Journal of AI and Data Mining
Vol 8, No 3, 2020, 427-438.
DOI: 10.22044/JADM.2020.8375.1975

Vehicle Type Recognition based on Dimension Estimation and Bag of Word Classification

R. Asgarian Dehkordi and H. Khosravi*

Faculty of Electrical Engineering and Robotics, Shahrood University of Technology, Shahrood, Iran.

Received 30 April 2019; Revised 08 October 2019; Accepted 10 February 2020
**Corresponding author: hosseinkhosravi@shahroodut.ac.ir (H. Khosravi).*

Abstract
The fine-grained vehicle type recognition is one of the main challenges in machine vision. Almost all the methods presented so far have identified the type of vehicle with the help of feature extraction and classifiers. Due to the apparent similarity between the car classes, these methods may produce erroneous results. This paper presents a methodology that uses two criteria in order to identify the common vehicle types. The first criterion is feature extraction and classification and the second one is to use the dimensions of car for classification. This method consists of three phases. In the first phase, the coordinates of the vanishing points are obtained. In the second phase, the bounding box and dimensions are calculated for each passing vehicle. In the third phase, the exact vehicle type is determined by combining the results of the first and second criteria. In order to evaluate the proposed method, a dataset of images and videos, prepared by the authors, is used. This dataset is recorded from places similar to those of a roadside camera. Most existing methods use high-quality images for evaluation, and are not applicable in the real world but in the proposed method, the real-world video frames are used to determine the exact type of vehicle, and the accuracy of 89.5% is achieved, which represents a good performance.

Keywords: *Bag of Words, Camera Calibration, Dimension Estimation, Vehicle Type Recognition.*

1. Introduction

In the recent years, identifying the exact type of vehicles has drawn the attention of the researchers in the field of Intelligent Transportation Systems (ITS) as a challenging subject. Currently, most surveillance systems use the plate numbers of vehicles to identify them. The Automatic Number-Plate Recognition (ANPR) systems are prone to faults in accurate recognition of the vehicle number or the number-plate may be damaged for some reason. Therefore, identifying the vehicle type can be of great assistance to the police in identifying the suspicious vehicles. Vehicle identification has other various applications such as vehicle quantity and type surveys, which can provide useful information about congestion and abundance of specific vehicle types.

So far, various reliable methods have been introduced for vehicle identification, each of which has provided an acceptable accuracy for its own dataset. These methods try to identify the type of vehicle by comparing its appearance with the vehicle classes defined in them. Similarities in some parts of different vehicles, such as similarities in grilles, logos, and lights, are one of the sources of errors in these algorithms. In addition, studies show that a majority of these methods carry out the identification using collections of images taken from close distances and usually the front of the vehicle. However, surveillance cameras are usually located several meters above the road surface, and the images they obtain present overhead views that are rather far from the vehicles. This leads to a reduction in the image quality, making vehicle identification difficult, and increases the error of these methods in practice [1-3].

The main objective of this paper is to present a novel method for the identification of common vehicles in a video received from a roadside camera in such a way that it can be used in real-world applications. The difference between the proposed method and the methods presented so far is that this

method uses two separate criteria to determine the type of a common vehicle:

a) The first criterion uses a method such as those given in [1, 4, 5] to measure resemblance. The collection of pictures used to train the algorithm corresponding to this criterion is obtained by videos recorded from camera roads. Although increasing the number of training data can lead to an improvement in the performance of these methods, the factors such as wind, changes in illumination, small vehicle image size, changes in vehicle appearance, and ambient noise often lead to a degradation of accuracy. For this reason, the second criterion below is used to mitigate the error.
b) The second criterion is to eliminate the image perspective and obtain the vehicle dimensions to determine the vehicle type. For this purpose, the first criterion is used for a few minutes to specify the general dimensions of the conventional vehicle classes on the motion plane obtained from the camera. Then these dimensions are used to identify the vehicle type in the subsequent video frames.

It is recommended in both stages above that the vehicle type and dimensions be determined using several frames rather than only one. Finally, the exact vehicle type is identified by combining the results obtained from the two criteria. Testing the results of the proposed method on sample recorded videos has demonstrated the solid structure and performance of the proposed method.

The rest of this paper is organized as follows. Section 2 reviews the relevant papers. Section 3 describes the proposed method. Section 4 discusses the results. Section 5 concludes the paper.

2. Previous work

Since the proposed method uses two particular criteria to accurately identify the vehicle type, this section will review the previous works associated with each one of these criteria. The methods corresponding to the first criterion that have been recently presented for vehicle type determination can be generally categorized into two groups: (A) those based on deep convolutional neural networks and (B) those based on feature extraction and classification. Another group of methods (C) is based on automatic camera calibration and vehicle dimension estimation. These three categories are investigated in the following.

A. Methods based on deep neural networks

It has been mentioned in [1] that if different cameras are used to provide training data and test images, the classification error increases. In order to reduce the dependency on the training data, captured from a specific camera, the proposed method uses web data to provide adequate-resolution vehicle images. The neural network employed in this research work has an architecture similar to ALEXNET [6], composed of five convolutional layers, three pooling layers, and three fully-connected layers. Faster R-CNN [7], a method used for identifying objects and requiring a large dataset for training, is used for vehicle detection in the images. Yang *et al*. [8] have collected a large dataset for accurately identifying the vehicle and have proposed a CNN-based method capable of vehicle identification from various angles. Yu *et al*. [9] have used two CNNs, one for vehicle detection and the other for type identification.

Sochor *et al*. [10] have presented a method that does not require placing cameras overhead and can use a camera placed in its natural location (the roadside). The data utilized in this method is composed of 21000 images obtained from an angle similar to that of a roadside camera. In this method, the bounding box is formed according to the vanishing points, and the unpacked vehicle image is obtained and applied to CNN for response improvement. The method indicated in [11] is somewhat similar to that in [10]. The database collected in [11] is composed of 116000 images obtained from surveillance cameras.

B. Methods based on feature extraction

Given the symmetry of the front-view image, Hsieh *et al*. [12] first extracted the SIFT points from the whole image, and then they identified all the symmetrical objects in the image by studying the symmetrical points and separated the vehicle from other objects by some processing. In this method, after the vehicle has been detected, its precise type is identified by extracting the HOG and SURF features and applying them to SVM. This method uses 2048 images for training and 4090 images for testing.

Biglari *et al*. [4, 5] presented a part-based method that tries to find the distinctive parts for each subset of vehicles. They mentioned [4] that deep CNNs required burdensome computations and large computer memory and time. For this reason, they used HOG for describing each part and SVM-based methods for classification. In this method, after detecting a vehicle in an image, it is applied to an algorithm that calculates a score to determine the type of vehicle. If the score is smaller than a threshold value, the vehicle is considered to be

unidentified. The database of this method is made up of 5991 vehicle images. With this small dataset, it has achieved a relatively good accuracy (97%).
Munroe and Madden [13] have used the distance between the headlights with the center of the vehicle image as a feature vector for vehicle type identification. In [14], the prominent forms of the rear lights and the number-plate have been used as features to overcome nighttime illumination problems. The SVM, KNN, and decision tree are used for classification. Sarfraz *et al*. [15] have proposed a probabilistic method that automatically learns a set of segments for the vehicle classes during the training stage, and determines the vehicle class in the test stage according to the extent of similarity between patches and the training model.

C. Methods based on dimension calculation and perspective elimination

So far, a few methods have been presented based on calculating the dimensions of the vehicles. In order to perform this task for moving vehicles, the perspective is required to be omitted. Many methods involving calibration and perspective correction use vanishing points [16-18].
In the method proposed by Dubska *et al*. [19], the vanishing points and the focal length are first determined using several frames from the input video. Then using these parameters, a bounding box is created for each vehicle. By projecting this box onto the motion plane, after several frames, the scale factor is obtained, and the vehicle speed and dimensions are computed. The error involved in the speed and distance estimation in this method is less than 2% for the particular dataset.
In [20] (the previous work of the authors), a fully automatic method for camera calibration and traffic analysis has been presented. In this method, the area of each vehicle on the road is specified after perspective elimination, and its dimensions are calculated by applying a transformation and using a metric factor. The 3D bounding box is created with a high accuracy in this method, and the error of dimension estimation is 1.5%. However, in the case of shadows or changes in ambient illumination, this method degrades significantly. Moreover, the procedure to determine the metric factor is time-consuming.

3. Proposed method

Figure 1 shows a flowchart of the proposed method. According to this figure, the proposed method is made up of three general phases, where the first and second phases are indeed a pre-processing for the third phase. In the first phase shown inside the orange box, the vanishing points and the camera focal length are determined using the motion path of the vehicles in the first few frames of the input video. These parameters are computed only once at the outset.
In the second phase, displayed within the blue box, the vehicle area is first determined by eliminating the undesirable noise effects and shadows, and the resulting vehicle images are processed in two parallel paths. In one path, the vehicle type is identified with the help of the BOW method, and in the other, the vehicle bounding box is determined using the vanishing points, and the perspective is eliminated by projecting the coordinates of the box vertices on the hypothetical motion plane to compute vehicle length, width, and height. Finally, by identifying several representative vehicles for each class and merging and evaluating the results of the two paths, the dimensions of common vehicles on the motion plane are obtained, and this phase ends.
After the completion of the first and second phases, the coordinates of the vanishing points and the dimensions of common vehicles are specified on the motion plane. According to this information, the dimensions of the passing vehicles can also be specified. Using the dimensions alongside the results of the BoW[1] classification, the vehicle type is identified accurately. This procedure is shown inside the green box of figure 1.
In short, we used two criteria for vehicle type identification; one based on BoW and the second based on the vehicle dimensions. The second criterion covers some errors of the first criterion. This somewhat broadens the choice of the first method. However, since the processing time is a very important issue while dealing with video, we preferred not to use the deep learning methods that require expensive GPU, a huge number of training samples, and a long time for training. Therefore, given the methods presented in [4] and [5] and due to the presence of well-defined vertices and edges in the structure of a vehicle, it would be possible to obtain an acceptable response using the features such as HOG [21], SIFT [22], SURF [23], and classifiers such as SVM [24]. For this reason and based on the conducted experiments, the Bag of Words method, which is a strong recognition algorithm, is selected as the first criterion.

[1] Bag of Words

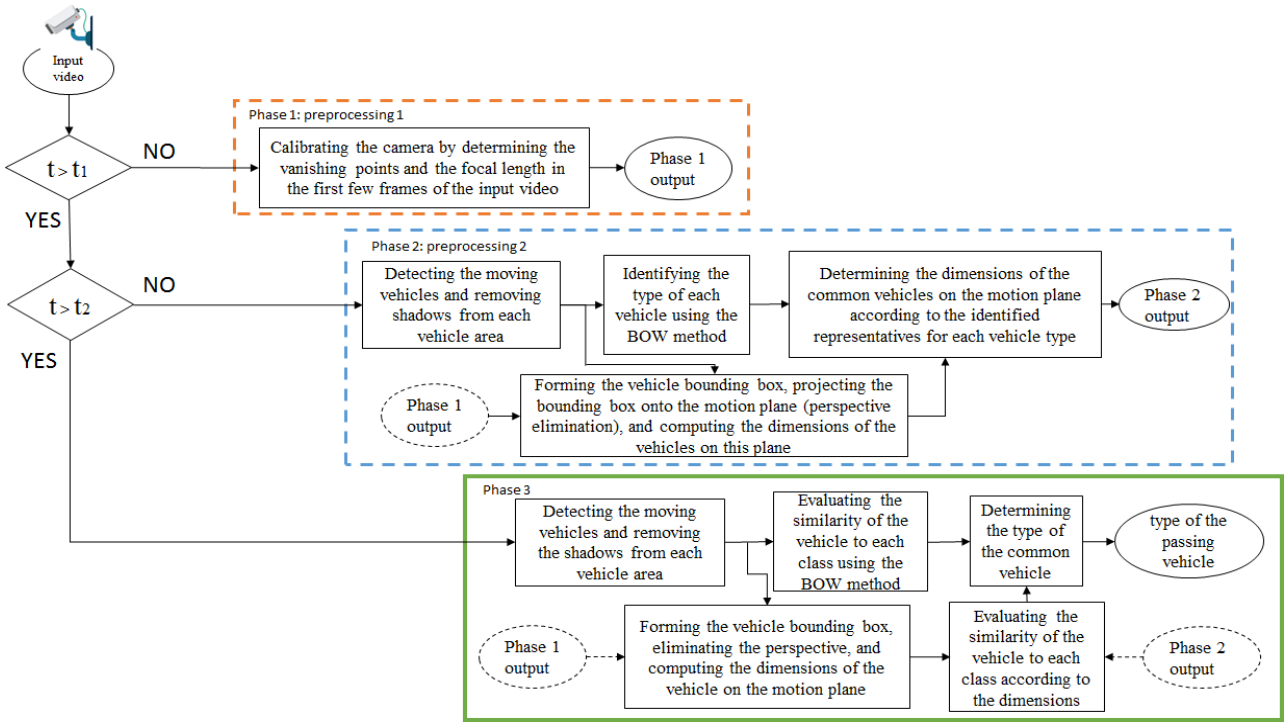

Figure 1. Flowchart of the proposed method.

3.1. Phase 1: Camera calibration according to road surface and motion of vehicles

In order to estimate the coordinates of the vanishing points (VP_1, VP_2, and VP_3) and the focal length (f) in the proposed algorithm, the method presented in [18] is used. In this method, VP_1 and VP_2 are determined according to the direction of motion of the vehicles. Then f is calculated using Eq. (1)

$$Vp_1 = \begin{bmatrix} u_x \\ u_y \end{bmatrix},\ Vp_2 = \begin{bmatrix} v_x \\ v_y \end{bmatrix},\ C = \begin{bmatrix} p_x \\ p_y \end{bmatrix} \tag{1}$$

$$f = \sqrt{-(Vp_1 - C)^T (Vp_2 - C)}$$

where C is the center coordinates of the image. Given the f value, the third vanishing point is obtained using Eq. (2).

$$Vp_1' = \begin{bmatrix} u_x \\ u_y \\ f \end{bmatrix},\ Vp_2' = \begin{bmatrix} v_x \\ v_y \\ f \end{bmatrix},\ c' = \begin{bmatrix} p_x \\ p_y \\ 0 \end{bmatrix}, \tag{2}$$

$$Vp_3' = (Vp_1' - c') \times (Vp_2' - c')$$

By choosing the first two components of VP_3' as the components of VP_3, the coordinates of the third vanishing point are determined. These parameters are computed usually during the first minute of the surveillance video ($t1 \leq 1$ min). More details can be found in [25].

3.2. Phase 2: Obtaining dimensions of common vehicles

As mentioned earlier, the objective of the proposed method is to lower the type identification error using two different criteria. The first criterion is to identify the vehicle from its appearance. For this purpose, the BOW method is used. The second criterion, which complements the first one, is identification using vehicle dimensions after perspective elimination.

The purpose of the second phase is to determine the perspective-free dimensions of common vehicles so that they may be used in the third phase as an auxiliary criterion for vehicle type identification. For this purpose, in about 5 minutes of the input video ($t_2 \geq 5$min), the vehicle types are recognized using the BoW method. Now we have at least 4 candidates for each class of vehicles. In the next step, outliers of each class, i.e. vehicles in which their dimensions differ significantly from other vehicles in that class, are removed. The average dimensions of the remaining candidates are saved as the target dimensions of that class.

3.2.1. Identification of moving vehicles and determination of exact shadow-free area for each vehicle

Given the moving nature of the vehicles, the foreground detection method can be used to identify the vehicle area. Various methods have been presented so far in this regard. According to the numerous tests conducted on these methods, the one presented in [26] is selected for this purpose. This method is fairly fast and exhibits acceptable resistance against noise. Furthermore, it updates the background model during the time.

One of the main issues of the foreground detection methods is their unacceptable response in the presence of shadows. If shadows are considered as part of the vehicle, in the subsequent steps, the algorithm will produce a bounding box that is larger than the actual bounding box. Thus, we must remove shadows before the subsequent steps to make our algorithm shadow resistant. Since shadow elimination is part of the main algorithm, it must have low computation. For this, the method presented in [27] is used for shadow elimination, based on the experiments. This method is capable of determining the vehicle area in the presence or absence of shadows.

3.2.2. Computing Vehicle Dimensions

In order to determine the real dimensions of vehicles, at first, we must construct the 3D bounding box. The details can be found in [20]. In the top picture of figure 2, the green lines are the tangents drawn to the first vanishing point, the red lines are the tangents drawn to the second one, and the blue lines are those drawn to the third one. The bottom pictures of figure 2 display the bounding box created for the vehicle. Points A, B, C, and E are the coordinates of some of the vertices of the bounding box. The distances AE, AB, and AC are used as the height, width, and length of the vehicle in the subsequent calculations.

After obtaining the bounding box and its dimensions, it is observed that the vehicle is reduced in size by moving toward the first vanishing point, and its obtained dimensions vary with time. On the other hand, the determined dimensions of a vehicle must be the same irrespective of the frame. In order to solve this problem, a motion plane parallel with the actual road surface is used, and by projecting the vehicle onto this plane, its dimensions can be calculated.

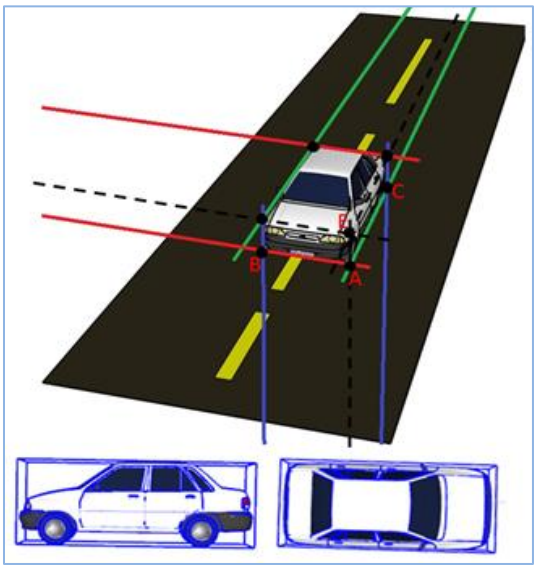

Figure 2. Formation of the vehicle bounding box.

The hypothetical motion plane ϕ must be parallel to the actual road; therefore, this plane is parallel to the axes aligned with the first and second vanishing points, and its normal vector can be determined by the cross product of these two vectors. The normal vector is, in fact, parallel to the direction of the third vanishing point. To project the vehicle onto the motion plane, the center of the image plane is taken as $P = [P_x\ P_y\ f]$ and the camera location as $O = [P_x\ P_y\ 0]$ (P_x and P_y are the coordinates of the image center). Figure 3 displays the camera, the image plane, and two arbitrary motion planes with different distances from the camera. By considering the motion plane ϕ, the points A_W, B_W, and C_W are obtained by projecting the corresponding points A, B, and C from the image plane onto the motion plane ϕ, and point E_W is obtained by projecting point E onto the normal vector (N) of the motion plane:

$$A_w = \phi \cap \overrightarrow{OA},\ C_w = \phi \cap \overrightarrow{OC},\qquad B_w = \phi \cap \overrightarrow{OA},\ E_w = N \cap \overrightarrow{OE} \tag{3}$$

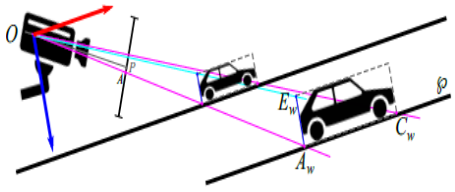

Figure 3. Representation of the camera, image plane, and two arbitrary motion planes.

After projecting the points, the length, width, and height of the vehicle are determined on the motion plane (ϕ) as below:

$$L = |A_w C_w|,\ W = |A_w B_w|,\ H = |A_w E_w| \tag{4}$$

It must be noted that these dimensions are not the actual metric dimensions of the vehicle. Using (4), it is possible to estimate the dimensions of the vehicle in each frame for the arbitrary plane. However, these values are almost identical in some frames and slightly different in the other frames. The reason for these differences is the incorrect foreground detection and the slight difference between the direction of the vehicle and that of the vanishing points in some frames. This issue must be resolved in such a way that a fixed and exact size is reported for a vehicle in all frames. Via the studies carried out for several vehicles in several videos, it was found out that if the dimensional values are slightly rounded after computation and the histogram of the length, width, and height of a vehicle is obtained for the frames where the vehicle is in the sight of the camera, the maximum of these histograms relates to the more accurate length, width, and height of the vehicle.

In order to create the dimension histogram of a vehicle, it must be tracked as long as it is in the front of the camera. Here, the optical flow [28] is used to identify the location of the vehicle, and by tracking point A (Figure 2) for each vehicle, its location in each frame is determined. As such, every vehicle that appears at the front of the camera is tracked, and histograms of its dimensions are created while it is in the field of view. The length, width, and height of each vehicle are then determined according to the maximum of these histograms.

3.2.3. Vehicle type identification using BOW

The Bag of Word method has been used as a successful method in many machine vision and pattern recognition algorithms [29-31]. In the proposed method, the technique presented in [32], which is based on BoW and SVM, is used to identify the type of vehicle. One of the advantages of this method is that it produces an acceptable response even with a small training dataset.

This algorithm assigns a score based on the similarity of the vehicle to each class, and introduces the class with the highest score as the winner. For instance, by applying this method to the image in figure 4, the scores and the class vector are obtained as in table 1.

Figure 4. Sample test image.

Table 1. Vehicle type identification using the first criterion.

Class name	Score
Pride131	**-0.0979**
Peugeot Pars	-0.1244
Peykan	-0.1245
Pegout 405	-0.1244
⋮	⋮

As it can be seen, Pride has earned the highest score among the reported values, and the type of vehicle is reported as Pride.

During the analysis of the successive frames, sometimes, one vehicle may be classified into different classes. In order to resolve this issue, we trace the classification results of the BoW method in successive frames and save them into an array. Finally, when the vehicle goes outside the camera view, its class is computed as the mode of that array (Eq. 5).

$$C = \text{argmax}(\sum_i class_i) \quad (5)$$

Here, i denotes the frames in which the vehicle is tracked and $class_i$ represents the vehicle class in each frame. The studies carried out showed that in this situation, the error was smaller compared to the situation where the vehicle type is identified using only one frame.

Although this technique slightly reduces the classification error, still there are some cases that the BoW fails in successive frames, and the final decision is incorrect. This is due to the physical similarity of some vehicles, for example, the similarity in grilles and headlights. Figure 5 displays the results obtained from applying the first criterion for cases where this method has not been able to correctly identify the vehicle type.

Figure 5. Some of the incorrect identifications made by the BOW method.

A comprehensive analysis of the errors in the first criterion shows that, in some cases, the identified class for a vehicle is similar to the actual class in terms of dimensions and body shape (e.g. the similarity between Pride 131 and Pride 132). However, in some cases, the winner class was completely different from the actual class (e.g. Pride 131 and Peugeot 405). For this reason, it is suggested in the proposed method that the vehicle dimensions be used in addition to the BoW method to report the vehicle type more accurately. Therefore, the dimensions of common vehicles are identified on the motion plane in the subsequent stage, and then these dimensions are used in the third phase to increase the accuracy of identification.

3.2.4. Dimension determination for common vehicles on arbitrary plane

In Sections 3.2.2 and 3.2.3, the dimensions of the passing vehicles on the motion plane and the type of each vehicle are determined. By combining this information, it is possible to determine the dimensions of a particular vehicle (e.g. Peugeot 405) on the motion plane. For this purpose, N samples for each desired vehicle type are identified in the initial frames, and by placing the dimensions of the N samples in a matrix, a 3 × N matrix is obtained for each vehicle class, the first, second, and third rows of which relate to the length, width, and height of the identified vehicles. Thus for each vehicle class such as Pride 131, Peugeot 405, and Peugeot Pars, N samples must be identified in the initial frames. Since the intention here is to identify common vehicles and there is a sufficient traffic on the roads monitored by surveillance cameras, identifying this number of samples is not considerably time-consuming. Equation (6) shows the formation of the matrix.

$$Candid\ \ Dimension_{m} = \begin{bmatrix} W_1 \ldots W_N \\ H_1 \ldots H_N \\ L_1 \ldots L_N \end{bmatrix} \qquad (6)$$

$$1 < m < class\ \ number$$

After the desired matrices for each class are formed, the length and width values that are not consistent with the other elements of the matrix are eliminated, and by averaging the remaining values in each row, the width, height, and length (W_m, H_m, L_m) of any particular vehicle type can be determined on the arbitrary motion plane. The accuracy of the BOW method was found out to be over 80% both in the experiments conducted herein and in the values reported by other research. Hence, there is still a probability that this method may have identified the vehicle type incorrectly. The reason for the fact that N vehicles are identified for each class in the proposed method and then samples that are inconsistent with other samples are eliminated ensures that incorrectly-identified samples do not affect the determination of the considered vehicle dimensions on the hypothetical motion plane. The number N must be sufficiently large, i.e. $N \geq 4$. As such, it can be ensured that the majority of identified samples correspond to the considered class, and the effect of incorrect samples on the dimensions is canceled by eliminating the outliers.

3.3. Phase 3: Type determination for passing vehicles

In the first and second phases, the parameters required for calibration and the dimensions of each vehicle class on the arbitrary plane were computed. These parameters are used in the third phase to accurately identify the types of passing vehicles. In this phase, the dependence of each vehicle on the classes is determined using the first criterion by assigning a score to each class; then according to these scores and by comparing the dimensions of the passing vehicle and the dimensional range of the common vehicles, the type of each vehicle is identified with a high accuracy. As it can be seen in figure 1, the steps of detecting the passing vehicles, computing the dimensions on the motion plane, and comparing the similarity using the BOW

method are common between the second and third phases. Consequently, the re-explanation of these steps is avoided, and only the steps unique to the third phase will be studied in the following.

3.3.1. Similarity evaluation between passing vehicle and each class via dimensional comparison

After eliminating the perspective and calculating the dimensions of the passing vehicle, it is time for comparing these dimensions with those corresponding to common vehicles. The error in computing the vehicle dimensions in [20] is 1.5%. The error is considerably reduced in the approach proposed by the present paper due to the lack of need for the metric factor (to convert the dimensional units to meters) and the individual evaluation of each vehicle class for determining its dimensions on the motion plane (Section 3.2.4). However, the values obtained for each vehicle fluctuate slightly around the values computed in Section 3.2.4. Hence, for evaluating the dependence of a vehicle on a given class, the following relationships must be considered.

$$\begin{aligned} L_m(1-K) < L_W &< L_m(1+K) \\ W_m(1-K) < W_W &< W_m(1+K) \\ H_m(1-K) < H_W &< H_m(1+K) \end{aligned} \quad (7)$$

In these relationships, L_m, H_m, and W_m are the dimensions corresponding to each class, and L_W, H_W, and W_W are those corresponding to the passing vehicle. If the K value is selected to be too large, many classes will be chosen as the main class of the vehicle, and the second criterion will lose its efficiency, and if it is selected to be too small, the vehicle may not be placed in the range of its actual class. Accordingly, and given the experiments carried out in Section 3.2.4 for the vehicles in the test videos, K is chosen to be 0.03, meaning that up to 3% error for each dimensional component is acceptable.

Therefore, if all conditions of (7) hold for a vehicle, it can be placed in that class. In this way, sometimes a vehicle can be placed in more than one class. For instance, in the Peugeot family, the dimensions of Peugeot 405 and Peugeot Pars are almost identical, and if the passing vehicle is a Peugeot, it is placed in both classes. Also, there is a small probability, due to illumination conditions or other reasons, for the dimensions of a passing vehicle to have a difference larger than 3% with their reference values. In this case, the second criterion will not place the vehicle in the correct class but the subsequent part of the proposed method is designed in such a way that this issue does not adversely affect its performance. Table 2 shows the dependence of the dimensions of the white Pride in figure 5 on some of the classes.

Table 2. Vehicle classification using the dimensions.

Class name	Dimensions matched
Peugeot Pars	No
Pride 132	**Yes**
Samand	No
Pride 131	**Yes**
Peugeot 405,SLX	No

3.3.2. Type determination for passing vehicle

As expressed in Section 3.2.3, in some cases, BoW is unable to correctly identify the type of the passing vehicle. Using the second criterion, vehicle dimensions reduced some errors but there are still some vehicles classified incorrectly. By evaluating the classification results of BoW, we found that almost in all misclassified cases, the second class with the highest score was the correct class. In order to improve the results even more, we tried to combine the BoW scores and the vehicle dimensions instead of applying each method, separately. Therefore, to determine the exact vehicle type, the winner classes for a vehicle are sorted according to their BoW scores. Then the first one that satisfies the dimensional conditions is considered as the vehicle type. Furthermore, if none of the first two choices of the BoW method satisfy the dimensional conditions, the vehicle class is considered to be the one identified by the first criterion (it is probable in this case that the computed vehicle dimensions are wrong, and by considering the first choice of the BoW method as the winner class, the adverse effect of the second criterion is reduced). Table 3 displays this process for the white Pride of figure 5, which is initially misclassified by BoW as Peugeot 405. The BoW scores are sorted in the ascending order.

Table 3. Vehicle classification using BoW and dimensions.

Class name	Bow scores	Dimensions verified?	Final class
Peugeot 405	**-0.1438**	No	-
Pride 131	-0.1471	**Yes**	☑
Pride 132	-0.1485	Yes	-
Peugeot Pars	-0.1537	No	-
⋮	⋮	⋮	⋮

As it could be seen in table 3, the vehicle type was correctly identified by combining the BOW method data and the dimensions.

4. Experimental results

The dataset used for testing the proposed method is composed of a collection of videos and images. The videos were obtained, with a police license, by recording from various roads in Iran. For these videos, the camera was located a few meters above the road surface, and its angle was adjusted similar to that of the road surveillance cameras. The dataset consists of approximately 7000 images, 3500 with a frontal view, and the other 3500 with rear view of the vehicle. Part of the frontal view images was provided by the police and the other part was obtained by extracting one or two images of each vehicle from the road surveillance videos. The images include numerous different vehicle groups but most of these images logically correspond to sedans and hatchbacks common in Iran, and other groups have a smaller share. Given the fact that identifying common vehicles is of greater practical importance (compared to less common vehicles), 9 classes of common vehicles were considered for identification. The topmost picture of figure 6 displays a sample frame from a traffic surveillance camera. Except for one or two vehicles, the rest of the vehicles can be placed in the classes considered for identification. The rest of the pictures in figure 6 show some of the images in the dataset and video frames (along with the box constructed for the vehicles by the algorithm).

Table 4 displays the accuracy of the BoW and several other methods for the test images. Bus-Truck, Pickup Truck, and others are all considered unknown groups that were taught to methods to increase the accuracy in determining the dimensions (Sections 3.2.3 and 3.2.4). The presence of unknown groups hinders the job of the algorithms. Despite the long distance between the camera and the vehicles, the average accuracy of the BoW method was 90.1%, which demonstrated the efficiency of this method.

Figure 6. First row: The image of an urban surveillance camera, second and third rows: sample training images, fourth and fifth rows: sample video frame.

Table 4. Comparison of methods accuracy.

Class name	BOW+ SVM	HOG+ SVM	BOW+ MLP	HOG+ MLP
Pride 131	**89**	84	87	82
Peugeot Pars	**95**	97	93	96
Peugeot206	**99**	97	96	95
Peugeot 405	**82**	79	80	73
Pride 132	**85.5**	83	84	80
Peykan	**96**	85	94	81
L90	**93**	90	90	89
Samand	**89**	90	88	89
Pride 141	**91.5**	89	89	85
Others	**85**	78	83	74
Bus-Truck	**86**	83	82	79
Pickup Truck	**90**	88	85	86

Shadow removing helps us to identify the exact vehicle area. It is not only essential for dimension estimation but also significantly increases the accuracy of the BoW method.
Figure 7 shows the area with and without shadows for a Peugeot 405, and the results of BoW identification for both are shown in table 5. It is clear that after shadow elimination, the vehicle was classified correctly.

Figure 7. Vehicle bounding box before (left) and after (right) shadow elimination.

Table 5. Importance of shadow elimination in BoW scores for the vehicle of figure 7 (Peugeot 405).

Class name	Scores **before** shadow removing	Scores **after** shadow removing
Pride 141	-0.1177	-0.1190
Samand	-0.1177	-0.1190
L90	-0.1177	-0.1190
Peykan	-0.1177	-0.1187
Pride 132	-0.1181	-0.1195
Peugeot 405	-0.1177	**-0.1075**
Peugeot 206	-0.1177	-0.1174
Peugeot Pars	-0.1177	-0.1179
Pride 131	**-0.1046**	-0.1175

Due to the applicability of the proposed method to video, the difference of test datasets, and the lack of access to the implementation of other methods, a fair comparison with other methods is unachievable. Table 6 displays the results of vehicle identification in the test videos. The test videos were obtained under the same conditions as the videos corresponding to training images. The values in the first row display the results of the first criterion + video buffering, and those in the second row show the results of the first criterion + the second criterion + video buffering. The results in the last column show the average accuracy for all the videos.
It has been mentioned in [4] that increasing the number of test sets increases the error, a relationship that also holds for the proposed method. Moreover, if the number of training classes increases, the error will likely rise. In any case, by identifying the common vehicle classes, the main requirement of traffic monitoring systems will be satisfied. Besides, as shown in table 6, using two criteria in the proposed method has resulted in a 16% decrease in error, and the 89.5% accuracy demonstrates the adequate performance of the proposed method on video.

Table 6. Identification accuracy for the test videos.

Video name	BoW + Video buffering	BoW + Buffering + Dimension verification
Test 1	76	**93**
Test 2	62	**87**
Test 3	77	**86**
Test 4	75	**91**
Test 5	79	**92**
Test 6	73	**88**
Average	73.66	**89.5**

5. Conclusion

A new approach was proposed that identifies the vehicle type in roadside videos using two criteria. The first criterion identifies the vehicle type based on the physical features, and the second criterion reduces the error of the first criterion using the dimensions of the vehicle. Studies of the results of applying this method to the test videos have shown a remarkable decrease in error due to the use of the second criterion. In order to train and test the proposed method, a set of tagged images and videos were collected that would be released to the public for future research works.
To improve the performance of the proposed method, one may measure the accuracy of the algorithm by testing other powerful methods such as CNNs instead of BoW. Also it is possible by enlarging the dataset, a task being undertaken as of the writing of this paper, to better evaluate the performance of this algorithm and improve its efficiency by recognizing its flaws.

References

[1] Wang, J., et al. (2018). Vehicle Type Recognition in Surveillance Images From Labeled Web-Nature Data Using Deep Transfer Learning, IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 9, pp. 2913-2922.

[2] Huang, Y., et al. (2015). Vehicle logo recognition system based on convolutional neural networks with a pre-training strategy, IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 4, pp. 1951–1960.

[3] Psyllos, A. P., Anagnostopoulos, C.-N. E. & Kayafas, E. (2010). Vehicle logo recognition using a SIFT-based enhanced matching scheme. IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 2, pp. 322–328.

[4] Biglari, M., Soleimani, A. & Hassanpour, H. (2018). A Cascaded Part-Based System for Fine-Grained Vehicle Classification. IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 273-283.

[5] Biglari, M., Soleimani, A. & Hassanpour, H. (2017). Part-based recognition of vehicle make and model. IET Image Processing, vol. 11, no. 7, pp. 483-491.

[6] Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097–1105.

[7] Ren. S, et al. (2015). Faster R-CNN: Towards realtime object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, pp. 91–99.

[8] Yang, L., et al. (2015). A large-scale car dataset for fine-grained categorization and verification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp. 3973-3981.

[9] Yu., Shaoyong, et al. (2017). A model for fine-grained vehicle classification based on deep learning. Neurocomputing, vol. 257, pp. 97-103.

[10] Sochor, J., Herout, A. & Havel, J. (2016). BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp 3006-3015.

[11] Sochor, J., Špaňhel, J., Herout, A. (2019) BoxCars: Improving Fine-Grained Recognition of Vehicles using 3-D Bounding Boxes in Traffic Surveillance. IEEE Transactions on Intelligent Transportation Systems. vol. 20, no. 1, pp. 97-108.

[12] Hsieh, J., Chen, L., & Chen, D. (2014). Symmetrical SURF and Its Applications to Vehicle Detection and Vehicle Make and Model Recognition. IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 1, pp. 6-20.

[13] Munroe, D. T., & Madden, M. G. (2005). Multi-class and single-class classification approaches to vehicle model recognition from images. Information Technology, pp. 93–102.

[14] Boonsim, N. & Prakoonwit, S. (2017). Car make and model recognition under limited lighting conditions at night. Pattern Analysis and Applications, vol. 20, pp. 1195–1207.

[15] Sarfraz, M. et al. (2011). A Probabilistic Framework for Patch-based Vehicle Type Recognition. VISAPP.

[16] Cathey, F. & Dailey, D. (2005). A novel technique to dynamically measure vehicle speed using uncalibrated roadway cameras. Intelligent Vehicles Symposium, pp. 777–782.

[17] You, X. & Zheng, Y. (2016). An accurate and practical calibration method for roadside camera using two vanishing points. Neurocomputing.

[18] Dubská, M. (2015). Fully Automatic Roadside Camera Calibration for Traffic Surveillance. IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 3, pp. 1162-1171.

[19] Dubsk´a, M., Sochor, J. & Herout, A. (2014) Automatic camera calibration for traffic understanding. BMVC, 2014.

[20] Asgarian Dehkordi, R. & Khosravi, H. (2017) Auto Detection of Vehicle Dimensions using Videos from Roadside Camera. 10th Iranian Conf. Machine Vision and Image Processing (Farsi edition), Isfahan.

[21] Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 886–893.

[22] Lowe, D. (1999). Object recognition from local scale-invariant features. Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, vol. 2, pp. 1150–1157.

[23] Bay, H., et al. (2008). Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359.

[24] Chang, C.-C. & Lin, C.-J. (2001). LIBSVM: A Library for Support Vector Machines. [Online].Available:http://www.csie.ntu.edu.tw/~cjlin/Libsvm.

[25] https://www.coursera.org/lecture/robotics-perception/how-to-compute-intrinsics-from-vanishing-points-jnaLs

[26] Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. Proceedings of the 17th International Conference on Pattern Recognition (ICPR), vol. 2, pp. 28-31.

[27] Lin, CT., et al. (2010). An Efficient and Robust Moving Shadow Removal Algorithm and Its Applications in ITS. EURASIP Journal on Advances in Signal Processing. vol. 2010, no. 39.

[28] Kamijo, S., et al. (1999). Traffic monitoring and accident detection at intersections," International Conference on Intelligent Transportation Systems, Tokyo, pp. 703-708.

[29] Van de Weijer, J., & Khan, F. S. (2013). Fusing color and shape for bag-of-words based object recognition. International Workshop on Computational Color Imaging, Springer, Berlin, Heidelberg. pp. 25-34.

[30] Bokharaeian, B., & Diaz, A. (2016). Extraction of Drug-Drug Interaction from Literature through Detecting Linguistic-based Negation and Clause Dependency. Journal of AI and Data Mining, vol. 4, no. 2, pp. 203-212.

[31] Momtazi, S., Rahbar, A., Salami, D., & Khanijazani, I. (2019). A Joint Semantic Vector Representation Model for Text Clustering and Classification. Journal of AI and Data Mining, vol. 7, no. 3, pp. 443-450.

[32] Bag of Visual Words for Image Classification. http://heraqi.blogspot.com/2017/03/BoW.html, accessed 18 March 2017.

نشریه هوش مصنوعی و داده کاوی

بازشناسی نوع خودرو با استفاده از تخمین ابعاد و طبقه‌بند کیف کلمات

رسول عسگریان دهکردی و حسین خسروی*

دانشکده برق و رباتیک، دانشگاه صنعتی شاهرود، شاهرود، ایران.

ارسال ۲۰۱۹/۰۴/۳۰؛ بازنگری ۲۰۱۹/۱۰/۰۸؛ پذیرش ۲۰۲۰/۰۲/۱۰

چکیده:

باز شنا سی نوع دقیق خودرو یکی از م سائل پرچالش در حوزه بینایی ما شین ا ست. تقریبا تمام روش‌هایی که تاکنون ارائه شده ا ست، نوع خودرو را به کمک ا ستخراج ویژگی و طبقه‌بند شناسایی می‌کنند. با توجه به شباهت ظاهری بین کلاس‌های اتومبیل، این روش‌ها ممکن ا ست نتایج نادر ست به بار آورد. در این مقاله رو شی ارائه شده ا ست که از دو معیار برای باز شنا سی انواع و سایل نقلیه متداول ا ستفاده می‌کند. معیار اول، ا ستخراج و طبقه‌بندی ویژگی‌ها و معیار دوم استفاده از ابعاد ماشین برای طبقه‌بندی است. این روش شامل سه مرحله است. در مرحله اول مختصات نقاط محوشدگی پیدا می شود. در مرحله دوم، مستطیل محیطی خودرو و ابعاد آن برای هر وسیله نقلیه عبوری محاسبه می‌شود. در مرحله سوم، نوع دقیق خودرو با ترکیب نتایج معیارهای اول و دوم تعیین می‌شود. برای ارزیابی روش پیشنهادی، از مجموعه‌ای از تصاویر و فیلم‌ها، تهیه شده توسط نویسندگان مقاله، استفاده شده ا ست. این مجموعه داده از زاویه‌هایی م شابه زاویه دوربین‌های کنار جاده ضبط شده ا ست. اکثر روش‌های موجود برای ارزیابی از تصاویر با کیفیت بالا اســـتفاده می‌کنند که در دنیای واقعی کاربرد ندارند اما در روش پیشـــنهادی از فریم‌های ویدیویی از دنیای واقعی برای تعیین نوع دقیق وســـیله نقلیه استفاده شده و دقت ۸۹٫۵٪ حاصل شد، که عملکرد خوبی را نشان می‌دهد.

کلمات کلیدی: کیف کلمات، تنظیم دوربین، تخمین ابعاد، بازشناسی نوع خودرو.