Transactions on Machine Intelligence 2(1) (2019) 33–39

ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

A SVM Based Approach for Real Time Detection and Classification of Vehicles at the Toll Gates Using Video Sequences

A. Mosayebi[1,*], H. Khosravi[2]

[1,2] Electrical Engineering Department, Shahrood University, Shahrood, Semnan, Iran

ARTICLE INFO

Article History:
Received 19 November 2018
Received in revised form
4 January 2019
Accepted 16 March 2019
Available online 16 March 2019

Keywords:
Vehicle Detection,
Classification,
Videos, Rois,
Svm, Toll Gates

ABSTRACT

This paper aims to present a real-time scheme for detection and classification of vehicles passing the toll gates in Iran. In our approach, a set of videos are captured using a stationary camera, placed on the roadside, at a little distance from the toll booth. The algorithm is designed in a way that there is no need for camera calibration. Based on our videos, 3 ROIs are defined, two of them are considered to determine if a vehicle is passing and the other one is the region containing the vehicle. This work starts with the training phase, in which, for each image in a manually gathered database, HOG vectors are extracted. Two SVMs are trained in this phase, one for distinguishing vehicles from non-vehicles, and one for classifying vehicles into light and heavy vehicles. After finishing the training, in the testing phase, firstly, foreground mask is obtained differencing two consecutive frames of the video. Then, those two aforementioned ROIs are checked in every frame and as soon as a vehicle is inside the interest region, that ROI is captured. Next, the captured frame is passed to the first SVM and it is classified as vehicle or non-vehicle. Those which are identified as vehicles are passed to the second SVM to be classified as light or heavy vehicle. Average true-positive and precision rates of the vehicle detection step are 92.5% and 97.5% respectively and the same rates, for the recognition step, are 98% and 0.99%.

1. INTRODUCTION

Artificial Neural Networks methods have a wide range of applications [1-6]. In the recent years, intelligent transport systems have been widely developed. Video based technologies, are used more and more in surveillance systems, and since they are as reliable as active and passive sensors, but cheaper and more convenient to use and maintain, are often preferred to such sensors. Vehicle detection and classification are examples of vision based systems, these systems can help police or any other organizations that needs secure automatic monitoring.

There is a vast variety of algorithms that have been presented in the field of vehicle detection and related issues. Using different views of vehicles, like side, front or oblique views, different purposes of the works, various restrictive conditions and many other issues resulted such huge variety. Some of the already proposed approaches

[*] Corresponding Author: moasyebiahmad@yahoo.com
Electrical Engineering Department, Shahrood University

http://dx.doi.org/10.47176/TMI.2019.33

related to the subject of this paper are brought in the following. In [7], a parallel MLP ensemble is trained with the Haar-like features. These features are extracted from manually cropped images of vehicles front views, and they are used to detect the front views of vehicles moving on the road. For classification, a prototype based method called Classified Vector Quantization (CVQ) is applied. An important point about this work is that other than classifying samples into four categories, a reject option is also considered and this reduces the error rate of the algorithm. Another approach [9], uses a combination of Haar-like features and Histogram of Oriented Gradients (HOG) with AdaBoost learning algorithm for detection as explained in [8]. For a two-class classification in that paper, HOG features with AdaBoost is used. In [10], license plate is detected and the ROI containing front view of each car is extracted based on the location and the size of the license plate. Gradient features (like HOG) are extracted, Two-Dimensional Linear Discriminant Analysis (2DLDA) method is applied, and a nearest neighbor classifier is used for classification. There's another approach [11], in which again license plate localization is prior to finding the front view of the vehicle. PCA is applied on raw pixels of the extracted ROI and self-clustering is used for classification. Some works like [12] are focused only on distinguishing between vehicles and non-vehicles. In this paper, vertical and horizontal edge maps are created and used to find regions where vehicles are most probable to be inside them. After finding such regions, using PCA for feature extraction and SVM for classification, the object within the region is recognized as vehicle or non-vehicle. Symmetry is an important property of vehicle front view and an algorithm is proposed based on this property in [13]. This scheme introduces a modified version of "Speeded Up Robust Features" (SURF) called "symmetrical SURF" by which the front view of the vehicle can be found, then it is divided into grids and for each grid SURF are extracted, and vehicle make and model are recognized training a SVM. What might make this approach superior to the other works is that there's no need for foreground extraction or training based methods (like AdaBoost) for detection, but it seems to fail in the cases of noisy or poor-quality videos.

As described already, there is a huge number of works and algorithms that have been implemented, but it seems that little work has been done on detection and classification of vehicles at the toll gates. Working on this sort of videos is a bit challenging due to some specific conditions like partial occlusions or various movement patterns of vehicles. Video sequences used for this work are captured by a stationary camera placed on the roadside, at a little distance from the toll booth and the captured images are rather oblique.

As the main framework of our system, firstly, a relatively large number of images are manually selected for training. These images consist of vehicle samples and non-vehicle samples that are used to train the first SVM (SVM1) and vehicle samples themselves consist of heavy-vehicle samples and light-vehicle samples that are used to train the second SVM (SVM 2). After this, the detection process starts, and every incoming frame of the video is converted to grayscale, then foreground mask is obtained differencing every two consecutive grayscale frames. Two ROIs are considered, one that we call "norm-ROI" and the other that we call "centroid-ROI". When sum of white pixels in the norm-ROI exceeds a certain number and at the same time, centroid in the centroid-ROI passes a certain border, a region of the frame which is defined as the third ROI, is captured, and the captured image is the input for the first SVM. First SVM decides whether the object in the captured image is a vehicle or a non-vehicle. If it's a vehicle, then it is passed to the second SVM and it decides whether it is a light or heavy vehicle. Overview of the proposed system is shown in Figure 1. Five set of videos (videos 01-05) are prepared as the database for our paper, that all of them are about 30 minutes long, except for video 05 that is about 15 minutes long. To test the algorithm, a step by step procedure is considered, meaning that for the first time, SVM1 is trained with the manually selected samples from video 01, tested with video 02 frames, and accuracy of the examination is calculated. Then the correctly distinguished samples of video 02 (vehicle or non-vehicle) are added to the training data so that results get better in the next step. This process continues until video 04 is tested and its proper samples are added to the training set. In the last step, heavy and light vehicles are also collected from the already built training set, SVM2 is trained and both detection and recognition accuracies can be calculated.

The rest of the paper is organized as follows. Section 2 explains the off-line training procedure and feature extraction, Section 3 is pre-classification processing, it describes how the foreground extraction and ROI checking is done and how the image, likely to contain a vehicle, is captured. Section 4, outlines the classification, section 5 discusses the experimental results, and section 6 presents the future work and conclusions.

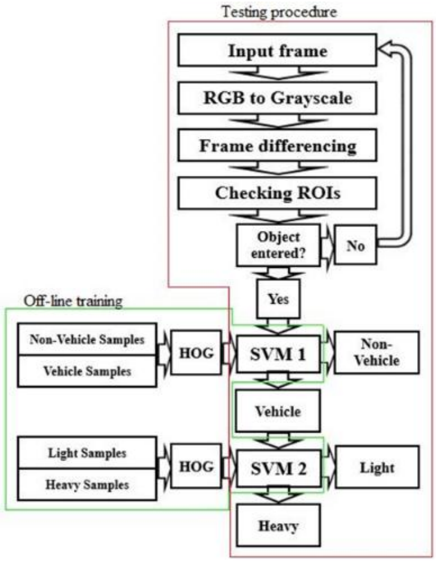

Fig. 1. Overview of the proposed system

2. OFF-LINE TRAINING

In this section, an explanation of how training data is prepared and how it is used for training SVM is given.

2.1. Training Data Preparation

As mentioned before, three ROIs are considered in this work, two of which are used to verify if an object exists and one is the region in image that is to be captured. For creating a dataset of positive (vehicle) and negative (non-vehicle) samples, the third ROI, in one of every five frames of the video 05 is captured and stored. Then about 2000 samples are manually collected as negative samples and about 200 samples as positive samples, HOG feature vectors are calculated for those images and SVMs are trained with those images. In Figure 2, some samples are brought. This training set gets more diverse step by step, and in the last step, positive samples are divided into two groups of light and heavy vehicles and passed to the second SVM as described already.

2.2. Feature extraction

Histogram of Oriented Gradients was first used for the purpose of human detection by Dalal and Triggs in 2005 [8]. HOG computes the occurrence of gradient orientation in localized portions of an image. For each image, there is a HOG vector. HOG descriptor deals with local cells in an image, and this makes this descriptor invariant to geometric and photometric transformations but it's not invariant to changes in orientation of the object. All vehicles in our database appear with the same orientation, so HOG works well for our purpose.

Fig. 2. Manually collected samples a) positive samples-heavy vehicles b) positive samples-light vehicles

c) Negative samples-non-vehicles

3. PRE-CLASSIFICATION PROCESSING

3.1. RGB To Grayscale Conversion

Every incoming RGB frame, has to be converted to grayscale for the frame differencing step using Equation (1).

$$Grayscale = 0.299\,R + 0.587\,G + 0.114\,B \quad (1)$$

3.2. Frame Differencing

Mixture of Gaussians and averaging frames are two examples of foreground extraction methods that are so common. But frame differencing is chosen for this work due to some reasons, first, it is less time-consuming than most of other methods, and second, in our videos, each vehicle stops for a couple of seconds to pay, so methods like GMM won't be able to reliably model the background and vehicles will be present in the modeled background unexpectedly. Every input frame, after being converted to grayscale, is subtracted from the previous grayscale frame, then, thresholding absolute of the resulting image, foreground mask is created.

3.3. Centroid

To calculate the centroid of the foreground mask in the centroid-ROI, x and y coordinates of the centroid have to be calculated. Primarily, first-order moments are calculated for each coordinate separately as in the Equations (2) and (3).

$$m_{10} = \sum_x \sum_y x\ image(x, y) \quad (2)$$

$$m_{01} = \sum_x \sum_y y\ image(x, y) \quad (3)$$

Then x and y coordinates are calculated as in the Equations (4) and (5), and the Equation (6) shows the resulting centroid.

$$x = m_{10}/(\sum_x \sum_y\ image(x, y)\ ) \quad (4)$$

$$y = m_{01}/(\sum_x \sum_y\ image(x, y)\ ) \quad (5)$$

$$Centroid = (x.y) \quad (6)$$

3.4. checking ROIs

Figure 3 shows the aforementioned ROIs in a sample frame of the video, it shows a scene when a vehicle is passing the interest region (defined with a blue rectangle), norm-ROI and centroid-ROI are shown with red and green rectangles respectively. For every frame, percentage of white pixels within the norm-ROI of the corresponding foreground mask is calculated, we call it "P". Centroid of the foreground mask within the centroid-ROI is calculated too, we call it "C" and its y coordinate "C.y". Based on these information, algorithm decides whether to capture the image in the third ROI (blue rectangle in Figure 3) or not, if P is greater than 0.1 and C.y>0, capturing takes place and the captured image is passed to the first SVM. Cropped image, possibly containing an object, is shown in Figure 4, and Figure 5 shows the foreground mask of a scene in which the two above-mentioned conditions are met. The red circle in Figure 5 shows the position of the calculated centroid.

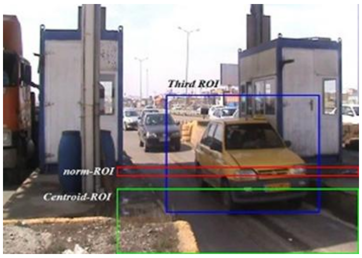

Fig. 3. Defined ROIs

Fig. 4. Cropped image

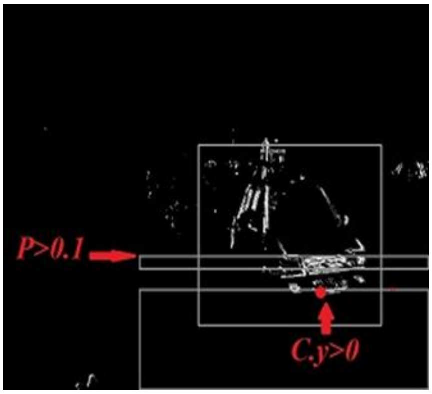

Fig. 5. Foreground mask: Conditions are met in a scene

4. CLASSIFICATION

HOG feature vector is calculated for the cropped image, if this vector is classified as a non-vehicle by the first SVM, classification will stop, but if it is recognized to belong to a vehicle, it will be the input for the second SVM. SVM2 classifies the vector as a light vehicle or a heavy vehicle.

5. EXPERIMENTAL RESULTS

For recording the videos a DCR SR46 Sony Handy-cam was set on a tripod on the road side. Size of each recorded frame is 720×576 and frame rate of the video is 25 frames per second. The step by step process of testing begins with training the first linear SVM (SVM1) with manually cropped images from video 01. This primary classifier is tested with images from video 02, accuracy is verified and the newly recognized true-positive (TP) and false-negative (FN) samples (vehicles) are added to the positive samples of the primary training set. True negative (TN) and false positive (FP) samples are added to the negative samples of the primary training set. A new SVM1 is built with the new training set and it is tested with images from video 03. These similar actions are taken in every step until video 04 is also used for testing. At this point, we have a training set with images of videos 01-04. Then positive samples (vehicles) of this training set are manually divided into two groups of heavy and light vehicles so that the second SVM (SVM2) can be trained using these samples. Eventually, both SVMs are tested with images from video 05. To report the results, two criterions are verified, first one is the True-Positive Rate (TPR) and the second one is the precision or Positive Predictive Value (PPV). Equations (7) and (8) present these two ratios. Results of each step are brought in the Table 1.

$$TPR = TP/(TP + FN) \tag{7}$$
$$PPV = TP/(TP + FP) \tag{8}$$

Table 1. Accuracy results

Step	Training data	Testing data	TPR svm1	PPV svm1	TPR svm2	PPV svm2
#1	From video 01	From video 02	95%	99%	-	-
#2	From videos 01-02	From video 03	97%	95%	-	-
#3	From videos 01-02-03	From video 04	88%	99%	-	-
#4	From videos 01-02-03-04	From video 05	90%	97%	98%	99%

Algorithm is executed on a laptop with a 2.3 GHz Intel core, 8 GB ram memory and 1 TB hard disk and processing time of the algorithm is about 6 milliseconds per frame that is real-time for this purpose.

6. CONCLUSIONS AND FUTURE WORK

The objective of this paper was to build a system of detection and classification of vehicles at the toll gates in Iran. The system totally had two main phases of training and testing. Training had a step by step process and the entire data was not used just once. Cases of misclassification or misdetection happened due to some unwanted conditions, like when some trucks would pass in the neighboring road or when front part of a vehicle is badly damaged. The first problem can be solved by separating very two neighboring lines with a higher wall. Although results prove that performance of the system is acceptable but there are some ways to improve it, like using other classifiers, using neural networks and so forth. Future work can be recognition of make and model of the light-vehicles.

7. ACKNOWLEDGEMENT

The author would like to express his sincere thanks to Dr. Hossei Khosravi, Director of the Department of Electronics of the Shahrood University for all his help and support and Mr Javad Farhadi and Mr Arash Farhadi who helped the author record all videos used in this work.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Rahmani-Seryasat, O., Haddadnia, J., & Ghayoumi-Zadeh, H. (2015). A new method to classify breast cancer tumors and their fractionation. *Ciência e Natura*, *37*, 51. doi:10.5902/2179460x19428
[2] Rahmani Seryasat, O., Haddadnia, J., & Zadeh, H. (2016). Assessment of a novel computer aided mass diagnosis system in mammograms. *Iranian Quarterly Journal of Breast Disease*, *9*(3), 31–41.
[3] Seryasat, O. R., & Haddadnia, J. (2017). Assessment of a novel computer aided mass diagnosis system in mammograms. *Biomedical Research*, *28*(7), 3129–3135.
[4] Haddadnia, J., Seryasat, O. R., & Rabiee, H. (2013). Thyroid diseases diagnosis using probabilistic neural network and principal component analysis. *Journal of Basic and Applied Science Research*, *3*(2), 593–598.

[5] Rahmani Seryasat, O., Kor, I., Ghayoumi Zadeh, H., & Shams Taleghani, A. (2021). Predicting the number of comments on Facebook posts using an ensemble regression model. *International Journal of Nonlinear Analysis and Applications*, *12*, 49–62. doi:10.22075/IJNAA.2021.4796

[6] Rahmani Seryasat, O., Ahmadi, S., Yousefi, P., Tat Shahdost, F., & Sanei, S. (2021). Recognizing phishing websites based on a bayesian combiner. *International Journal of Nonlinear Analysis and Applications*, *12*, 809–823. doi:10.22075/ijnaa.2021.5457

[7] Zhang, B., & Zhou, Y. (2012, Οκτώβριος). Reliable vehicle type classification by Classified Vector Quantization. *2012 5th International Congress on Image and Signal Processing* (CISP), Chongqing, Sichuan, China. doi:10.1109/cisp.2012.6469857

[8] Mita, T., Kaneko, T., & Hori, O. (2005). Joint Haar-like features for face detection. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. Beijing, China. doi:10.1109/iccv.2005.129

[9] Youpan, H., Qing, H., Xiaobin, Z., Haibin, W., Baopu, L., & Zhenfu, W. (2013). Algorithm for vision-based vehicle detection and classification. Στο *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on* (pp. 568–572).

[10] Hua, H., Qian, Z., Yulan, J., & Shuming, T. (2008). A 2DLDA Based Algorithm for Real Time Vehicle Type Recognition. Στο *11th International IEEE Conference on* (pp. 298–303).

[11] Yu, P., Jin, J. S., Suhuai, L., Min, X., & Yue, C. (2012). Vehicle Type Classification Using PCA with Self-Clustering. *2012 IEEE International Conference on* (pp. 384–389). ICMEW.

[12] Bao, T., Nam, G. H., & Byung Ryong, L. (2009). Vehicle detection and recognition for automated guided vehicle. *ICCAS-SICE* (pp. 671–676).

[13] Jun-Wei, H., Li-Chih, C., Duan-Yu, C., & Shyi-Chyi, C. (2013). Vehicle make and model recognition using symmetrical SURF. *Advanced Video and Signal Based Surveillance (AVSS)* (pp. 472–477).