LINEAR SVM CLASSIFICATION USING BOOSTING HOG FEATURES FOR VEHICLE DETECTION IN LOW-ALTITUDE AIRBORNE VIDEOS

Xianbin Cao[1,2], Changxia Wu[1], Pingkun Yan[3], Xuelong Li[3]

[1]University of Science and Technology of China, Hefei, 230026, P.R. China;
[2]BeiHang University, Beijing, 100083, P.R. China;
[3]Center for OPTical Imagery Analysis and Learning (OPTIMIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shanxi, P.R. China

ABSTRACT

Visual surveillance from low-altitude airborne platforms has been widely addressed in recent years. Moving vehicle detection is an important component of such a system, which is a very challenging task due to illumination variance and scene complexity. Therefore, a boosting Histogram Orientation Gradients (boosting HOG) feature is proposed in this paper. This feature is not sensitive to illumination change and shows better performance in characterizing object shape and appearance. Each of the boosting HOG feature is an output of an adaboost classifier, which is trained using all bins upon a cell in traditional HOG features. All boosting HOG features are combined to establish the final feature vector to train a linear SVM classifier for vehicle classification. Compared with classical approaches, the proposed method achieved better performance in higher detection rate, lower false positive rate and faster detection speed.

Index Terms—Vehicle detection, boosting HOG feature, linear SVM, urban environment

1. INTRODUCTION

During the past two decades, a number of methods for detecting vehicles in low-altitude airborne platforms under urban environment have been developed [1][2]. The existing methods can be roughly classified into two categories. The first category is based on image processing techniques. For example, Shastry and Schowengerdt [3] proposed a frame-by-frame video registration technique using a KLT feature tracker to automatically determine control points correspondence. Angel and Hickman [4] used the location, orientation and scale information of vehicles predicted in a reference frame to align the adjacent frames. This method performs reasonably well on freeway segments, where the speeds of vehicles are generally uniform. On the other hand, machine learning based detection methods have been becoming more and more

popular recently [5][6]. Zhao and Nevatia [5] combined several appearance features including the shadows of vehicles to detect moving and static vehicles. A Bayesian network, which is trained by vehicle and non-vehicle samples, is setup to classify the objects. Cucchiara et al. [7] provided a traffic monitoring method by combining a pixel-based processing technique and a knowledge-based reasoning method, where occlusion reasoning helps to estimate each vehicle in the occlusion region effectively. However, most of the existing methods are either time consuming or only applicable to vehicle detection in the highway situation, which is a relatively simple environment.

In this paper, we propose a new and efficient method for vehicle detection in low-altitude airborne platform under
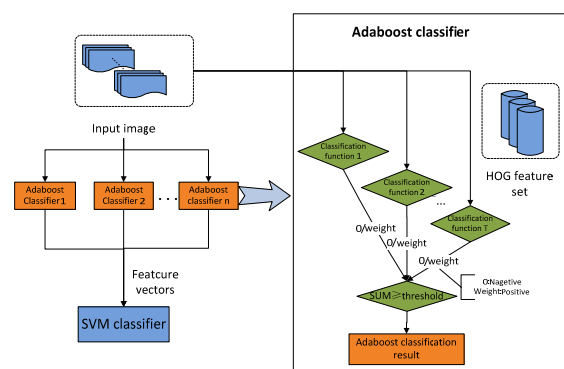

Fig.1. The flow chart of the SVM classifier: input images are divided into cells, each of which corresponds to a strong classifier trained by adaboost using HOG features. And all adaboost outputs are combined to establish the final SVM feature vector.

urban environment. In our work, a boosting histogram of oriented gradients (boosting HOG) feature is proposed to use together with a linear support vector machine (SVM) for vehicle detection. The output of the adaboost training on each cell is a strong classifier and the strong classifiers from all the cells of an input image patch are combined to construct a feature vector. A linear SVM classifier is then trained over the feature vectors for the final vehicle classification. Our experimental results demonstrated that

compared with other algorithms, the proposed method can reach better performance in terms of higher detection rate, lower false positive rate and faster detection speed.

2. VEHICLE DETECTION

In our work, two most important parts of the method are feature description and classifier training. The overall flowchart of our method using linear SVM classifier with boosting HOG features is shown in Fig. 1. The detail processing is presented in the following subsections.

2.1. Boosting HOG features

The key component in vehicle detection is to select good feature description that is able to distinguish vehicles from other objects. HOG feature shows better performance in characterizing object shape and appearance and it is not sensitive to illumination change. However, the high dimensionality of the HOG features will lead the training and classification to slow down dramatically. We propose boosting HOG features for reducing the dimensionality of the feature vectors to speed up the detection process. The classification is performed by using the discrete adaboost algorithm. In our proposed method, as the HOG feature is a histogram with bins indicating local gradient distribution, those bins are considered as a set of weak classifiers. By applying thresholds to all the bins, each bin gives a binary output, which is taken as the output of the corresponding weak classifier. Let $H_N=\{b_1, \ldots, b_N\}$ denote a HOG descriptor with N bins. By comparing each bin b_i to its corresponding threshold θ_i , we can obtain the weak classifiers corresponding to the bins. Formally, let $h_{ji}(x)$ denote the ith weak classifier, which can be written as:

$$h_{ji}(x) = \begin{cases} 1, & \text{if } p_i f_{ji}(x) < p_i \theta_{ji}; \\ -1, & \text{otherwise.} \end{cases} \quad (1)$$

where x represents input image; $f_{ji}(x)$ denotes the value of the i th bin in the jth cell,; θ_{ji} is a threshold used to make a decision for $f_{ji}(x)$; p_i is a polarity parameter used to change the direction of the inequality, which can be either -1 or +1. Each bin in a cell can be considered as a weak classifier, and by traditional discrete adaboost training, all the bins in a cell can be gathered to form a strong classifier. The output of the corresponding adaboost classifier is used as an entry to establish a boosting HOG feature vector. An example with 4 bins in each cell is shown in Fig.2.

Like the classical HOG features, cells are grouped together to form bigger blocks over a larger spatial regions. To achieve better detection performance, a multi-scale approach is employed. The algorithm starts densely scanning images with a sliding window to capture local appearance information to differentiate vehicles from other objects. However, only local information is inadequate to extract global shape features of the vehicles. Therefore, blocks in various sizes are used to achieve the aims of detecting vehicles quickly and efficiently. Blocks are obtained by scanning the samples with a sliding window. The step size in horizontal direction for scanning is called block-stride. Three types of blocks are carefully designed, as shown in Fig.3: I) *block-size=8 × 8, cell-size=4 × 4, block-stride=4*, and totally 21 blocks over a *32 × 16* region; II) *block-size=8 × 16, cell-size=4 × 8, block-stride=4* and a total of 15 blocks over a *32 × 16* region; III) *block-size=16 × 16, cell-size=8 × 8, block-stride=8* and there are 3 blocks over a 32 ×16 region. Due to the larger size of the blocks, the global information of vehicles can be obtained. Compared with the method proposed by Dalal and Triggs [8], where there are 9 bins in each cell, the HOG features of each block have the length of 36. On the other hand, since there are four cells in a block of all these three types and a cell corresponds to an adaboost output, a boosting HOG feature keeps only 4 dimensional for each block.

2.2. Linear SVM training

The output of the adaboost training on each cell is a strong classifier and the strong classifiers from all the cells of an input image patch are combined to construct a feature vector. A linear SVM classifier is then trained over the feature
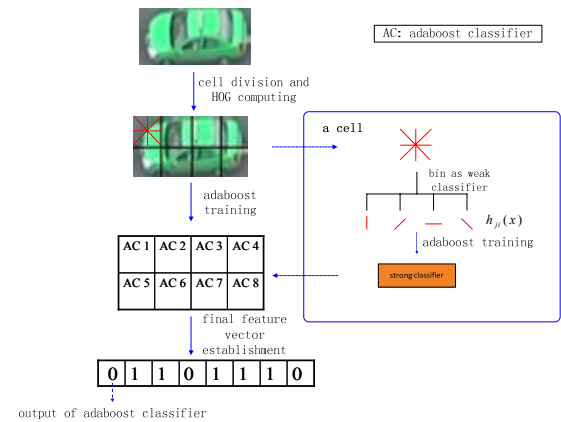

Fig. 2. The flow chart to obtain the boosting HOG feature.

vectors for the final vehicle classification. To train the classifiers to get good performance, 2048 positive samples are manually made and 1,000,000 negative samples were generated automatically by programs using image frames without vehicles. All the samples were scaled to the size of $32 \times 16 (width \times height)$ for training as shown in Fig. 4. In order to achieve better detection performance, various sized blocks are designed to train a series of adaboost classifier to detect vehicles in different sizes. All three types of block size are employed in constructing boosting HOG feature, and small block size which is rich in local information is first used to train boosting HOG feature. Since some easy cases can usually be correctly classified, the training process

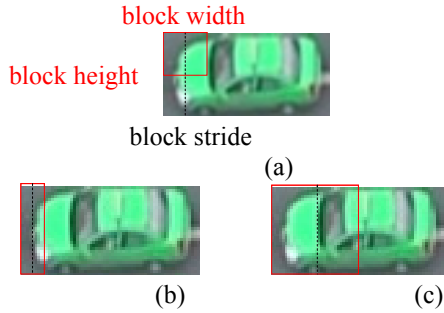should focus on the difficult samples. In order to do so, the

block width

block height

block stride

(a)

(b)   (c)

Fig. 3. three types of block with the same positive sample, (a) corresponds to type I), (b) corresponds to type II) and (c) corresponds to type III): red solid lines are edges of block size and black dashed lines are block stride in horizontal direction.

false positive and false negative samples after the first round of training are selected to train the classifiers one more time using type II blocks. Type III blocks are adopted to train boosting HOG feature at last and all trained boosting HOG features are combined to train the final SVM classifier.

Fig. 4. Some positive and negative samples used for training the classifiers.

The procedure of SVM training for moving vehicle classification is summarized as follows.

Step 1) Train the adaboost classifier using the training samples with type I block size and compute the boosting HOG features.

Step 2) With the obtained features, train the SVM classifier with all the samples. Test the trained SVM on the same training samples. Find all the false positive and false negative samples.

Step 3) Use the false positive samples and false negative samples to train the adaboost classifiers with type II blocks as in step 1).

Step 4) Train the adaboost classifier with type III blocks similar as in step 1).

Step 5) Combine the obtained adaboost classifiers to build the feature vector and train the final linear SVM classifier to classify vehicles.

Some vehicle detection results obtained by the SVM classification are shown in Fig. 5, where the detected vehicles are marked with red rectangles. As we can see, most of the vehicles have been correctly detected.

3. EXPERIMENTS AND DISCUSSION

In this section, experimental results of the proposed method on urban traffic videos are presented to show the performance of our approach.

3.1. Datasets and Experiment Platform

The experiments were performed on more than 3 hours of video in urban traffic environment and highway traffic scenes. Some snapshots of the videos with vehicle detection
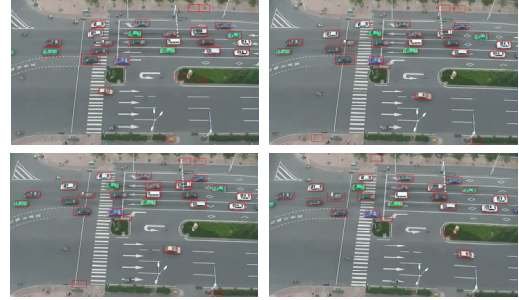
Fig. 5. Output of the SVM classification in different frame

results are shown in Fig. 6. The test videos of urban traffic were captured around the height of 60 to 90 meters using a digital video camera (Sony-DCR-HC21E). And the video of highway is captured by an unmanned aerial vehicle. Ten urban video segments with 300 frames each were manually labeled in our experiment. The size of the video frames is *720x400* pixels. The experiments were carried out on a computer with 3.2 GHz CPU and 4 GB double data rate RAM.

3.2. Performance evaluation

The detection speed in terms of frames per second (fps), the detection rate (DR), and the false positive rate (FPR) were used to quantitatively evaluate the performance of our system. The DR and FPR are defined as

$$DR = \frac{TP}{TP + MP} \text{ and } FPR = \frac{FP}{TP + FP},$$

respectively, where TP is the average number of detected regions corresponding to vehicles, MP is the average number of missed vehicles, and FP denotes the number of detected regions which are actually not vehicles.

3.3. Experimental Results and Comparison

In order to objectively evaluate the performance of our method, the performance of three other commonly used algorithms has been included for comparison. The first algorithm is the SIS (simplistic image subtraction) method [9]. This method simply subtracts all the pixels in the first frame from the ones in the second frame. ISFR (Image subtraction method with frame registration) performs image subtraction after frame registration [3], which is extended from the simplistic subtraction method but can compensate camera motion. SCOHF (SVM classification with the original HOG features) uses linear SVM classifier but with the original HOG features to perform vehicle classification.

All the experiments were carried out on the same hardware platform as described above. The detection performance of the four methods is presented in Table 1. The results show that our method achieved the best performance compared with the other three methods. Although the simplistic image subtraction method shares the advantage of fast detection

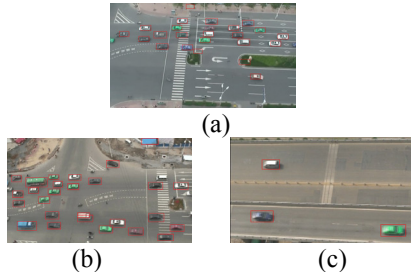(a)

(b)   (c)

Fig. 6. (a) (b) The vehicle detection results in urban traffic scenes; (c) The vehicle detection result in highway traffic scene.

speed with the proposed method, it produces a large number of false alarms. On the other hand, the registration based method, which takes too much computational time, cannot satisfy the need of real-time applications on resource-limited platforms. In contrast with the SVM classification method using the original HOG features, the great reduction

Table 1. The comparison of the detection speed of the four methods.

	SIS	ISFR	SCOHF	Our method
DR	75%	80%	88%	90%
FPR	80%	15%	13%	10%
Detection Speed (fps)	>30	0.18	0.15	4.76

of feature dimensionality in our proposed method is able to speed up the detection process significantly, from 0.15fps to 4.76fps. Moreover, our method achieved higher detection rate with fewer false positives compared to the other three methods mainly due to the reason that the used linear SVM classifier, which is trained using the boosting HOG feature vectors, has higher classification ability.

As Table 1 shows, the SVM classification based on boosting HOG features can achieve comparable performance with that based on original HOG features. On the other hand, the boosting HOG feature vector has $84+60+12=156$ dimensions compared with $21 \times 4 \times 9 = 756$ in the original HOG features, and thus the SVM classifier based on boosting HOG leads to a much simplified feature vector. Therefore, in our method, the linear SVM classifier is speeded up by using more compact feature vectors and outperforms both of the other two methods.

4. CONCLUSIONS AND DISCUSSIONS

In this paper, we have presented a method to detect moving vehicles. A linear SVM classifier is trained for moving vehicle detection in airborne videos. For the sake of lowering the time cost of the original HOG feature, boosting HOG features are proposed and used for SVM training and classification. Our experimental results showed that we achieved the detection accuracy of 90% with fast detection speed. In our future work, we will design an integrated classification solution for vehicle detection by mining various features to further improve the detection performance of the proposed method.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] S.M. Khan, H. Cheng, D. Matthies, H. Sawhney, "3D Model Based Vehicle Classification in Aerial Imagery," *IEEE Conference on Computer Vision and Pattern Recognition*, 2010

[2] B. Coifman, M. McCord, R.G. Mishalani, M. Iswalt and Y. Ji, "Roadway Traffic Monitoring from an Unmanned Aerial Vehicle," *IEEE International Conference on Intelligent Transportation Systems*, vol. 153, pp.11-20, 2006.

[3] A.C. Shastry, R.A. Schowengerdt, "Airborne Video Registration and Traffic-flow Parameter Estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, issue 4, pp.391-405, Dec. 2005.

[4] A. Angel and M. Hickman, "Methods of Analyzing Traffic Imagery Collected From Aerial Platforms," *IEEE Transaction on Intelligent Transportation Systems*, vol. 4, no. 2, June 2003.

[5] T. Zhao and R. Nevatia, "Car Detection in Low Resolution Aerial Image," *IEEE International Conference on Computer Vision*, vol.1, pp. 710-717, 2001.

[6] H. Tao, H. Sawhney and R. Kumar, "Object Tracking with Bayesian Estimation of Dynamic Layer Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.1, pp.75–89, January 2002.

[7] R. Cucchiara, M. Piccardi and P. Mello, "Image Analysis and Rule-based Reasoning for a Traffic Monitoring System," *IEEE Transactions on Intelligent Transportation System,* vol.1, no. 2, pp.119–130, 2000.

[8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp. 886–893, June 2005.

[9] A. Angel, M. Hickman, "Methods of Traffic Data collection Using Aerial Video", *IEEE 5th International Conference on Intelligent Transportation Systems*, 3-6 September 2002.