*ICIC Express Letters*
Volume **12**, Number **6**, June **2018**

ICIC International ©2018 ISSN 1881-803X
pp. **575–581**

# RESEARCH OF PRECEDING VEHICLE IDENTIFICATION BASED ON HAAR-LIKE FEATURES AND ADABOOST ALGORITHM

Yongtao Liu, Jie Qiao and Lang Wei

School of Automobile
Chang'an University
Middle-section of Nan'er Huan Road, Xi'an 710064, P. R. China
liuyongtao86@163.com; 850573553@qq.com; qch_1@chd.edu.cn

Received November 2017; accepted February 2018

Abstract. *In order to improve the efficiency of the preceding vehicle identification, a preceding vehicle identification algorithm combined with Haar-like features and Adaboost algorithm is proposed. Based on massive amounts of offline training sample set, effective vehicle contour and texture characteristics are extracted, and Haar-like characteristics are used to describe the goal. Meanwhile, Adaboost machine learning algorithm is adopted to train classifier, the sample characteristics of cascade classifier are built, and the test object is used to detect the vehicle existence. The experimental result shows that the algorithm based on Haar-like features and Adaboost algorithm in the paper has a high detection rate of above 91%, and average detection speed of 28ms/f. It can well adapt to the uncertain factors such as environmental interference and vehicle type, improve the robustness of the preceding vehicle detection, and also meet the requirement of the safe driving in the longitudinal dimensionality.*

**Keywords:** Haar-like feature, Adaboost, Training sample set, Identification

1. **Introduction.** With the rapid development of social economy and the rapid increase of vehicles in China, the number of road traffic accidents has always been high. According to the statistics of road traffic management department, the vehicle collision is one of the main forms of major road traffic accidents. Therefore, detecting the preceding vehicle longitudinal position in real time with Advanced Driver Assistance Systems (ADAS) has an important practical significance on improving the safety of running vehicles.

There are several vehicle identification methods which are based on the vehicle shadow extraction [1], the rear mathematical modeling [2], and the vehicle contour recognition [3], etc. The identification method based on vehicle shadow is greatly influenced by the outside light factors. The identification method based on the rear of the mathematical model obtains deep information of the image through the corresponding point calibration, but subjected to the equipment and objective reasons of calibration, accurate transformation matrix between the coordinate systems is unable to get, and the applicability is restricted. The method based on Adaboost machine learning algorithms [4] is strongly immune to interference from the outside environment and has good robustness, but the sample training time of this method is long and the false alarm rate is high.

The conventional methods based on Adaboost and Haar-like features for vehicle recognition have the problem of poor recognition performance based on cascade Adaboost classifier as well as the problem of much time consumed for training traditional Adaboost. The method proposed in the paper uses integral image method to extract the extended Haar-like features, and uses improved Adaboost classifier method to select a small number of critical features from a very large set of Haar-like features while training Adaboost. The proposed approach has better performance both in recognition and time consuming than traditional methods through the experiment, and shows promising perspective in vehicle recognition and rear-end collision prevention. For the paper, the vehicle rectangular features are extracted through the integral image method, and then the vehicle cascade classifier is constructed based on the improved Adaboost algorithm and characteristic samples. Meanwhile, the preceding vehicle identification is done combined with Haar-like features and Adaboost algorithm. At last, the experiment of preceding vehicle identification is performed with the proposed approach.

2. **Vehicle Rectangular Features Extraction.** The function of machine learning is to use the known training samples to detect the unknown data. As training samples, the target feature extraction plays an important role in target recognition. Compared with extracting targets by using the pixel values, Haar-like rectangular feature is a rapid, simple effective way for vehicle contour and texture features extraction, it mainly because of using Haar-like rectangular features can reduce the differences between the similar samples, and its calculating speed is higher than that based on pixel values. When using Haar-like rectangular features to scan image, traversal search is done in the unit of the pixel size in the basic area. Therefore, it will produce a lot of features when scanning an image, resulting in a decline in calculating speed.

Rainer Lienhart puts forward a kind of extensional Haar-like rectangular feature [5], also a fast calculating method aiming at extensional rectangular features is proposed. Extensional Haar-like rectangular characteristics are shown in Figure 1, with a total of 4 categories and 15 kinds of features to describe the edge and structure characteristics of the vehicle.

For the vehicle samples of $32 \times 32$ pixels, there are at least tens of thousands of traditional rectangular features, if every rectangular feature is calculated, the characteristics

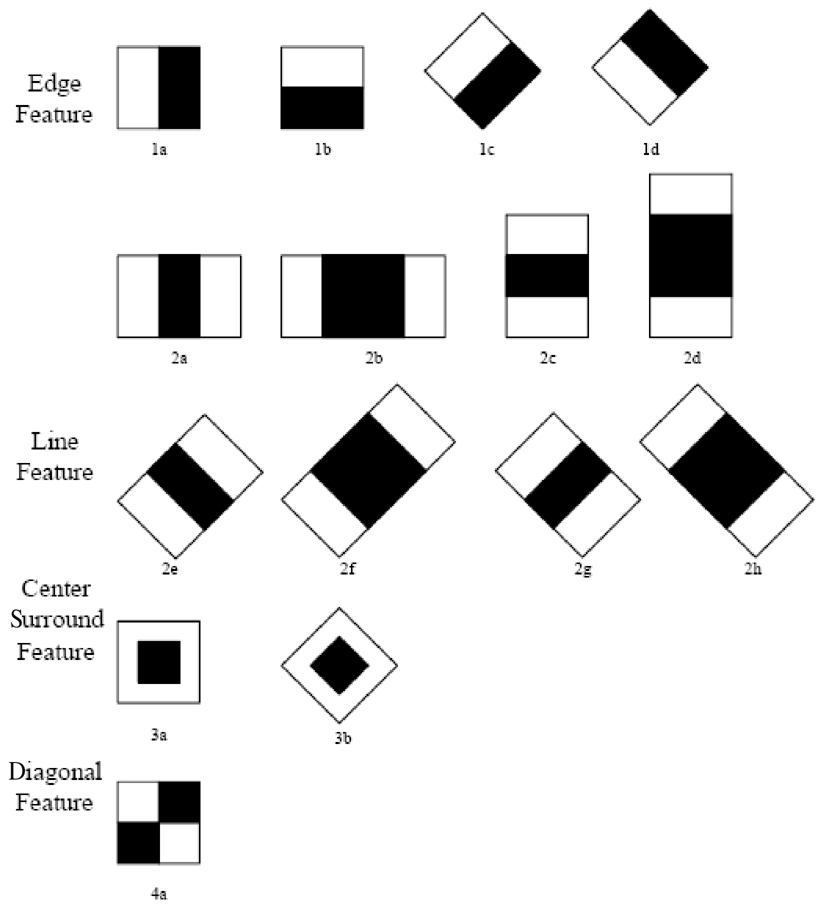


FIGURE 1. Extensional Haar-like rectangular features

volume will be enormous, so the integral image method is adopted to improve the efficiency of the algorithm in this paper. In actual application, the first step is to convert image pixels, and each point after conversion represents the sum of the pixel values from the upper left corner to the rectangular area, and the second step is to extract integral image. Equation (1) shows the integral image calculation method.

$$i(x,y) = \sum_{x' \le x, y' \le y} i(x', y') \tag{1}$$

where $i(x,y)$ represents the pixel integral value of point $(x,y)$, and $i(x',y')$ is the pixel value of $(x,y)$ in the original image.

## 3. Construction of Vehicle Cascade Classifier Based on Offline Training.

**3.1. Classifier training based on Adaboost algorithm.** Adaboost algorithm is used to train the strong classifier in this paper. Adaboost classifier extracts a corresponding effective characteristic by each loop [6], each valid characteristic generates a corresponding weak classifier, and finally, many weak classifiers are combined into a strong classifier. Besides, strong classifier training steps established in this paper are as follows.

Step1, building sample set, $\{(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)\}$. When the negative sample $Y$ is 0 and the positive sample $Y$ is 1, all images containing the target vehicle are classified as the positive samples, and all images containing the surrounding environment of the vehicle are classified as the negative samples. For each image, its eigenvalue is stored in the eigenvector $X$, $X = \{x_1, x_2, \ldots, x_n\}$. Each element in the eigenvector $X$ represents the eigenvalue of the image.

Step2, generating weak classifier. Each Haar-like feature can generate a weak classifier, and the specific form is as follows,

$$h_j(x) = \begin{cases} 1 & p_j f_j < p_j \theta_j \\ 0 & \text{else} \end{cases} \tag{2}$$

where $h_j$ represents the weak classifier, $f_j$ represents the *No.* $j$ eigenvalue, $\theta_j$ is the threshold, $p_j$ is the parity validation factor, and Adaboost can automatically select a threshold to separate positive and negative samples.

Step3, initializing weights. When $y_i$ takes 1 as positive sample, the weighting factor is,

$$w_{t,j} = \frac{1}{2l} \tag{3}$$

When $y_i$ takes 0 as negative sample, the weighting factor is,

$$w_{t,j} = \frac{1}{2m} \tag{4}$$

where $l$ represents the number of positive samples, and $m$ represents the number of negative samples. For each loop body $t = 1, 2, \ldots, T$ ($T$ is the number of iterations, determining the number of weak classifiers), normalize weights as follows,

$$w_{t,j} = \frac{w_{t,j}}{\sum\limits_{j=1}^{n} w_{t,j}} \tag{5}$$

Step4, for the *No.* $j$ characteristic, calculating the generating weak classifier $h_j$, and its error relative to the current weight is derived as follows,

$$\varepsilon_j = \sum_{i=1}^{n} w_i \left| h_j(x_i) - y_i \right| \tag{6}$$

Step5, updating single sample weight.

$$w_{t+1,i} = w_{t,i} \beta^{1-e_j} \tag{7}$$

Finally, the strong classifier is as follows,

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \log(1/\beta_t) h_t \geq \frac{1}{2} \sum_{t=1}^{T} \log(1/\beta_t) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

3.2. **Construction of cascade classifier based on characteristic samples.** Most of the image areas usually do not contain the target vehicle, if a high-performance classifier is used to test the target directly, a lot of time will be wasted in the non-vehicle areas. Therefore, a cascade classifier is adopted in this paper to quickly rule out the non-vehicle areas to improve the target detection speed.

As shown in Figure 2, the classical waterfall cascade classifier is applied in this paper, each layer is an Adaboost classifier trained from Adaboost algorithm, and each Adaboost classifier contains several weak classifiers. The sample figure to be identified is tested by cascade classifier layer by layer, if it is found as negative sample in any layer, the strong classifier in the back cannot pass, and the recognition efficiency of the positive samples can be improved eventually.

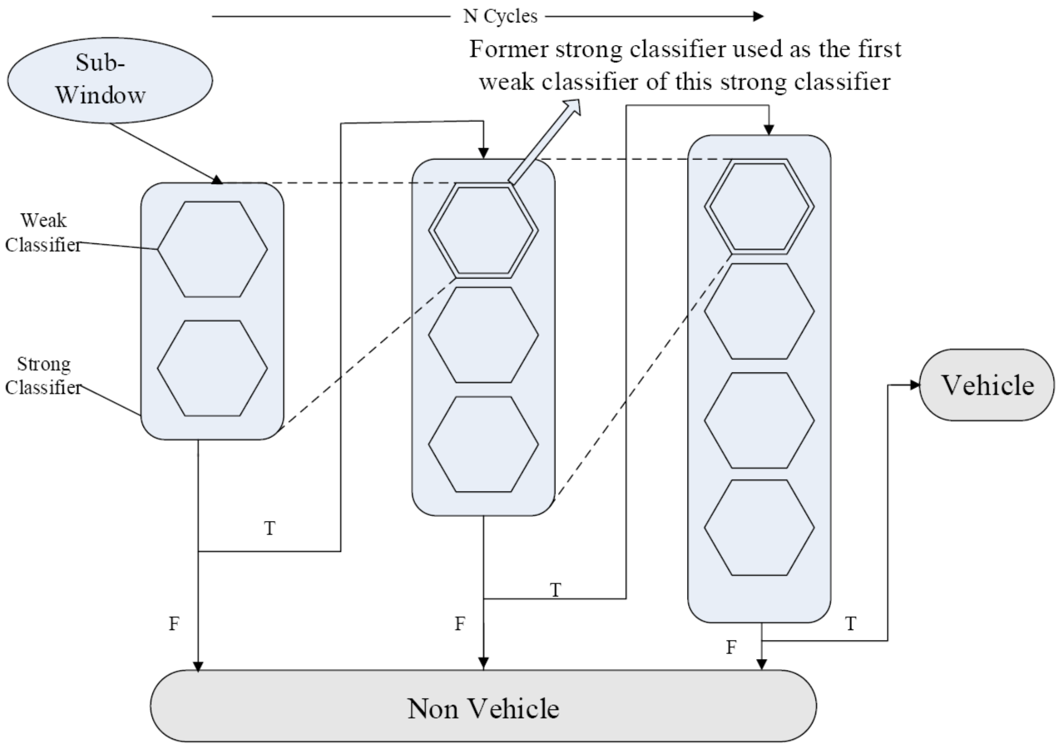


FIGURE 2. Training process of cascade classifier

When using cascade classifier, the detection rate is expected to be high, and the false alarm rate is as small as possible. The detection rate and false alarm rate on each layer of the classifier can be got as long as the expected detection rate and false alarm rate are known. Set the false alarm rate of cascade classifier as follows,

$$F = \prod_{i=1}^{K} f_i \tag{9}$$

where $F$ is the total false alarm rate, $K$ is the number of layers in the classifier, and $f_i$ is the false alarm rate of each layer. Equation (10) shows the detection rate.

$$D = \prod_{i=1}^{K} d_i \tag{10}$$

where $D$ is the total detection rate, $K$ is the number of layers in the classifier, and $d_i$ is the detection rate of each layer.

Considering the reliability of the vehicle detection and the real time of the algorithm, 8-layer cascade classifier is adopted in this paper, finally the detection rate is above 0.9, the error detection rate of each layer is 0.5, and the detection rate of each layer is above 0.99.

4. **Preceding Vehicle Identification Combined with Haar-Like Features and Adaboost Algorithm.** The preceding vehicle identification algorithm combined with Haar-like features and Adaboost proposed in this paper is divided into two steps: offline training and online identification. For the offline training, a sample set is established, the Haar-like features are extracted on a large number of vehicle samples using the method mentioned in the above section, and the cascade classifier is built on the basis of Adaboost algorithm. For the online identification, the features of the samples in the sample library are extracted one by one, and the vehicle existence is detected with putting the characteristics of Haar-like into Adaboost cascade classifier. The algorithm structure is shown in Figure 3.

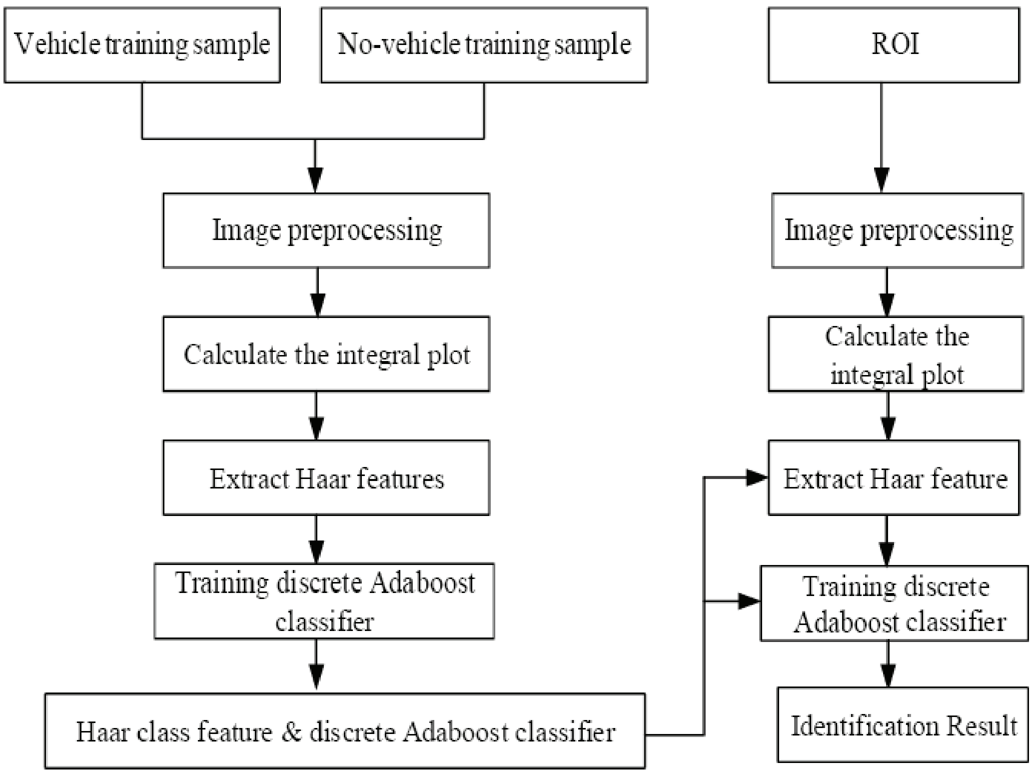


Figure 3. Structure of the vehicle identification algorithm

Training samples are divided into positive samples and negative samples, positive samples are the images containing target vehicle samples with different sizes and locations. The rest images which only contain the background areas are the negative samples; however, these images should contain the environment that often appears around the vehicles. In this paper, 500 positive samples and 1500 negative samples are collected in the training sample set, as shown in Figure 4(a) and Figure 4(b).

In order to ensure the numbers of the characteristic matrix are consistent, the invariance of matrix must be used to do normalization on the images, the transformation function is derived from the matrix, and this function can be adopted to normalize the images of the sample set. Besides, the sample is normalized to a sample set in the size of $30 \times 30$ pixel. It also means that the vehicle cascade classifier can only identify the target vehicles in the

(a) Part of the positive sample set (b) Part of the negative sample set

FIGURE 4. Training sample set

size of $30 \times 30$ pixels in the process of the actual vehicle identification; therefore, during the identification process, the Gaussian image pyramid is used to scale the vehicles to the range that the cascade classifier can recognize.

5. **Experimental Test.** In order to verify the effectiveness and feasibility of the vehicle identification algorithm proposed in this paper, the off-line experiment method was adopted to do the test. 100 frames from the vehicle video on the highway were selected as the test sample, and part of the target vehicle identification results were shown in Figure 5. Experimental results show that the accuracy rate of the vehicle detection algorithm proposed in this paper is above 91%. Compared with the traditional vehicle identification algorithm based on Adaboost algorithm, the accuracy rate has increased by 10%, the average detection speed is 28ms/f. The effects of the driveway shadow on both sides of the roads and other vehicles which are not in the same lane can be avoided. As a result, the detection robustness of the system has improved a lot.

FIGURE 5. Identification results of the preceding target vehicle

6. **Conclusions.** Based on massive amounts of offline training sample set, effective vehicle contour and texture characteristics are extracted in this paper. Adaboost machine learning algorithm is adopted to train the classifier, the characteristics sample of cascade classifier is built, and the test object is used to detect the vehicle existence. The experimental result shows that the algorithm based on Haar-like features and Adaboost algorithm in the paper has a high detection rate of above 91%, and an average detection speed of 28ms/f. The effects of the driveway shadow on both sides of the roads and other vehicles which are not in the same lane can be avoided. For further studies, more road environmental factors must be considered to optimize system performance.

**Acknowledgment.** This work is partially supported by the Fundamental Research Funds for the Central Universities in China (No. 310822171116 and No. 310822171001). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## REFERENCES

[1] N. Al-Najdawi, A. Abu-Roman, S. Tedmori and M. Al-Najdawi, An adaptive approach for real-time road traffic congestion detection using adaptive background extraction, *International Arab Journal of Information Technology*, vol.13, no.6, pp.1075-1083, 2016.

[2] S. Kamkar and R. Safabakhsh, Vehicle detection, counting and classification in various conditions, *IET Intelligent Transport System*, vol.10, no.6, pp.406-413, 2016.

[3] C. Zhang, Y. Wang, X. Lu and K. Wang, Front-vehicle detection algorithm based on hypothesis and verification of improved HOG feature, *Journal of Electronic Measurement and Instrumentation*, vol.29, no.2, pp.165-171, 2015.

[4] Y. Hu and Y. Liu, Research and implementation of license plate recognition system based on FPGA, *Electronic Measurement Technology*, vol.38, no.12, pp.54-58, 2015.

[5] S. Zhang, C. Bauckhage and A. B. Cremers, Efficient pedestrian detection via rectangular features based on a statistical shape model, *IEEE Transactions on Intelligent Transportation Systems*, vol.16, no.2, pp.763-775, 2015.

[6] J. Hu, Syncretize automated detection of driver fatigue based on AdaBoost classifier with EEG signals, *Frontiers in Computational Neuroscience*, vol.11, no.72, pp.1-10, 2017.