed algorithm the road image was mapped using direct linear transformation (DLT) method. A vehicle passed a straight line which was limited by four points and the camera measured the movement of the vehicle. The positions of the four points were used to map their data from the image plane to the road plane. Finally, after calculating the vehicle displacement at the ground plane and the time interval between the frames, they calculated the vehicle speed. The reported accuracy is 96.14%. J. Sochor *et al.* also used a single camera for speed measurement [5]. Vehicle tracking was performed using optical flow. They considered two reference start and end lines on the road plane and count the number of video frames in which the vehicle was present between these reference lines. The measured speed has an average error of 0.63 km/h with a standard deviation of 4.5 km/h. M. G. Moazzam *et al.* measured the speed of vehicles by calculating the number of frames it takes for the vehicle to travel the distance between two specified lines on the road plane [6]. The calculation average error is 10%. A new method was presented by A. Tourani *et al.* based on measuring the vehicle displacement vector on the image plane in terms of pixels per frame [7]. They converted the velocity vector into km/h with two distinct pixels-to-meters and frames-to-seconds transfer functions. Their reported speed measurement accuracy is 94.8%.

Commonly, vehicle speed estimation methods are classified into two main categories: binocular stereo camera-based and monocular single camera-based.

Accurate vehicle feature detection in video frames is a prerequisite for video-based vehicle speed measurement. License plate has a regular appearance, uniform contour, and relatively rich texture details and can thus be easily detected. License plate is also unique and suitable for the synchronization of speed measurement and information identification in the future. Many existing traffic surveillance systems record license plates of vehicles in traffic violations, and the infrastructure is already available. Although the license plate is not attached to the ground plane, the 3D spatial position of the license plate center can be directly calculated using our binocular stereovision system. The relative 3D displacements can be obtained. Accordingly, accurate vehicle speed can be derived. Therefore, we choose license plate as the object to be detected in our system. Traditional video object detection methods include background modeling, frame difference, and optical flow. The two former methods are based on video processing. They are suitable for detecting moving objects with fixed background. However, license plate is always a part of a moving vehicle and cannot be detected separately. Meanwhile, the latter method can track objects with fixed or moving background. It can also track license plate as a separate part but cannot detect license plate without manual help. Morphological methods are widely used in image object detection by relying on color, edge, shape, and texture attributes extracted with the Canny detector, Sobel operator, template-matching conditional random field, or wavelet. An edge operator is used to extract the vertical edge of license plate. Edge density information is utilized to detect license plate. A rectangular sliding window is adopted to detect image regions of high gradient density, such as license plate. All these morphological methods are cumbersome, time consuming, and unsuitable for LPD in complex background.

Modern object detection methods have developed with the progress of convolutional neural networks (CNNs) [8]. Since AlexNet won ILSVRC 2012 by a notable margin 2012, object detection methods with CNN have received considerable attention. CNN-based object detection methods can detect various objects accurately and intelligently with the trained networks and models. The R-CNN series (e.g, RCNN, Fast-RCNN, and Faster-RCNN) algorithms are typical examples of two-stage algorithms. They have high accuracy but need long computation time. You-only-look-once (YOLO) and SSD are typical examples of one-stage algorithms that perform faster than the two-stage ones. For example, the detection speed of SSD [9] in VOC2007 dataset can reach 59 frames per second (FPS), whereas that of Faster R-CNN can only reach 7 FPS. All these CNN-based object detection methods are suitable for LPD. A CNN network is trained and fine-tuned to detect license plate. A CNN-based MD-YOLO framework for multidirectional LPD is proposed. Region-based convolutional neural network is trained to detect license plate.

The next part of the recommended system is vehicle tracking. Deep SORT was developed by Nicolai Woke and Alex Bewley [3] shortly after SORT [10] to address the shortcoming problems associated with a high number of ID switches. The solution proposed by deep SORT is based on using deep learning to extract features of objects to increase accuracy in the data association process. In addition, a linking strategy was also built called Matching Cascade to help link objects after disappearing for a while more effectively.

3. Proposed Tachometer System

In the proposed vehicle tachometer system, the input monitoring video source is obtained from the industrial surveillance camera system XNP-6550RH/VAP located in Phap Van - Cau Gie highway with the following parameters: Sony ICX445 CCD camera, 1/3 inch, 1288x964 maxi-mum resolution, effective 1.25 Mpixels, 32 MB onboard buffer, and 512 KB data flash memory for data acquisition Whether. The local server system is equipped with 6-core Intel E5-2620 v3 CPU @ 2.40 GHz, 32 GB RAM, Nvidia Geforce GTX 1080 8 GB standalone graphics card, and 1T solid-state hard drive for data storage and a laptop, with Core i7 CPU, 8 GB RAM, Nvidia Geforce 830 M 2GB discrete graphics card.

3.1. Vehicle Feature Extraction

The proposed system is implemented by utilization of the algorithms mentioned above, e.g vehicle feature detection and vehicle tracking, and in combination with vehicle speed estimation. In vehicle feature detection, an LPD -SSD model is trained based on an optimized SSD network. The trained model is then used to detect all license plates in the captured video. For vehicle tracking, the detected license plate areas in consecutive frames are matched to investigate tracking independency between the license plates under consideration. Finally, the license plate in the frame pairs is matched to extract all matching-point pairs and utilized for estimation of the vehicle speed. In the tachometer, the ratio between the distance in meters and the distance in pixels has been pre-determined. The points closest to the center of the license plate are then chosen as the exact coordinates of the current frame pairs, the distance the vehicles have traveled in a given period (i.e. a frame interval or several frame intervals) is determined and then vehicle speed is estimated from the detected distance over time.

Fig. 2 describes the processing steps in vehicle license plate detection using LPD-SSD.

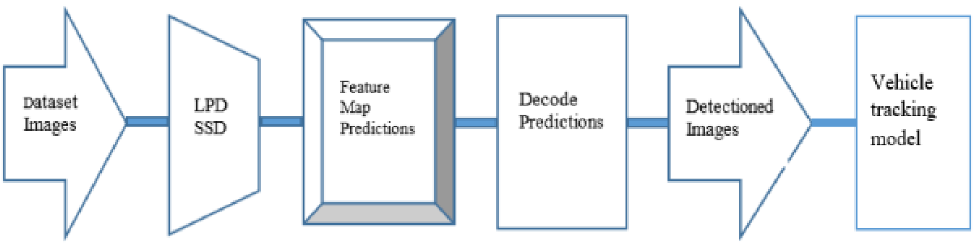

Fig. 2. Flowchart of vehicle license plate detection.

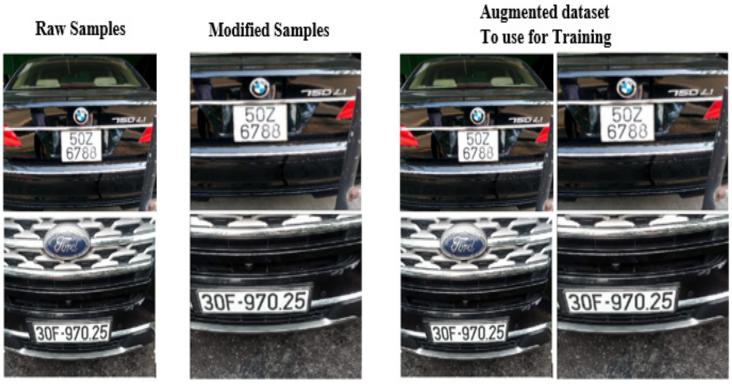

Fig. 3. Dataset before and after data augmentation

3.1.1. Data augmentation

This is a technique applied in dataset setup and normalization processes to enhance data quality. Currently, in deep learning, data augmentation plays a very important role. Since the data set has little or lacks diversity, it will be difficult to train the model to produce a good model used for future prediction. As illustrated in Fig. 3, a data augmentation technique is adopted to diversify and enrich the datasets for improvement in prediction efficiency through the trained model. The data augmentation helps fix the "not enough data" problem, prevents overfitting, and moreover makes the model perform better on never-before-seen models. In addition, no additional effort is required in data collection or data labeling, which may require a higher cost or impractical solution.

The following image and signal processing algorithms are required to increase dataset diversity:

- Image cropping, rotation, and flipping;
- Blur and sharpen filtering;
- Affine transformation: to preserve parallel lines in an image;
- Noise addition or removal: such as salt and pepper noise, Gaussian noise;
- Color shifting: lighting or contrast adjustment.

Fig. 4 illustrates the diversifying process as mentioned above to set up datasets, where the cropped images containing license plate area have been resized to a given original size and followed by brightness adjustment stage and then rotated to the horizontal direction.

3.1.2. Prepared data set

Fig. 5 illustrates examples of dataset that includes:

- 2500 images (4032x2268 resolution) with high resolution and image quality captured by the Nikon d3200 SLR in parking lots and roads street at 5 - 30 m distance;
- 3000 images (1600x1200 resolution) taken from the traffic data set supported by the Voice of Vietnam Radio for research students;
- 2500 images (1920x1080 resolution) extracted from the campus security monitoring system of apartment buildings in Hanoi city, in which the vehicles can be far away from the camera and the number plate size is very small.

After cropping and resizing random images picked from the total 8000 images above in the augmentation, the dataset size has been expanded by a factor of three, i.e., the augmented dataset finally includes 24000 images.

a) b) c) d) e)

Fig. 4. Example of calibration results of the augmented data. a) Original image. b) Resized image. c) Cropped image. d) Brightness adjusted image. e) Rotated image

Fig. 5. Examples of the setup dataset

3.2. Application of LPD- SSD Based Model

To detect vehicles from the surveillance video frames, vehicle features are first extracted by using LPD - SSD as the network structure [1] as shown in Fig. 6.

The LPD-SSD based model is trained with half of the augmented dataset which consists of 24000 images captured from different angles. The other half is used to test the trained model and its performance is recorded and assessed afterward. Another 600 license plate images that are independent of the augmented dataset are used to validate the final trained model as demonstrated in Fig. 7.

After 20000 training iterations, the performance of the trained model has been recorded and shown in Table 1, where the first step of LPD-SSD studied in [1] (LPD-SSD^{1st}) demonstrates superior accuracy of 98.2% on the augmented dataset compared to that without augmentation. The real-time processing speed of the two models is equal to 5.4 FPS. As the processing speed increases, the consumption power of the CPU and GPU in this system also increases correspondingly.

Table 1. The accuracy of LPD-SSD network applied to the original and the augmented datasets respectively

Network	Accuracy	FPS
LPD-SSD^{1st} [1] without augmentation	97.8%	5.4
LPD-SSD with augmented dataset	98.2%	5.4

Since the location of the mounted surveillance camera and its distance to the vehicles moving on highways are incredibly important parameters, the proper setup of LPD - SSD network topology may help adapt this system to any transportation environment with allowed speed under 80 km/h which is typical in Vietnamese transportation.

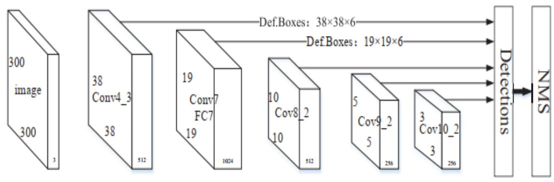

Fig. 6. LPD-SSD network architecture

a) License plate detected at 30 m distance

b) License plate detected at 5 m distance

Fig 7. License plates from the surveillance video stream on Phap Van - Cau Gie highway are detected by LPD-SSD model (green) on the augmented datasets captured at ranges between 5m-30m in two-way lanes

3.3. Vehicle Tracking

Fig. 8 depicts sequential steps in the vehicle tracking process using Deep SORT [3] and Fig. 9 introduces its state model accordingly.

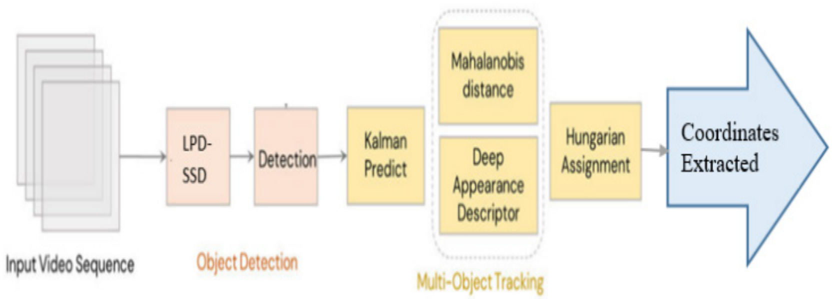

Fig. 8. Vehicle tracking using Deep SORT

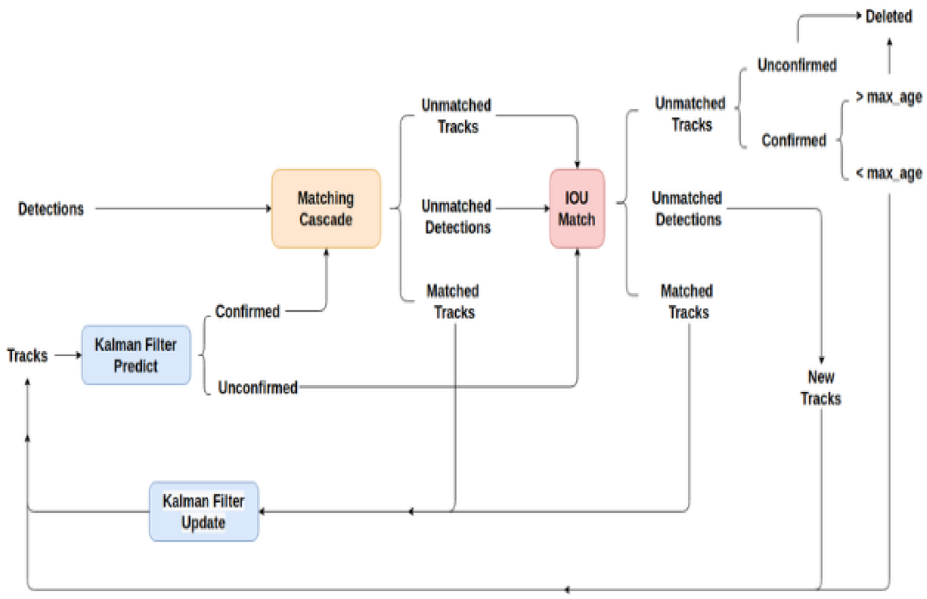

Fig. 9. State model of the Deep SORT

These steps are described in detail below:

- **Step 1**: Use LPD-SSD to detect objects in the current frame.
- **Step 2**: The Deep SORT utilizes the Kalman filter to predict new track states based on past tracks. These states are initially assigned tentative values. If this value is still guaranteed to be maintained for the next 3 frames, the state will change from poll to confirmed and try to stay tracked for the next 30 frames. Since the frame rate of the surveillance camera applied in this problem is 30 fps with speed ranges from 50 km/h - 90 km/h or 13.88 m/s - 25 m/s, where possible detection range is about 25-30 m distance, the utilized number of frames containing vehicles still can guarantee no media distortion. Conversely, if the tracking is lost in less than 3 frames, the status will be removed from the tracker.
- **Step 3**: Using the verified tracks, introduce a matching cascade to associate all candidate detections based on distance and feature metrics such as the shape or the combined background colors for a given license plates.
- **Step 4**: Unlinked tracks and detections are passed to the next filter layer. Use the Hungarian algorithm to resolve the assignment problem with the IOU cost matrix for the next 2nd association.
- **Step 5**: Classification of the candidate detections and tracks.
- **Step 6**: Use the Kalman filter to recalibrate the track value from the detections associated with the track and initialize new tracks.

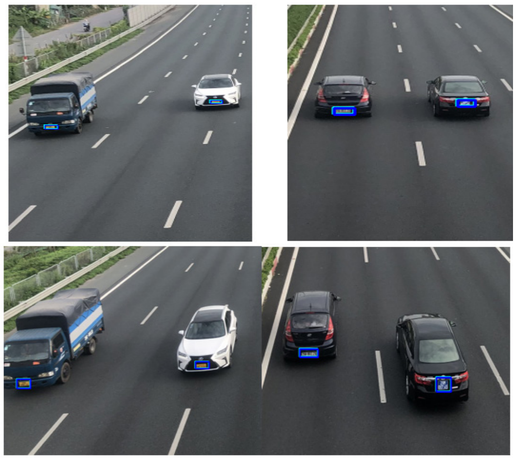

Fig. 10. License plates tracked by Deep SORT

Fig. 10 demonstrates the tracked license plates from the video stream performed by both LPD-SSD model (light blue) bounding box (BB) and Deep SORT model (dark blue) for combination.

3.4. Speed Estimation

3.4.1. Geometric model of camera in distance estimation problem

The essence of a camera is its lens system [11]. Fig 10 demonstrates the position of an object according to the incident ray passing through O which is the center of the lens. The image obtained on the screen reflects the object thanks to the optical phenomena in the lens which is real and inverted to the object. The closer the object is to the viewfinder, the larger size it has. This viewfinder receives lighting source passing through lens system of the camera, and then an image reflecting the object will be created on the screen at the focal length *f*. The pixel position *(i, j)* of a reflected point from the object into the image plan can be evaluated from the object's coordinates in 3D space as in [12]:

$$i = f\frac{X}{Z} \text{ and } j = f\frac{Y}{Z} \quad (1)$$

where, *X, Y*, and *Z* are the pixel coordinates in the OXYZ space; *f* is the focal length of the camera lens system. The camera location must be set up over the surface of the road with its optical axis inclined downward to the roadway to cover the road plane. Since the proposed solution focuses on speed detection of vehicles on both highways and inner-city streets with a mixed traffic flow of motorcycles, cars, vans, and other means, there is the need to identify each type of means out of each other in multiple road lanes. Therefore, this paper utilizes the direction angle (DA) of the first primary axis (FPA) for each coming vehicle detected in the video sequence and captured by the surveillance camera mounted on the roads [13]. The background subtraction is first performed for the captured frame sequence and then each vehicle is located and the DA of FPA is evaluated for decision making in the identification task [10]. The study in [13] modelled the location of the surveillance camera installation on the road as shown in Fig. 11 below.

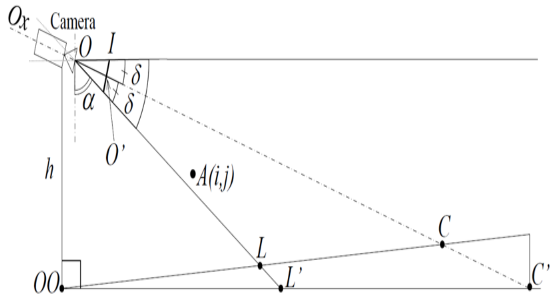

Fig. 11. A model of a surveillance camera calibration mounted on roads

Camera calibration is one of the important aspects of the study. The vehicle's location in video images is 2D, however, the vehicles in real world are 3D since vehicles cannot leave the road surface. Therefore, vehicle motion is also in 2D which allows the formulation of the vehicle's coordinate transformation in 2D-to-2D mapping. In this section, the calculation of the pattern function between vehicle coordinates in the image and real-world coordinates is performed. How the video camera is installed when the video images are captured from the road traffic and what characteristics are involved in that must be determined. As shown in Fig. 11, the camera is set at the height of h above the road surface with its optical axis sloped at an angle δ from the road. The relation between the camera lens angle and the view domain covered by a camera can be determined using frequent geometrical equations. In short, the camera calibration parameters are set as follows:

1) Camera elevation angle range is α towards the vertical axis and $\alpha + \delta$ is the maximum adjustable elevation angle of this camera;
2) An obtained image *I* at resolution is $m \times n$ pixels;
3) The camera is located at the height h evaluated from the road surface;

If the camera's viewing angle along the Ox-axis direction intersects the road surface at point C, which is always set up as the center of the captured image since point C corresponds to the center of the camera's image sensor. Point L is the closest position on the road where the camera can capture. Considering those parameters in this model setup, the distance from the camera to the object can be essentially evaluated.

If O' is assumed to be the center of the camera's image sensor, the angle $\widehat{O'OA}$ between OO' and a given pixel A(*i,j*) on the image plan I of size $m \times n$ may be determined as:

$$\widehat{O'OA} = tan\frac{\sqrt{\left|i-\frac{m}{2}\right|^2+\left|j-\frac{n}{2}\right|^2}}{f} \quad (2)$$

Utilizing the properties of circles, rectangles, and trigonometry, all pixels in the image plan I will be determined corresponding to the angle range scanned by the camera.

Since this problem focuses only on the evaluation of vehicle speed, a vehicle is assumed to move in a straight direction, and therefore only the pixels along the horizontal frame boundary are the target under consideration to determine the movement distance of the vehicle followed by estimation of the vehicle speed. Let denote Δp the size of one pixel on the image sensor, its representation can be written as:

$$\Delta p = \frac{f \tan\delta}{\frac{m}{2}} \quad (3)$$

From the known angles α and δ, the position of pixel *A(i,j)* on the resulting image I at the distance from the center of OO' can be found as follows:

If pixel *A(i, j)* **for** $i \geq \frac{m}{2}$ then:

$$d = h.tan\left(\alpha + \left(\delta - arctan\left(\frac{\left(i-\frac{m}{2}\right)\Delta p}{f}\right)\right)\right) \quad (4)$$

And **if** pixel *A (i, j)* **for** $i < \frac{m}{2}$ then:

$$d = h.tan\left(\alpha + \left(\delta + arctan\left(\frac{i\Delta p}{f}\right)\right)\right) \quad (5)$$

Vehicle position in each frame is first determined by the coordinates of the center in each detected BB.

3.4.2. Frame in the video

From the frame map shown in Fig 12, the time between the two consecutive frames is 1/30 second. When vehicles move toward and get closer to the camera or move away from the camera into the flow of traffic, a pair of consecutive or non-contiguous frames may be utilized with the frame displacement in time denoted by *Δt*. Therefore, the calculation of the average velocity *v* of the direction is simply expressed by the following formula:

$$v = \frac{\Delta s}{\Delta t} \quad (6)$$

$$\text{where } \Delta s = |d_1 - d_2| \quad (7)$$

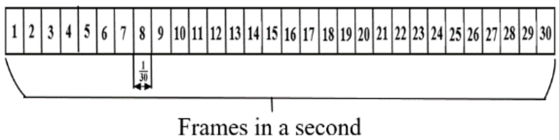

Fig. 12. Representation of the number of frames per second (at 30 fps).

Fig. 13. Demonstration of detected vehicle speed on highways

Fig. 13 demonstrates the estimated speed of all vehicles that appeared in the surveillance camera on Phap Van - Cau Gie highway.

4. Experimental Results

The original dataset is set up with 24000 image frames extracted from the traffic surveillance videos that are monitored by the camera system mounted on Phap Van - Cau Gie highway in Vietnam and authorized by Ha Noi City Public Transport Management Center under Ha Noi Department of Transport.

4.1. Distance Measure and Estimation

4.1.1. Measure the rating

Detect license plates: Used terms of summoning (Recall - R_e) and accuracy (Precision - P_r) and Jaccard Index coefficient (JI) [11] are defined in formulas (8), (9), and (10) to hit license plate detection performance price as following:

$$R_e = \frac{tp}{tp+fn} \quad (8)$$

$$P_r = \frac{tp}{tp+fp} \quad (9)$$

$$JI = \frac{area\ (B_p \cap B_{gt})}{area\ (B_p \cup B_{gt})} \quad (10)$$

where *tp* is called a true finding if JI ≥ 0.5, and *JI* coefficient is calculated by the ratio between the upper intersection union of the detected rectangle by the B_p algorithm and the rectangular region containing manually defined objects B_{gt}, *fp* is a bad development if as *JI* < 0.5 and *fn* are not detected for status. Table 2 shows the precision and recall of vehicle license plates detection for different distances with the proposed method.

Table 2. Precision and recall for distances with the proposed method

Distance (m)	**Recall** (%)	**Precision** (%)
5	99.7	99.1
15	98.9	98.6
30	98.3	98.2

4.1.2. Measure distance estimation error

To evaluate the distance estimation error, the actual distance from the camera to the vehicle will be determined manually, and then compared with the estimated one by marking the position distance on the road surface. The vehicle distance estimation is performed using degrees standard error (RMSE) as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\theta_i - \hat{\theta}\right)^2} \quad (11)$$

where $\hat{\theta}$ is the field-measured distance to vehicle; θ is the predicted distance according to the camera model.

Table 3 shows the evaluation results of vehicle distance estimation error. Based on the absolute base value of the RMSE, the distances of the points with large deviations should be eliminated to improve the accuracy of running track estimation. In this study, the points with an absolute value of RMSE greater than 1 will be removed. The ground coordinates of the point closest to center of the detected license plate center are selected as the exact spatial location of the vehicle of interest in the current video frame pair.

Table 3. Evaluation results of vehicle distance estimation error

Point	Δs (m)	RMSE (m)
1	0.712	0.43
2	0.715	0.52
3	0.725	0.61
4	0.740	0.38
5	0.746	0.49
6	0.753	0.6
7	0.773	0.48
8	0.783	0.55
9	0.794	0.58
10	0.813	0.41

4.2. Results and Performance Evaluation

To evaluate performance of the proposed system, the vehicle speed data respectively estimated by the proposed system, measured by GPS system mounted on the vehicle, and the odometer are recorded and compared. The GPS data is collected using a speedometer for iPhone. This speedometer software is integrated with GPS speedometer using data of GPS + GLONASS dual satellite navigation system, installed for iPhone 12 pro max iOS version 15.3.1 to measure speed and display results on the phone screen in real-time.

On base cars and cars, in particular, the tachometer system will measure the vehicle's revving speed with a transit speed sensor located in the tailgate. As a rule, the speedometer of a car can never display a speed lower or higher than 10% of the actual speed. The satellite navigation system is used to calculate the vehicle's speed by measuring the distance travelled and the time according to the information obtained from the GPS (In fact, the satellite speed always has an error of less than 2% compared to the actual speed). Typically satellite speedometers may monitor the actual vehicle speed which is much more accurate than that using odometer.

The system setup in the real scene to record vehicle speed is shown in Fig. 14, and the performance is introduced in Table 4 and corresponding charts in Fig. 15 in terms of speed data and the corresponding errors.

Table 4. Vehicle speed data collected by different methods and error comparison

Time step	Speed measured by odometer (km/h)	Speed measured by GPS (km/h)	Measured by proposed method (km/h)	Speed error measured by odometer and proposed (km/h)	Speed error measured by GPS and proposed (km/h)	Speed error rate measured by odometer and proposed (%)	Speed error rate measured by GPS and proposed (%)
1	83.4	76.1	76.9	6.5	0.8	7.79	1.05
2	84.5	76.1	77.2	7.3	1.1	8.64	1.45
3	85	79.2	78.3	6.7	-0.9	7.88	-1.14
4	84.7	79.2	79.9	4.8	0.7	5.67	0.88
5	86	80	80.6	5.4	0.6	6.28	0.75
6	87.4	82.8	81.3	6.1	-1.5	6.98	-1.81
7	90.6	83.1	83.5	7.1	0.4	7.84	0.48
8	91.4	85.2	84.6	6.8	-0.6	7.44	-0.70
9	91.3	84.7	85.7	5.6	1	6.13	1.18
10	94	87.3	87.8	6.2	0.5	6.60	0.57

Table 5. Speed detected and errors with the LPD - SSD^{1st} method

Time step	Speed measured by odometer (km/h)	Speed measured by GPS (km/h)	Speed measured by proposed (km/h)	Speed error between odometer and proposed (km/h)	Speed error between GPS and proposed (km/h)	Speed error rate between odometer and proposed (%)	Speed error rate between GPS and proposed (%)
1	NA	43.6	44.6	NA	1	NA	2.29
2	NA	46.8	45.2	NA	-1.6	NA	-3.42
3	NA	50.2	49.4	NA	-0.8	NA	-1.59
4	NA	49.7	50.5	NA	0.8	NA	1.61
5	NA	48.1	49	NA	0.9	NA	1.87
6	NA	46.5	47.1	NA	0.6	NA	1.29
7	NA	45.4	45.5	NA	0.1	NA	0.22
8	NA	44.3	45.1	NA	0.8	NA	1.81
9	NA	42.1	43.2	NA	1.1	NA	2.51
10	NA	42.3	43	NA	0.7	NA	1.65

Fig. 14. Visual depiction of the setup system to record speed

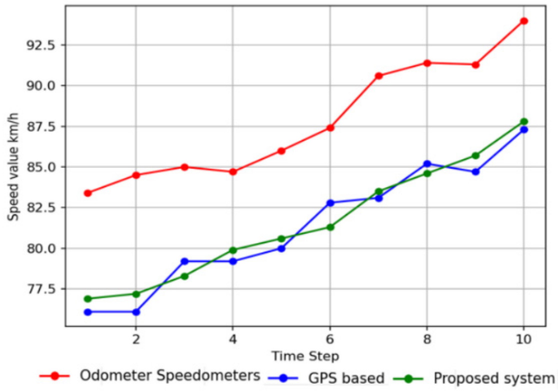

Fig. 15. Comparison of the detected speed by the proposed system versus the estimated Odometer speedometers one and by the GPS

Experiments are performed to compare the results with the LPD - SSD[1st] method. Table 5 is the specific values that LPD - SSD1st has measured and calculated. This value is used for comparison with the proposed method because the range of speed values in this table is the highest that LPD - SSD has determined.

The maximum error and error rate of LPD-SSD[1st] which appear in the time step 2 in Table 5, are -1.6 km/h and -3.42%, correspondingly, the speed measured by GPS is 46.8 km/h, whereas the speed measured by their proposed system is 45.2 km/h. The maximum error and error rate of our proposed system, which appear in the time step 6 in Table 6, are -1.5 km/h and -1.81%, correspondingly. And the speed measured by GPS is 82.8 km/h, whereas the speed measured by our proposed system is 81.3 km/h. Besides in Fig 16, the speed measurement range in our proposed system is 5-30m, while in the LPD - SSD system is 1-15m. The results imply that for a given range of speed between 50-90 km/h, the detected speed by the proposed system has produced errors around [-1.5 - 1.1] in a detectable distance of up to 30m as given in Table 6. That means the detected speed by the proposed system with an accuracy greater than 98% can adapt to transport regulations in Vietnam.

Table 6. Error comparison for vehicle speed measurement

Network	Max speed error measured by odometer and proposed (km/h)	Max speed error measured by GPS and proposed (km/h)	Detectable distance (m)
LPD - SSD[1st] [1]	NA	[-1.6, +1.1]	1 - 15
PROPOSED SYSTEM	[5.67, 8.64]	[-1.5, +1.1]	5 - 30

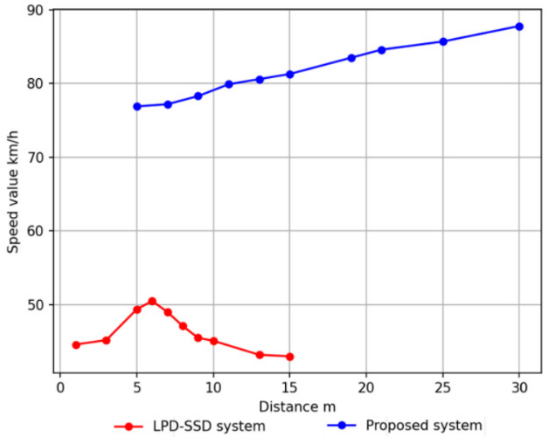

Fig. 16. Compare the speed and distance determined by the proposed method (blue) and the LPD - SSD[1st] method (red)

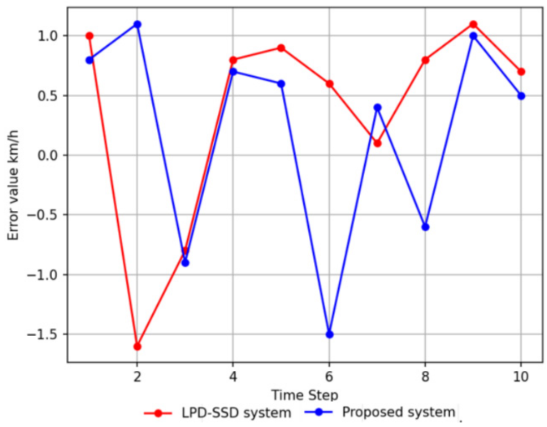

Fig. 17. Compare the error (km/h) of the proposed method (blue) and the LPD - SSD[1st] method (red)

5. Conclusion

In this study, data augmentation has been adopted for data normalization in the speed detection problem. The speed detection error in the tachometer within the range of [5.67, 8.64] and [-1.5 km/h, +1.1 km/h] and working distance up to 30 m prove that the proposed system has achieved relatively high efficiency in terms of transportation environment in Vietnam highways.

However, the range of detectable speed should be improved to work with other highways which allow maximum speeds of up to 120 km/h. The future work of this study will aim to handle this issue in video-based speed detection problems.

References

[1] Lei yang, Menglong Li, Xiaowei Song, Zhi Xiang Xiong, Chunping Hou, Boyang Que, Vehicle speed measurement based on binocular stereovision system, Electronic ISSN: 2169-3536, IEEE 30 July 2019, Page(s) : 106628 - 106641 https://doi.org/10.1109/ACCESS.2019.2932120

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, Speedup robust features (surf), Computer Vision and Image Understanding, 110(3): 346-359, 2008. https://doi.org/10.1016/j.cviu.2007.09.014

[3] Nicolai Woke, Alex Bewley, Dietrich Paulus, Simple online and realtime tracking with a deep association metric, 2017 IEEE International Conference on Image Processing (ICIP), September 2017. https://doi.org/10.1109/ICIP.2017.8296962

[4] A. S. Gunavan, D. A. Tanjung, F. E. Gunavan, Detection of Vehicle Position and Speed using Camera Calibration and Image Projection Methods, 4th International Conference on Computer Science and Computational Intelligence 2019, pp. 255-265, September 2019.

[5] J. Sochor et. al, Comprehensive Dataset for Automatic Single Camera Visual Speed Measurement, IEEE transactions on intelligent transportation systems, pp(99) 1-11, May 2018.

[6] M. G. Moazzam, M. R. Haque, M. S. Uddin, Image-Based Vehicle Speed Estimation, Journal of Computer and Communications, Vol. 7, pp 1-5, 2019.

[7] A. Tourani et al, Motion-based Vehicle Speed Measurement for Intelligent Transportation System, Inter. Journal of Image, Graphics and Signal Processing, Vol. 4, pp. 42-55, 2019.

[8] Igor Ševo, Aleksej Avramovic, Convolutional neural network-based automatic object detection on aerial images, 2016 IEEE Geoscience and Remote Sensing Letters, 13(5):740-744, 2016. https://doi.org/10.1109/LGRS.2016.2542358

[9] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, SSD: Single shot multibox detector, European Conference On Computer Vision, pp 21-3729, Dec 2016, arXiv:1512.02325v5 [cs.CV], https://doi.org/10.1007/978-3-319-46448-0_2

[10] Alex Bewley, ZongYuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft, Simple online and realtime object tracking, 2016 IEEE International Conference on Image Processing (ICIP), September 2016. https://doi.org/10.1109/ICIP.2016.7533003

[11] Chris Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 246-252, June 1999.

[12] Nguyen Viet Hung, Nguyen Thi Thao, Do Huy Khoi, Nguyen Tien Dung. Modeling method of vehicle speed detection based on image processing, Journal of Science and Technolgy, Thai Nguyen University, ISSN 1859 - 2171, Issue 169, 9/2017, pp 4-39

[13] David A. Forsyth, Jean Ponce, Computer Vision A Modern Approach, second edition, 2012.

[14] Massimo Piccardi, Background subtraction techniques: a review, IEEE International Conference on Systems, Man and Cybernetics, pp. 3099-3104, 2004.

[15] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, The pascal visual object classes (voc) challenge, International Journal of Computer Vision, vol. 88, no. 2, pp. 303-338, 2010. https://doi.org/10.1007/s11263-009-0275-4