# A New Approach to Track Multiple Vehicles With the Combination of Robust Detection and Two Classifiers

Weidong Min, Mengdan Fan, Xiaoguang Guo, and Qing Han

***Abstract*— It plays an important role to accurately track multiple vehicles in intelligent transportation, especially in intelligent vehicles. Due to complicated traffic environments it is difficult to track multiple vehicles accurately and robustly, especially when there are occlusions among vehicles. To alleviate these problems, a new approach is proposed to track multiple vehicles with the combination of robust detection and two classifiers. An improved ViBe algorithm is proposed for robust and accurate detection of multiple vehicles. It uses the gray-scale spatial information to build dictionary of pixel life length to make ghost shadows and object's residual shadows quickly blended into the samples of the background. The improved algorithm takes good post-processing method to restrain dynamic noise. In this paper, we also design a method using two classifiers to further attack the problem of failure to track vehicles with occlusions and interference. It classifies tracking rectangles with confidence values between two thresholds through combining local binary pattern with support vector machine (SVM) classifier and then using a convolutional neural network (CNN) classifier for the second time to remove the interference areas between vehicles and other moving objects. The two classifiers method has both time efficiency advantage of SVM and high accuracy advantage of CNN. Comparing with several existing methods, the qualitative and quantitative analysis of our experiment results showed that the proposed method not only effectively removed the ghost shadows, and improved the detection accuracy and real-time performance, but also was robust to deal with the occlusion of multiple vehicles in various traffic scenes.**

***Index Terms*— Multiple vehicles tracking, robust detection, detection accuracy, support vector machine (SVM), convolutional neural network (CNN), occlusion.**

## I. Introduction

INTELLIGENT vehicle is a hot research topic in recent years. Providing early-warning signals, monitoring and exercising control are some examples of major research in intelligent vehicles [1]. Video-based traffic sign detection, tracking, and recognition are the important components for the intelligent transport systems [2]. Vision based vehicle detection and tracking methods are important in intelligent vehicle. Vehicle detection methods are mainly divided into foreground detection method, feature detection method based on prior knowledge, and classifier based method [3]. The purpose of vehicle tracking is to predict the location of the vehicle to reduce the search space of the vehicle detection and save computation time. Common tracking algorithm has Meanshift, Particle Filter and Kalman Filter algorithms [4]. The development of tracking method is relatively mature. However, the object detection method has a great impact on the object tracking results.

Vehicle tracking in condition of occlusion, appearance or illumination change has been a challenging task over decades [5]. Visual tracking is a central topic in computer vision. However, the accurate localization of target object in extreme conditions, such as occlusion, scaling, illumination change, and shape transformation, still remains a challenge [6]. In this paper, the situation that the fast moving pedestrians and non-motor vehicles in traffic video could be classified as foreground vehicles is called the interference of pedestrians and non-motor vehicles. The major difficulties of tracking multiple vehicles are listed as follows.

1) Occlusion and separation of multiple vehicles could lead to tracking failure.
2) Interference of pedestrians and non-motor vehicles reduces the accuracy of tracking results.
3) Occlusion of multiple vehicles could result in loss of the tracking rectangle and vehicle label drift. In other words, an already detected and labeled vehicle loses tracking and is detected again as a new vehicle and henceforth re-labeled using another vehicle label.

It is necessary to solve the above problems and accurately calculate the position and size of the vehicles to achieve stable multiple vehicles tracking. Some methods have been studied for solving the above problems. For instance, Zhang *et al.* [7] researched the uncertainty bound for the multiple Gaussian functions, termed Multiple Gaussians Uncertainty (MGU), which generalizes the uncertainty principle for the single Gaussian function. It pushed forward the research of target detection. In [8], a system selectively fed the complementary data emanating from the two vision sensors to different algorithmic modules which together implemented three sequential components. Although some methods partly solved the problems, the existing methods still have two deficiencies. The first one is that the detecting results got by the existing methods

Manuscript received February 28, 2017; revised July 25, 2017; accepted September 10, 2017. Date of publication December 21, 2017; date of current version December 26, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61762061 and Grant 61662044 and in part by the Natural Science Foundation of Jiangxi Province, China, under Grant 20161ACB20004. The Associate Editor for this paper was Q. Wang. *(Corresponding author: Weidong Min.)*

W. Min, M. Fan, and Q. Han are with the School of Information Engineering, Nanchang University, Nanchang 330031, China (e-mail: minweidong@ncu.edu.cn).

X. Guo is with the School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin 300387, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2756989

are generally incomplete or unclear. The second one is that the occlusion and interference caused wrong multiple vehicles tracking [9]. For instance, an approach for tracking varying number of objects through both temporally and spatially significant occlusions was presented in [10]. It is noteworthy that the ViBe is a relatively stable and light-weighted algorithm among various methods, which is widely used in multiple vehicles tracking. The ViBe is of low complexity and low memory usage. It has better detection performance than other background subtraction methods in many literatures [11]. However, it also has the above two deficiencies. Especially, its first deficiency is mainly reflected in the appearance of the ghost shadows at times [12].

To alleviate the above problems, a new approach to track multiple vehicles with combination of robust detection and two classifiers is proposed in this paper. The two classifiers method is proposed because the ViBe only is difficult to distinguish two different moving vehicles that block and interfere with each other. The machine learning classifiers can further classify the features of the target and solve the occlusion problem. The main contributions of this paper are summarized as follows.

1. An improved ViBe algorithm for robust and accurate detection of multiple vehicles is proposed. It uses the gray-scale spatial information to build dictionary of pixel life length to make the ghost shadows and target's residual shadows quickly blended into the samples of the background. Because the fixed threshold does not distinguish the foreground very well, it also combines with the OTSU [13] method to set the dynamic threshold.

2. A method using two classifiers to attack the problem of failure to track vehicles with occlusions is designed. It classifies tracking rectangles with confidence values between two thresholds through combining local binary pattern (LBP) with SVM classifier and then using CNN [14] for the second time to remove interference areas between vehicles and other moving objects.

The rest of the paper is organized as follows. Section 2 gives a review of the related works. Section 3 describes our improved method for tracking vehicles accurately. Section 4 discusses our robust method to track multiple vehicles with combination of two classifiers, followed by the experiment results in section 5. This paper is concluded in section 6.

## II. Related Works

As mentioned above, vehicle detection methods are mainly divided into the foreground detection method, the feature detection method based on prior knowledge (such as symmetry, color and shadow, etc.), and the classifier based method [3]. In [15], an object detection system based on mixtures of multi-scale deformable part models was described. The system was able to represent highly variable object classes and achieved state-of-the-art results in the PASCAL object detection challenges. The key insight of Aggregated Channel Features [16] is that one may compute finely sampled feature pyramids at a fraction of the cost without sacrificing performance. For a broad family of features, features computed at octave-spaced scale intervals are sufficient to approximate features on a finely-sampled pyramid. A model for fine-grained vehicle classification based on deep learning was proposed to handle complicated transportation scenes [17]. This model consisted of two parts, vehicle detection model and vehicle fine-grained detection and classification model.

The current researches mainly focus on improving single method mentioned above. The existing methods still have two deficiencies. The first one is that the detecting results got by the existing methods are generally incomplete or unclear. The second one is that the occlusion and interference caused wrong multiple vehicles tracking. This paper combines the foreground detection method with the classifier based method to resolve those problems.

The foreground detection method is divided into the frame difference, optical flow and background subtraction method [18]. The frame difference method obtains the moving object contour by using the difference between two adjacent frames in the video sequences. It is suitable for the situation that has multiple moving vehicles or a moving video camera [19]. For instance, to overcome the limitations of the MeanShift method, a new approach was proposed through integrating the MeanShift algorithm and the frame difference method in [20]. In [21], an improved three-frame differencing algorithm and a foreground detection method combining background subtraction with the improved three-frame differencing were proposed. The disadvantage of this method is that it cannot extract the whole area of the object. Only the boundary can be extracted. Even more, this method depends on the choice of the time interval between frames. The fast moving target could be detected as two separate objects and the slow moving target could easily missed when the choice is not appropriate [22]. The optical flow method determines the position of each pixel by the change and correlation of the pixel intensity data in the image sequence [23]. A varied solution of the physics-based optical flow equation was studied for extraction of high-resolution velocity fields from flow visualization images in [24]. In [25], the implementation of a simple wavelet-based optical-flow motion estimator dedicated to continuous motions such as fluid flows was described. The optical flow method is not suitable for real-time processing due to high computational cost. In practical applications, because the optical flow method is affected by illumination, the optical flow field cannot be solved correctly [26]. The background subtraction method is the most commonly used motion detection method. It sets up background templates for image sequences. The target information is obtained by getting the difference between the current frame and the background model. The Gaussian Mixture Model (GMM) [27], codebook model [28] and ViBe are the mainstream approaches of the background subtraction method. The GMM gets good performance in nonlinear differential calculation and parameter space, but the calculation and parameter settings are complex. The codebook model adopts the method of quantization and clustering. It doesn't need to set the parameters, but it is sensitive to light and consumes a lot of memory [29].

The ViBe algorithm has no special requirements for the type of video stream, the color space and the scene type and needs less memory. It is a kind of light detection algorithm [10]. However, if the first few frames have moving object which is changed from static to dynamic or from dynamic to static, the ghost shadows appear. If there is a dynamic change in the background, the accuracy of the object detection will decline because the pixel difference is fixed. For these problems, some researchers have put forward their own improvements. For instance, in [30], a Belief Propagation approach for moving object detection using a 3D Markov Random Field (MRF) model was presented. This approach dealt with moving object detection problems like objects camouflaged by similar appearance to the background, or objects with uniform color that frame difference methods can only partially detect. The method in [31] generated an automated way to evaluate the team behavior of trainees in a delivery simulation course using video-processing techniques with emphasis on multiple people tracking. It transferred the person location problem during the occlusion into finding the local maximum points on a smooth curve, so that visual persons in the partial or complete occlusion could still be precisely captured.

The classifier based method is an important detection method. The original method used a single machine learning classifier to detect the object [32], which detected vehicles from a nighttime driving scene taken by an in-vehicle monocular camera. The method classified vehicles according to the features of the blob using support vector machine (SVM) (Support Vector Machine). The paper [33] constructed an original feature vector composed of 7 basic features and an extended feature vector composed of 17 features according to the basic features and semantic environment. The SVM was used to recognize the vehicle-borne LiDAR point clouds of street trees. Later, the hybrid classifier was adopted gradually. A hybrid vehicle detection scheme which integrated the Viola-Jones (V-J) and linear SVM classifier with HOG feature (HOG + SVM) methods was proposed for vehicle detection from low-altitude Unmanned Aerial Vehicle (UAV) images [34]. A moving region extraction method based on Gaussian model was used to reduce the scanning area of the window, excluded some background noise and improved test speed [35]. The action of cascading classification on samples achieved good results for the detection of moving vehicles. Recently, deep learning methods were applied to vehicle detection. Chen, et al. proposed a method that consists of detection and tracking modules using a region proposal network based on Convolutional Neural Network (CNN) feature maps to achieve a high level of robustness [36]. A denoizing-based CNN called DCNN was proposed in [37]. A CNN with one fully-connected layer was pre-trained first for feature extraction. After that, features of this fully-connected layer were used to pre-train a Stacked Denoizing Auto-Encoder (SDAE) in an unsupervised way. Then, the pre-trained SDAE was added into the CNN as the fully-connected layer.

In summary, the ViBe is a lightweight and robust algorithm among the foreground detection methods while the CNN is an advanced algorithm with high accuracy among the classifier based methods. The consideration of the proposed method in this paper is to fast track objects using the ViBe, and then use the CNN as an aid to further detect objects from complex environments. However, the ghost shadows appear at times when using the ViBe, while the CNN is time consuming, difficult to meet requirement of real-time vehicle detection. To alleviate the disadvantages of the ViBe and CNN, a new approach is proposed in this paper to track multiple vehicles with combination of the improved ViBe and two classifiers. The method tracks object vehicles using the improved ViBe, then uses two classifiers as an aid to further detect the objects from the complex environments. Using two classifiers enables the method to have both time efficiency advantage of the first classifier SVM and high accuracy advantage of the second classifier CNN.

## III. An Improved Method for Tracking Vehicles Accurately

### A. An Improved Tracking Method

Detecting results got by the existing methods are generally incomplete or unclear [9]. Failing to obtain all the characteristic pixels introduce holes in the moving entity. For instance, since the overlap of objects between two adjacent frames is hard to be detected using the frame difference method, it usually causes the cavities inside of objects. The ViBe method [10, 11] is a relatively stable and light-weighted algorithm among various methods, which deals with the above problems better. However, the ghost shadows appeared at times when using the ViBe. The fixed threshold of the ViBe may have a good performance on single background segmentation, but the accuracy of the foreground detection may be decreased if the background is in multi-modal scene.

Our proposed algorithm is improved on the basis of the original ViBe method to solve the above problems. The improvements will be discussed in details after we first provide a brief overview of the ViBe method in the following.

The random selection and neighborhood propagation mechanism of the ViBe are used to establish and update the template. The algorithm mainly consists of four aspects, i.e. background template definition, template initialization, foreground detection, and template updating.

1) Background template definition. First, $N$ background sample values for each pixel of the image sequence are established. Let's define the European color space value of position of $x$ pixel as $V(x)$, and $V_i$ as pixel values in the background sample space, i∈{1, $N$}. The background template is shown in formula (1).

$$B_G(x) = \{V_1, V_2, V_3, \ldots . V_N\} \tag{1}$$

2) Template initialization. Because similar pixels have similar spatial and temporal distributing characteristics, randomly selected pixel value has its background sample value from its M neighborhood. As shown in formula (2), $t$ represents time, while $N_G(x)$ represents the neighborhood pixels.

$$B_G^t(x) = \left\{V^t(x \,|x \in N_G(x))\right\} \tag{2}$$

3) Foreground detection. According to the corresponding background template $B_G(x)$ of each pixel, the location of the new pixels $V(x)$ is classified into the foreground or the background. Given a time $t$, the background template is represented as $B_G^{t-1}(x)$, while the

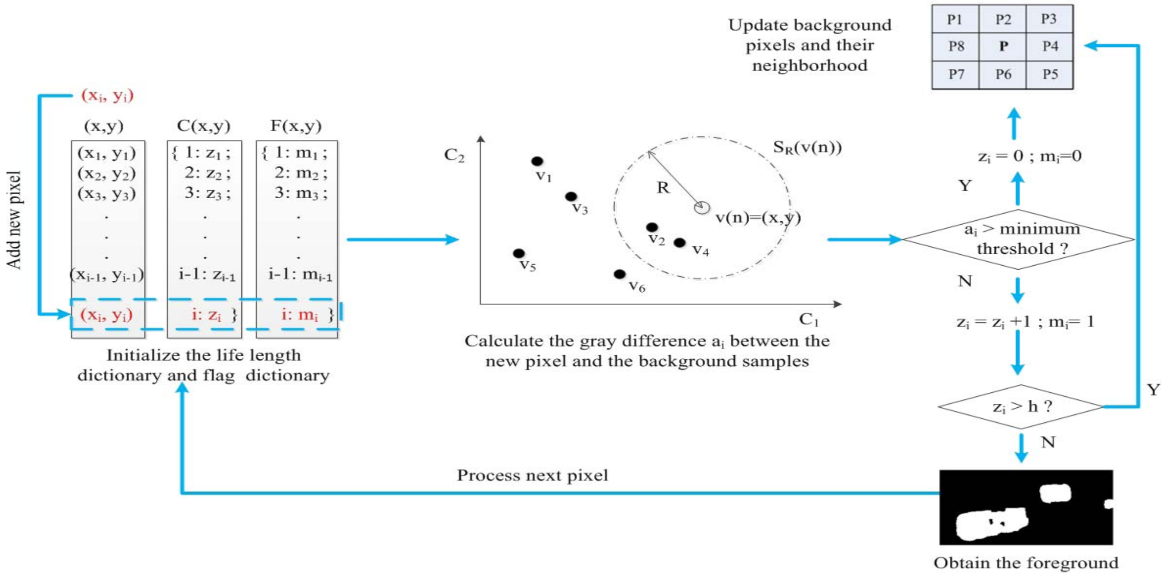


Fig. 1. The flowchart of reducing the ghost shadows of our improved ViBe method.

pixel value is $V^t(x)$. The foreground objects of an input image are segmented according to formula (3) where R is a fixed threshold for segmentation. If the number of candidate sets in the background is greater than the given minimum value, the new pixel is judged as a background and is updated into the template.

$$V^t(x) = \begin{cases} foreground \left| B_G^{t-1}(x) - V^t(x) \right| > R \\ background \left| B_G^{t-1}(x) - V^t(x) \right| < R \end{cases} \tag{3}$$

4) Background updating. Background is updated by random sampling method. When a pixel is classified as background, it has a probability of *1/φ* to update the background template. The position value of the neighborhood also has the probability of 1/$\varphi$ to be updated. The probability that such a sample value is not updated at time $t$ is $(N$-1$)/N$. Assuming that the time is continuous, after *dt* time, the probability that the sample value is still preserved is shown in the formula (4).

$$P(t, t+dt) = e^{-\ln\left(\frac{N}{N-1}\right)dt} \tag{4}$$

In this paper, an improved ViBe algorithm for detecting and tracking multiple vehicles accurately is proposed. It uses the gray-scale information to build life length dictionary of pixels to make the ghost shadows or object's residual shadows to be quickly blended into the samples of the background. It also combines with the OTSU method to set the dynamic threshold to ensure that the foreground detection is still accurate for multi-modal scene. More details are given in the following.

### B. Method of Reducing the Ghost Shadows

Our new method for reducing the ghost shadows and getting more complete detecting results is presented in the following. The life length dictionary is used to save the time information of the foreground. Background and neighborhood are updated for the second time according to the threshold. The detailed process steps are as follows.

1) Initialization of the algorithm. Set the number of sampling template as n, then establish the background template, finally define the life dictionary $C\ (x,\ y)$ and the flag bit dictionary $F\ (x,\ y)$ of foreground pixels $(x,\ y)$.
2) Analysis of the new pixel. Calculate the gray difference with n sample points.
3) Detection of the object. The number of pixels with gray difference less than the threshold R is recorded as the base number. If the base is greater than the given minimum threshold value, then the pixel is judged as background. After $C\ (i,j)$ and $F\ (i,\ j)$ are set as the initialization value 0, then go to step 4) and step 5). If the base number is less than the minimum threshold value, then the pixel is conservatively judged as foreground. We set $F\ (i,\ j)$ =1 and increase the length of life dictionary $C\ (i,j)$ by 1. If the count reaches the threshold $h$ of life length, then the pixel should be integrated into the background template as the second update, followed by going to step 4). After the new pixels are traversed and processed, turn to step 2).
4) Using $\alpha$ as the base of updating the random number, a sample point randomly selected from n sample points of background template is updated with the new background.
5) A sample point randomly selected from the pixels in the M neighborhood of the selected sample pixel is replaced.

The above steps are also shown in Fig. 1.

This method is different from the other improved methods. It is the combination of life cycle of pixels and random updating and neighborhood expansion. The advantage of this method lies in removing misjudgment area and retaining the advantages of the ViBe such as fast speed and low complexity.

Fig. 2 shows the detection results of dataset PETS2009. The original graphs are Frame 45, Frame 215 and Frame 220 of

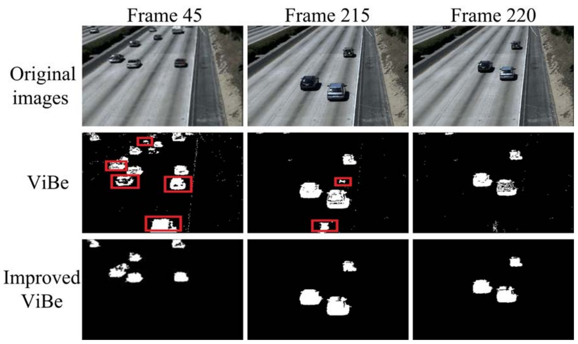


Fig. 2. Comparison of the ghost shadows based on dataset PETS2009 in Frame 45, Frame 215 and Frame 220.

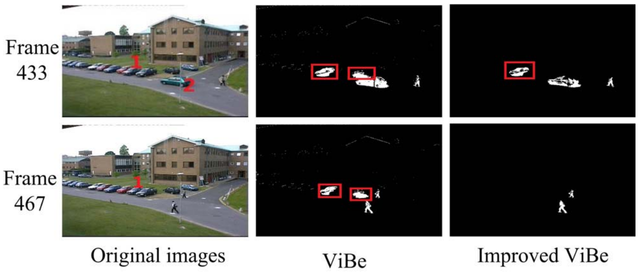


Fig. 3. Comparison of the legacy ghost shadows based on PETS2009 dataset in Frame 433 and Frame 467.

the freeway scene. The ghost shadows are labeled with red rectangles. In the detecting results of the ViBe, Frame 45 has obvious ghost shadows. Frame 215 remains 2 ghost shadows. It cannot eliminate the ghost shadows only until Frame 220. But our improved algorithm has eliminated the ghost shadows completely in Frame 45. The outlines of the object vehicles are also more complete in our algorithm.

Fig. 3 shows the detection results of experiments on dataset PETS2009. The original graphs are Frame 433, Frame 467 of the campus traffic scene video. The ghost shadows are labeled with red rectangles. In Frame 433, the vehicle 1 enters the parking space and tends to be stationary, while vehicle 2 is ready to leave from static state to driving state. There are obvious ghost shadows in the positions of vehicle 1 and vehicle 2 using the ViBe, whereas the improved algorithm has completely eliminated the ghost shadows and made the outlines of the object vehicles more complete. Frame 467 remains two ghost shadows using the ViBe, whereas the improved algorithm has completely eliminated the ghost shadows.

The experimental results show that this method can suppress ghost effectively.

### C. A Dynamic Threshold Method

Fixed threshold is used in the ViBe to judge whether the pixel should be added in background template. It is unstable in scenes with changes. This is also one of the deficiencies of the ViBe algorithm. It is necessary to consider the different dynamic situations during the process of classifying foreground and background for traffic video.

In our improved method, the OTSU is introduced to improve the original algorithm. The OTSU is a classical image segmentation method. The principle is the idea of clustering [13]. It makes the difference of gray values between the two parts maximum so that image pixels can be easily classified into two parts according to the gray level. We calculate variance to find a suitable gray level. The use of the OTSU method for image segmentation minimizes the probability of misclassification.

The background sample value of each pixel is constructed by using the ViBe, while the update of the sample is determined by the probability which retains valid sample values. According to these good characteristics, at time t, the first sample value is selected as the background frame. Then difference between the background frame and current frame $I\ (x,\ y)$ is calculated. Differential image $D\ (x,\ y)$ only contains the foreground and background, which is suitable for the OTSU.

Gray level of $D\ (x,\ y)$ is set to $L$. The pixels can be classified into two categories named $\{0,\ 1,\ T\}$ and $\{T,\ T+1,\ \ldots,\ 255\}$, respectively. The variance between the two categories is calculated according to formula (5)

$$\sigma^2 = P_0\,[u_0 - u]^2 + P_1\,[u_1 - u]^2 \tag{5}$$

Here, $P_0$ and $P_1$ represent the probability of the occurrence of two types of pixels, respectively. Variable $u$ represents the average gray value, while $u_0$ and $u_1$ represent the average gray value of the two types of pixels, respectively. The greater the value of formula (5) is, the better the threshold of segmentation is. Through using Formula (6), the optimal value of the inter-class variance $t*$ is the optimal dynamic threshold after the whole image is traversed. Because the first frame of the selecting sample is used as the background frame, in order to avoid too large or too small threshold, the bias threshold of $t*$ is limited. The upper and lower thresholds are represented as $T_{min}$ and $T_{max}$.

$$t* = Arg \max_{0 \le i \le 255} \sigma^2 \tag{6}$$

After using the improved method to distinguish the moving object, the post-processing is improved in the following two aspects.

1) Expansion and corrosion are employed to eliminate thin objects, smooth boundaries and fill small holes of the foreground.
2) The moving object contour in the binary image is calculated. The smaller density area of foreground image is released at first [40]. The center of the region meeting the minimum area condition is searched. The holes of this kind of regions are filled through image growth method in the regions. These make the moving object more obvious and complete.

## IV. A Robust Method to Detect Multiple Vehicles With Combination of Two Classifiers

In order to solve the problem that occlusion and interference causes wrong multiple vehicles tracking, the two classifiers method is designed. Using two classifiers enables the method to have both time efficiency advantage of the first classifier SVM and high accuracy advantage of the second classifier CNN.

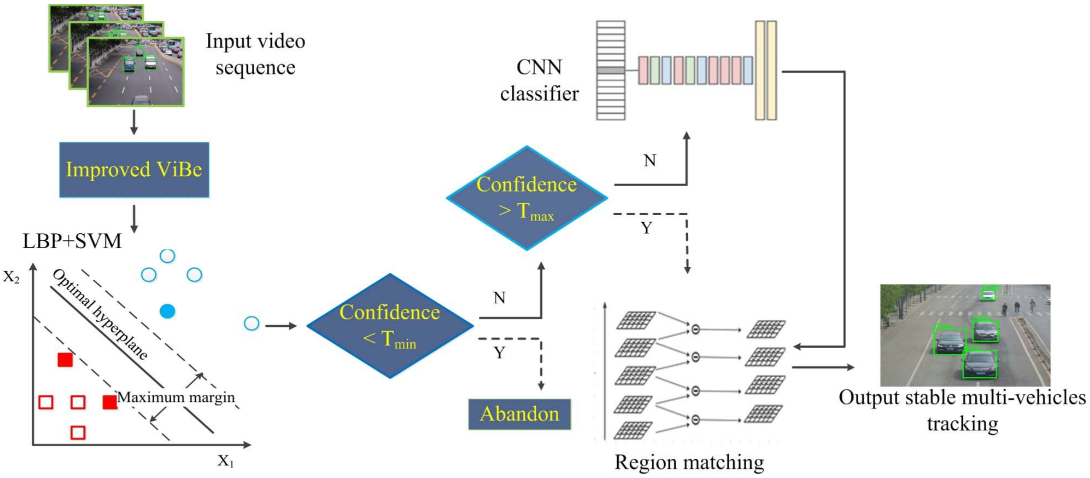
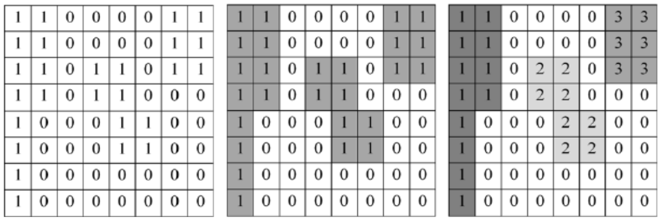


Fig. 4. A vehicles tracking method with combination of robust detection and two classifiers.



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 3 | 3 |
| 1 | 1 | 0 | 0 | 0 | 0 | 3 | 3 |
| 1 | 1 | 0 | 2 | 2 | 0 | 3 | 3 |
| 1 | 1 | 0 | 2 | 2 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 2 | 2 | 0 | 0 |
| 1 | 0 | 0 | 0 | 2 | 2 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Fig. 5. Connected regions.

The improved ViBe is used to extract the connected areas to detect moving vehicles firstly. Then the SVM classifier combined with LBP features [15] is used to scan on the image. If the confidence level is less than the low threshold $T_{min}$, it is determined not to be the object vehicle. The area rectangle is discarded directly. If the confidence value is greater than $T_{min}$ and less than the high threshold $T_{max}$, then the connected region may be the candidate vehicle tracking frame. The CNN classifier with CNN features [17] is used as the second classification to remove the interference region. If the confidence value is greater than $T_{max}$, then it can be judged as the correct tracking area. It is added into connected region table directly. After the foreground detection and classification of vehicle tracking rectangle, the connected region matching algorithm is used to do the correlation analysis for the motion of vehicle's front and rear frame to achieve stable multiple vehicles tracking.

The major steps of the above two classifiers algorithm are listed in Algorithm 1 as follows. Our proposed method is also shown in Fig. 4.

### A. Extraction of Connected Region

The first step of multiple vehicles tracking is the segmentation of the binary foreground and the extraction of the minimum enclosing rectangle (noted as MER in this paper) so as to lay the foundation for the next tracking recognition and get the correct and complete connected regions.

After the improved ViBe is used to detect the moving vehicles in traffic video, the connected regions are obtained.

**Algorithm 1** Two Classifiers Algorithm

```
Input:
(1) Low and high confidence value $T_{min}$, $T_{max}$;
(2) Video $x$ that was pre-positioned by the improved ViBe.
Steps:
for (int $i$ =0; $i$ < $x$; $i$++)
    LBP feature is extracted with rotation invariance LBP = min{ROR(LBP)};
    LBP feature is classified by SVM, where result is $\bar{x}$;
    if (confidence of $\bar{x}$ < $T_{min}$) then
        Abandon
    elseif ($T_{min}$ <= confidence of $\bar{x}$ <= $T_{max}$) then
        $\bar{x}$ is classified by CNN, where the result is $y$
        if( $y$ == true) then
            Match$O_i$, $i$ =1, ...$N_1$with $N_i$, $i$ =1, 2, ..., $N_2$
        else
            Abandon
    else // here, confidence of $\bar{x}$ > $T_{max}$
        Match $O_i$, $i$ =1, ...$N_1$ with$N_i$, $i$ =1, 2, ..., $N_2$
```

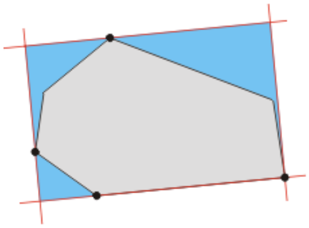


Fig. 6. An example of minimum enclosing rectangle (MER) .

The connected region is a region with the same category of pixels connected in the image. There are two kinds of neighborhood relations. They are the 4 neighborhood and the 8 neighborhood. In this paper, the 8 neighborhood is used, as shown in Fig. 5. First of all, the double scan method is used to scan the foreground image according to the relation between the adjacent pixels. All connected regions can be labeled by scanning the image sequence twice.

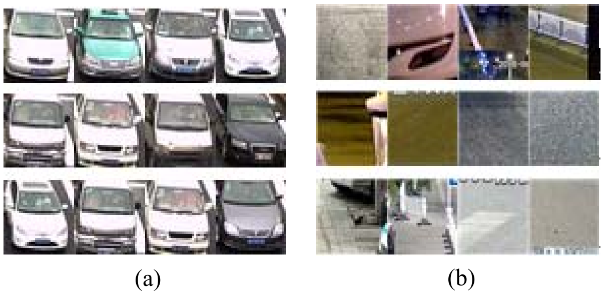


Fig. 7. Examples of some positive and negative samples.

1) In the first pass, a location label is assigned to each pixel. One or more different label may be assigned to a set of pixels in the same connected region during the scan. Therefore, we need to normalize these labels which belong to the same connected region with different labels.
2) The second scan is to classify the pixels with the same relationship into a connected region and assign the same label.
3) After all the connected regions are marked, small connected regions are removed according to the regional connectivity threshold at first. Then the MER for remaining connected region. Because the camera angle is generally fixed, only the MER of the object exceeding a certain area threshold is accepted. The MER of a connected region is shown in Fig. 6.

### B. First SVM Classification Based on LBP Feature

After extracting the effective moving masses, it is necessary to find out whether this is an effective vehicle tracking area, so as to avoid the interference of non-motor vehicles or pedestrians. The SVM classifier combined with LBP operator can well identify the tracking regions. Because of the superiority of the SVM classifier in the nonlinear high dimensional space [41], the LBP operator is combined with the SVM classifier. Using LBP has the advantages of rotation invariance and gray invariance [42]. The detailed vehicle detecting process steps of the SVM classification based on LBP feature are as follows.

1) Vehicle samples and negative samples of unrelated scenes are collected, as shown in Fig. 7 as a simple example. Then the LBP features of objects are extracted.
2) These features are applied to the SVM classifier to form a new SVM classifier.
3) Connected area rectangles are used as input for scanning one by one. The confidence value of each tracking rectangle is calculated.
4) The confidence threshold is set. When the confidence level is less than $T_{min}$, then the area rectangle is discarded. If the threshold is greater than $T_{max}$, then the rectangle is added into the regional connectivity table.
5) The connected regions with the confidence level between $T_{min}$and $T_{max}$ may be the candidate vehicle tracking rectangle. They are chosen to be classified twice.

### C. Second CNN Classification Based on CNN Feature

The confidence level of the tracking rectangle has been effectively obtained by the first cascade classification. According to the confidence level, the second classification is needed for further judging those possible false detection in the first classification. The CNN is one of the most advanced algorithms among the classifier based methods. The greatest advantage of this approach is its high accuracy [43]. But the computational complexity of CNN is higher than that of ordinary machine learning algorithms [44], and the algorithm is time-consuming. It not only achieves the integration of multiple local features, but also improves the accuracy of vehicle tracking. The process of the second classification based on CNN method is as follows.

1) The CNN feature of the object vehicle is extracted from a large number of vehicle samples and negative samples.
2) These features are applied to the CNN classifier to form a new CNN classifier.
3) Connecting regions with confidence level between $T_{min}$ and $T_{max}$ are classified by the CNN.

The most reliable tracking rectangle of a moving vehicle is obtained through using the two classifiers. The vehicle occlusion causes loss of tracking rectangle and the phenomenon of vehicle label drift in the tracking process. Therefore, we further use the method of region matching [45] to complete tracking.

### D. Multiple Vehicles Tracking Based on Region Matching

After using the SVM classifier and the CNN classifier, the most confident tracking rectangle of the moving vehicle is obtained. In order to avoid the loss of the tracking rectangle and the phenomenon of the vehicle label drift caused by the occlusion so as to make the label correctly, this paper uses the region matching method to analyze the connected region of tracking rectangles. The detailed process steps are as follows.

1) Creating history vehicle table *Old-list* at first according to multiple frames to store history vehicles, i.e. the vehicles already detected. There are coordinate information of the existing vehicle tracking rectangle and lost frames $L_{Oi}$ of each tracking rectangle in *Old-list*, where $O_i$, $i = 1, \ldots N_1$, represents the history vehicles.
2) Creating a new vehicle table *New-list* for the current frame to store the new coordinates of detecting vehicle tracking rectangle. Here, $N_i$, $i = 1, 2, \ldots, N_2$, represents a new observation track vehicle.
3) Creating an observation vehicle table *Current-list*. *Current-list* contains the coordinates of the vehicle tracking rectangle to be observed, noted as $C_i$, $i = 1, \ldots N_3$. Multiple vehicles tracking is the matching process between $N_i$ and $O_i$.
4) Calculating the matching degree of $N_i$ and $O_i$. The overlap ratio of the MER is calculated using the vehicle coordinates, as defined as $f$ as follows.

$$f = \frac{S_{o_i \cap N_i}}{S_{o_i} + S_{N_i}} \tag{7}$$

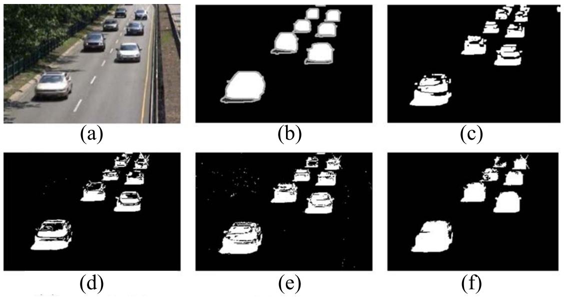


Fig. 8. Comparison of the legacy ghost shadows based on Highway video sequences in Frame 788.

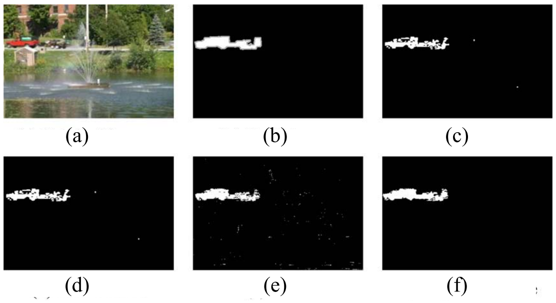


Fig. 9. Comparison of different algorithms based on Fountain video sequences.

Variable $S_{Oi \cap Ni}$ represents the overlapping area between $N_i$ and $O_i$.Variable $S_{Oi}$ and $S_{Ni}$ represent the MER of $S_{Oi}$ and $N_{i,}$ respectively. Variable $f$ represents the similarity of matching.

5) Traversing *New-list and Old-list* and updating the *Current-list* and *Old-list*.

## V. Experiments

### A. Qualitative Analysis

Qualitative analysis was done first from the vehicle detection and the vehicle tracking aspects. Our qualitative experiments were carried out using four sets of traffic videos according to previous experience. After that, statistical indicators and methods of classification algorithms were introduced. Different background subtraction based tracking methods such as the ViBe, the GMM, the LBP-OTSU and the improved ViBe were used in the comparison experiments. Experimental traffic videos of qualitative analysis are chosen from dataset PETS2009 and dataset changedetection.net 2012.

In this paper, the parameters of the algorithm are set as follow. Variable n represents the number of background samples. Variable m represents the number of neighborhood. Variable h represents the threshold of life length dictionary. The corresponding values of the parameters are as follows. Variable n is set as 20. Variable m is set as 8. Variable h is set as 15. The updating probability of the template is set as 16. Variable $T_{min}$ is set as 20 and $T_{max}$ as 35. The operating environment is PC (Intel (R) i5-4210M (TM) CPU @ 2.60 GHZ, with 8.00GB install memory (RAM), Windows 8.1 (64bits) operating system. The algorithms were developed in VS2010, using OpenCV 2.4.6.

*1) Qualitative Analysis of Multiple Vehicles Detection:* Experiments are based on the comparisons of the GMM and the LBP-OTSU to improve the contrast. Fig. 8 (a) to Fig. 8 (f) indicate the original image, true value map and the detection results of the GMM, the LBP-OTSU, the ViBe and the improved ViBe, respectively, in Frame 788 of Highway video sequence. The scene of Fig. 8 is not uniform because the tree leaves of background produce irregular flicker and the vehicle moves faster. The deviation of different detection algorithms is obvious. The increase of the objects leads to the increase of the weight of the Gaussian template and the outline of the vehicles appearing different degrees of fracture. The using of LBP-OTSU improved the adaptability of the pixel threshold. The LBP operator has certain inhibition on the light, but the noise suppression effect is poor so that there are more holes in the detecting results. The detection effect of the ViBe is reduced due to the rapid movement of the vehicles, but its anti-interference ability is not strong. The proposed algorithm effectively eliminates the ghost shadows. The results of vehicle detection of our method are more complete with much less false detected pixels for vehicles.

The second scenario of comprehensive comparison is shown in Fig. 9. The Fountain sequence of the dataset changedetection.net 2012 is selected as a challenging dynamic background traffic video in this paper. Fig. 9 (a) is Frame 737 of a sequence. Fig. 9 (b) is the truth bitmap. Fig. 9 (c), Fig. 9 (d), Fig. 9 (e), Fig. 9 (f) are the GMM detection bitmap, the LBP-OTSU detection bitmap, the improved ViBe detection bitmap and the original detection bitmap, respectively.

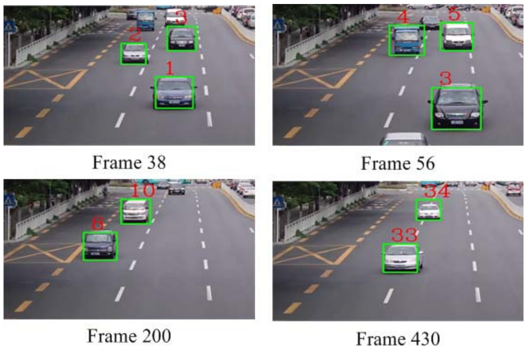
There are noises caused by the fountain and swaying branches next to the road. When a vehicle passes, the detected vehicle contours using the GMM and the LBP-OTSU are incomplete while the noise interference is obvious. In this paper, the adaptive threshold is used in our improved algorithm. Compared with the ViBe, the detection results of our improved method are clearer. It overcame the noise and improved the accuracy of the object vehicles.

*2) Qualitative Analysis of Multiple Vehicles Tracking:* This paper uses the two traffic crossroads videos and a video gathered by ourselves. Aiming at the interference of pedestrians and non-motor vehicles, the case of vehicles moving from separation to occlusion situation or the case of vehicles moving from occlusion to separation are analyzed. The experimental results are shown in Fig. 10.

The scene of Fig. 10 is a traffic station. The multiple vehicles tracking is relatively stable in the case that the vehicle does not have a large occlusion and the interference of pedestrians and non-motor vehicles.

Fig. 11 shows the tracking results of the case that vehicles moving from separation to occlusion. The tracking results are stable using our method.

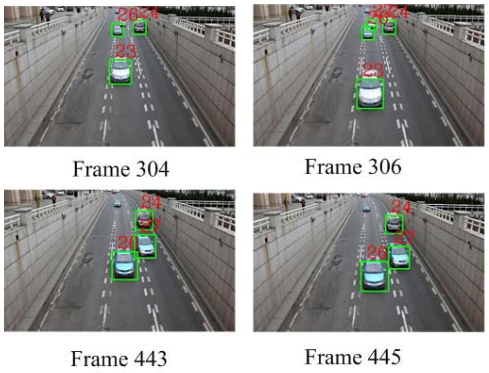
From Fig. 11, in Frame 304, the No. 28 vehicle occludes a car behind it. In Frame 306, the two cars are in serious mutual occlusion. The car previously behind is given a new number 27 while the No.28 vehicle is in stable tracking. In Frame 443, several cars have mutual occlusions. In Frame 445, vehicles separate from each other while the whole process is in the normal tracking.



Frame 38 Frame 56

Frame 200 Frame 430

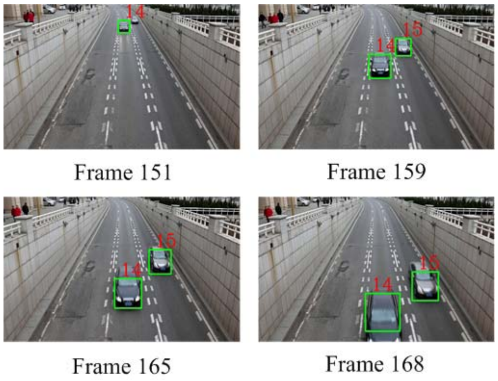
Fig. 10. Multiple vehicles tracking results in general cases.



Frame 304 Frame 306

Frame 443 Frame 445

Fig. 11. Multiple vehicles tracking results of the case that vehicles moving from separation to occlusion.



Frame 151 Frame 159

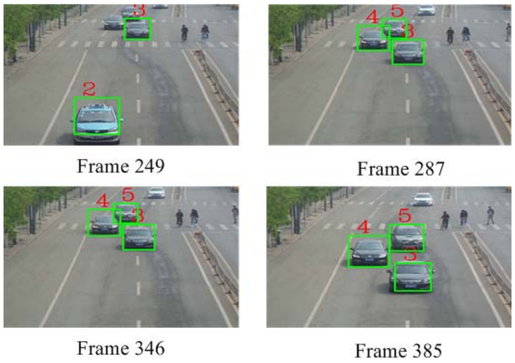
Frame 165 Frame 168

Fig. 12. Multiple vehicles tracking results of the case that vehicles are moving from occlusion to separation.

Fig. 12 shows the results of vehicle tracking from separation to occlusion. The tracking results are still stable using our proposed method.

From Fig. 12, in Frame 151, a taxi appears behind the No.14 vehicle. In Frame 159, those two cars appear serious occlusion. In Frame 165, the two cars still keep close distance. In Frame 168 multiple objects are not affected by the separation of the three vehicles.

The scene of Fig. 13 is in the traffic video gathered by ourselves. There are interference of pedestrians, bicycles and occlusion of vehicles in the video. But the tracking results are still stable using our proposed method.



Frame 249 Frame 287

Frame 346 Frame 385

Fig. 13. Multiple vehicles tracking results of the case that vehicles are moving with the interference of pedestrians and bicycles.

Fig. 13 shows that No.3 car at the middle of the crossing area with interfere pedestrians, bicycles and mobile in Frame 249. The pedestrians, bicycles and mobile that are moving at a relatively fast speed introduce interference and noises. In Frame 287, the connected region is separated. The proposed method tracked the emerging cars quickly. Bicycles and pedestrians were not mistakenly detected. In Frame 346, the vehicles gradually change from the whole occlusion to the local occlusion. Although the pedestrian and bicycle are more and more obvious, the proposed method has good robustness to avoid the interference of pedestrians and bicycles, as well as the occlusion of vehicles.

### *B. Quantitative Analysis*

In this paper, the improved ViBe is used to track vehicles, then two classifiers are used as an aid to further detect the objects from the complex environments. Tracking object vehicles using the improved ViBe has advantage of low complexity, lightweight, low memory usage. This method has better detection performance than other background subtraction methods in many literatures [11]. Using the two classifiers enables our method to have both time efficiency advantage of SVM and high accuracy advantage of CNN. Only objects that are not recognized as vehicles by SVM + LBP are put into the CNN classifier.

To further verify the robustness of the proposed method, the method was compared with the existing advanced methods using quantitative analysis. The quantitative evaluations for both detection and tracking real-time were done to verify the reliability of our method. The detection accuracy of the proposed method was compared with other advanced object detection methods using ROC curves. The tracking evaluation focuses on comparing the tracking position error of the improved ViBe with other existing advanced background subtraction based tracking methods. On the other hand, the improved ViBe not only contains tracking algorithm, but also contains foreground detection algorithm. The foreground detection evaluation was also done to prove the efficiency of the improved ViBe.

*1) Quantitative Analysis of Multiple Vehicles Detection:* Recently, some advanced methods of object detection appeared, such as DPM model, CNN feature based method,

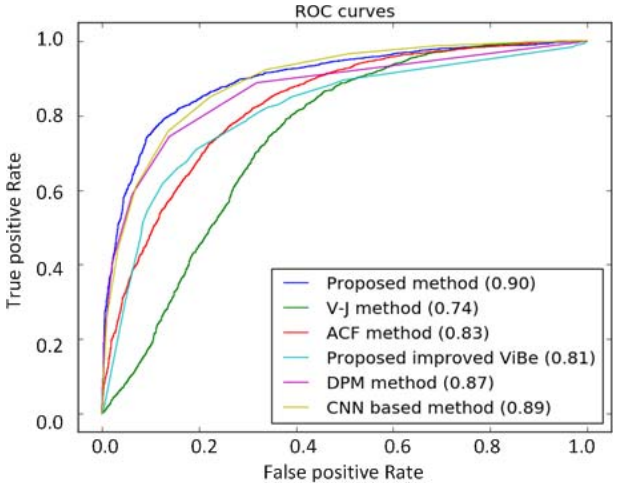


Fig. 14. Average ROC of object detection methods on standard dataset.

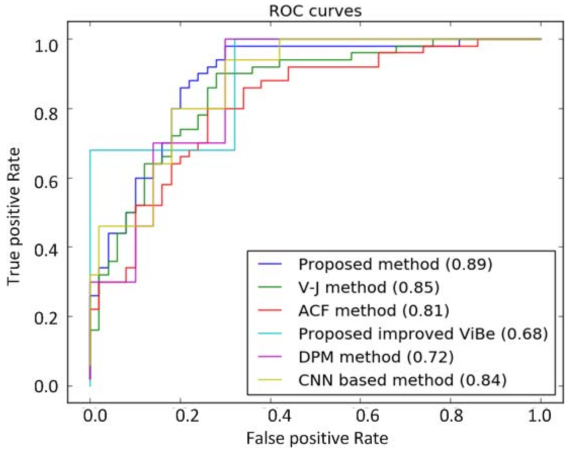


Fig. 15. Average ROC of object detection methods on self-collected dataset.

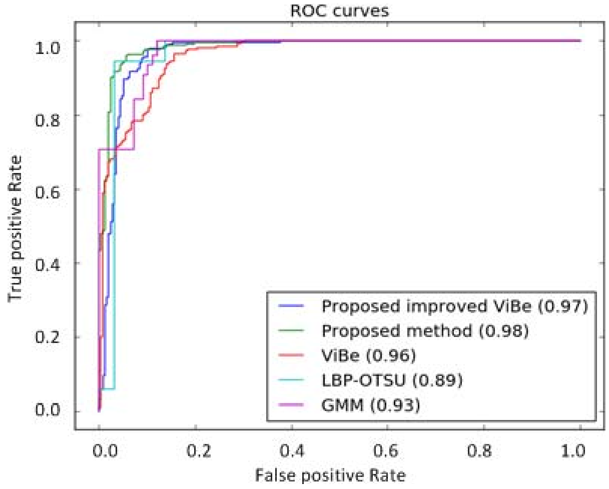


Fig. 16. Average ROC of foreground detection methods on videos.

ACF detector, and Viola-Jones method. Although these methods have some variations in recent years, their major ideas are still similar. Therefore, our method is compared with their popular variants in recent years. We have conducted our comparison experiments against the DPM model [15], the CNN feature based method [36], the ACF detector [16], and Viola-Jones (V-J) method [34].

First of all, we conducted comparative experiments using both the standard dataset and our self-collected dataset. Cross validation method was used to draw the average ROC curve of different algorithms. Ten thousand images of category "car" and ten thousand other kinds of images which were shuffled and collected in the cifar10 [46] dataset were combined as the standard dataset. Five hundred self-collected pictures of traffic vehicles and 500 non-vehicle pictures which contain sky, pedestrians, motorcycles and traffic signs elements were combined as the self-collected dataset. Part of the self-collected dataset has been shown in Fig.7. The standard dataset was used to verify the universality of the method. Because the self-collected images are more complex and close to the real road condition, they were used to verify the superiority of the proposed method. Forty percent of cross validation was used in the experiments to get the average ROC curves.

It is shown in Fig.14 and Fig.15 that the proposed method has the largest AUC area on the two testing dataset. AUC area is a common classifier evaluation index. The larger the area of AUC is, the better the performance of the method [47]. As a foreground detection method, the detection efficiency of the ViBe on static datasets is not very high. The effect of the CNN based method is close to the proposed method. The superiority of the CNN method in the field of vehicle detection is further verified. The other methods also worked well, but in the specific field of vehicle detection, the proposed method and the CNN method are better than other methods.

On the other hand, technically, the ViBe is an objects tracking method. Generally, tracking methods such as the ViBe not only contain tracking algorithm, but also contain foreground detection algorithm. In this paper, the ViBe is improved. The ViBe is a foreground detection method based on background difference. Therefore other advanced background subtraction methods were used as the comparison methods in our experiments. Average ROC curves of foreground detection methods on 1000 frames of image sequence extracted individually in Highway and Fountain videos were calculated. The results are shown in Fig.16.

Detection time of vehicle detection in traffic video is the running time of detection algorithm on the test dataset. In order to verify whether the proposed method has good real-time performance as expected, the average time consumption of different algorithms on the test dataset were recorded. In the experiments, the two datasets mentioned above, i.e. the standard dataset with 10,000 images and the self-collected dataset with 500 images, were divided into 60% of training dataset and 40% of test dataset for cross validation. The results are shown in TABLE I.

As seen in Table I, the time consumption per frame of the proposed method is far less than the time consumption of the method that only uses the CNN classifier. Although the time consumption per frame of the proposed method is greater than those of the other methods, our method is still quite fast and capable of running in real-time. The CNN method has the best

TABLE I
AVERAGE TIME CONSUMPTION PER FRAME

| Algorithm | Time consumption (ms) | |
|---|---|---|
| | Standard dataset | Self-collected dataset |
| Proposed method | 0.865799 | 0.340000 |
| V-J method | 0.179200 | 0.259995 |
| ACF method | 0.064499 | 0.065000 |
| Proposed improved ViBe | 0.346249 | 0.020995 |
| DPM method | 0.051750 | 0.115000 |
| CNN based method | 89.061749 | 3.050000 |

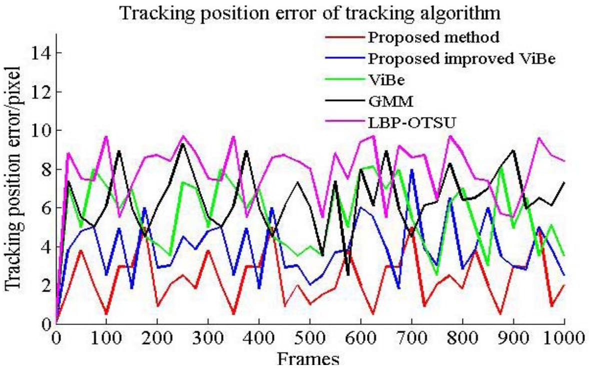


Fig. 17. Tracking error of tracking algorithms on Highway video.

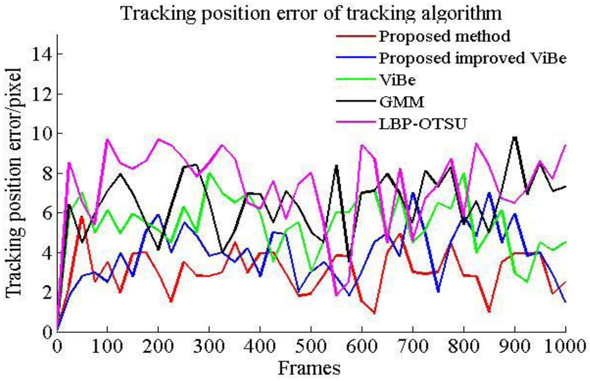


Fig. 18. Tracking error of tracking algorithms on Fountain video.

accuracy according to Fig. 14, but the experiments in Table I indicate that the CNN is time consuming and potentially has problem of running in real-time for large-scale dataset unless having better support from hardware. Therefore, compared with other methods, our method not only has high detection rate, but also consumes much less time than the CNN method, thus ensuring the efficient real-time performance.

*2) Quantitative Analysis of Multiple Vehicles Tracking:* The tracking evaluation focuses on comparative experiments between the improved ViBe method and some other existing tracking methods based on background subtraction. Tracking position errors of our method and these methods were obtained and compared. The true centroid coordinates were calibrated every 100 frames. The average of the absolute value of the offset error and the x coordinate error were added and calculated for average value. The experimental results are shown in Fig. 17 and Fig. 18.

As demonstrated in Fig. 17 and Fig. 18, the tracking position error of the proposed method has been substantially improved compared with other methods. Our proposed method, same as other algorithms, did not completely fail to detect any vehicle objects, but meanwhile having good accuracy and time efficiency according to the experiments discussed above. This further validates the robustness of our proposed method.

## VI. CONCLUSION

In this paper, a new method to track multiple vehicles with combination of robust detection and two classifiers is proposed. We propose the improved ViBe for robust and accurate detection of multiple vehicles. It effectively restrains dynamic noise and removes the ghost shadows and object's residual shadows quickly. In this paper we also designed a method using two classifiers to attack the problem of failure to track vehicles with occlusions. The two classifiers method has both time efficiency advantage of the SVM and high accuracy advantage of the CNN. Several quality evaluation criteria based on statistics index of classification algorithms are adopted. The comparative experiments were conducted to evaluate the quality and performance using these criteria between the proposed method and some popular algorithms. The qualitative and quantitative experiments showed that our improved method removed the ghost shadows, improved the detection accuracy and overall performance, and was robust to deal with the occlusion of vehicles in various traffic scenes. In the future, we will further combine with the deep learning technologies to carry out the research of vehicle detection, tracking and recognition.

## REFERENCES

[1] G. Tagne, R. Talj, and A. Charara, "Design and comparison of robust nonlinear controllers for the lateral dynamics of intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 796–809, Oct. 2015.

[2] Y. Yuan, Z. Xiong, and Q. Wang, "An incremental framework for video-based traffic sign detection, tracking, and recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1918–1929, Jul. 2017.

[3] B.-F. Wu, C.-C. Kao, J.-H. Juang, and Y.-S. Huang, "A new approach to video-based traffic surveillance using fuzzy hybrid information inference mechanism," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 485–491, Mar. 2013.

[4] X. Wang, L. Xu, H. Sun, J. M. Xin, and N. Zheng, "On-road vehicle detection and tracking using MMW radar and monovision fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2075–2084, Jun. 2016.

[5] J. Fang, Q. Wang, and Y. Yuan, "Part-based online tracking with geometry constraint and attention selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 854–864, May 2014.

[6] Q. Wang, J. Fang, and Y. Yuan, "Multi-cue based tracking," *Neurocomputing*, vol. 131, no. 7, pp. 227–236, May 2014.

[7] B. Zhang, A. Perina, Z. Li, V. Murino, J. Liu, and R. Ji, "Bounding multiple Gaussians uncertainty with application to object tracking," *Int. J. Comput. Vis.*, vol. 118, no. 3, pp. 364–379, Feb. 2016.

[8] J. Han, E. J. Pauwels, P. M. D. Zeeuw, and H. N. Peter, "Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 225–263, Jun. 2012.

[9] R. K. Satzoda and M. M. Trivedi, "Multipart vehicle detection using symmetry-derived analysis and active learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 926–937, Dec. 2016.

[10] Y. Huang and I. Essa, "Tracking multiple objects through occlusions," in *Proc. CVPR*, Jun. 2005, pp. 1051–1058.

[11] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[12] T. Kryjak, M. Komorkiewicz, and M. Gorgon, "Real-time implementation of foreground object detection from a moving camera using the ViBe algorithm," *Comput. Sci. Inf. Syst.*, vol. 11, no. 4, pp. 1617–1637, Jun. 2014.

[13] X.-C. Yuan, L.-S. Wu, and Q. Peng, "An improved Otsu method using the weighted object variance for defect detection," *Appl. Surface Sci.*, vol. 349, pp. 472–484, May 2015.

[14] Y. Luo, C. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of PCA and LBP with SVM," *Int. J. Light Electron Opt.*, vol. 124, no. 17, pp. 2767–2770, Aug. 2013.

[15] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[16] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, Aug. 2014.

[17] S. Yu, Y. Wu, L. Wei, Z. Song, and W. Zeng, "A model for fine-grained vehicle classification based on deep learning," *Neurocomputing*, vol. 257, no. 27, pp. 97–103, Feb. 2017.

[18] C. Cuevas, R. Martínez, and N. García, "Detection of stationary foreground objects: A survey," *Comput. Vis. Image Understand.*, vol. 152, pp. 41–57, Jul. 2016.

[19] N. Jiang and W. Liu, "Data-driven spatially-adaptive metric adjustment for visual tracking," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1556–1568, Apr. 2014.

[20] H. Yin, Y. Chai, S. X. Yang, and X. Yang, "Fast-moving target tracking based on mean shift and frame-difference methods," *J. Syst. Eng. Electron.*, vol. 22, no. 4, pp. 587–592, Aug. 2011.

[21] M. Fei, J. Li, and H. Liu, "Visual tracking based on improved foreground detection and perceptual hashing," *Neurocomputing*, vol. 152, pp. 413–428, Nov. 2015.

[22] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. Van Den Hengel, "A survey of appearance models in visual object tracking," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, p. 58, Sep. 2013.

[23] T. C. Huang, T. H. Wu, Y. H. Lin, W. Y. Guo, W. C. Huang, and C. J. Lin, "Quantitative flow measurement by digital subtraction angiography in cerebral carotid stenosis using optical flow method," *J. X-Ray Sci. Technol.*, vol. 21, no. 2, pp. 227–235, Feb. 2013.

[24] B. Wang, Z. Cai, L. Shen, and T. Liu, "An analysis of physics-based optical flow," *J. Comput. Appl. Math.*, vol. 276, pp. 62–80, Jul. 2015.

[25] P. Dérian, P. Héas, C. Herzet, and E. Mémin, "Wavelets and optical flow motion estimation," *Numer. Math., Theory, Methods Appl.*, vol. 6, no. 1, pp. 116–137, Jan. 2013.

[26] M. Lucena, J. M. Fuertes, and N. P. de la Blanca, "Optical flow-based observation models for particle filter tracking," *Pattern Anal. Appl.*, vol. 18, no. 1, pp. 135–143, Sep. 2015.

[27] M. Camplani and L. Salgado, "Adaptive background modeling in multicamera system for real-time object detection," *Opt. Eng.*, vol. 50, no. 12, pp. 1–17, Dec. 2011.

[28] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imag.*, vol. 11, no. 3, pp. 172–185, Jun. 2005.

[29] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, Dec. 2014.

[30] Z. Yin and R. Collins, "Belief propagation in a 3D spatio-temporal MRF for moving object detection," in *Proc. CVPR*, 2007, pp. 1–8.

[31] J. Han and P. H. N. de With, "Real-time multiple people tracking for automatic group-behavior evaluation in delivery simulation training," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 913–933, Nov. 2011.

[32] N. Kosaka and G. Ohashi, "Vision-based nighttime vehicle detection using CenSurE and SVM," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 10, pp. 2599–2608, Oct. 2015.

[33] S. Dahiya, "A Gaussian filter based SVM approach for vehicle class identification," *Int. J. Modern Edu. Comput. Sci.*, vol. 12, pp. 9–16, Dec. 2015.

[34] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on Viola–Jones and HOG + SVM from UAV images," *Sensors*, vol. 16, no. 8, pp. 1325–1332, Aug. 2016.

[35] X. Hu, X. Ye, D. Zhang, and L. Wu, "Vehicle detection technology based on cascading classifiers of multi-feature integration," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 10, p. 1750032, Apr. 2017.

[36] L. Chen, X. Hu, T. Xu, H. Kuang, and Q. Li, "Turn signal detection during nighttime by CNN detector and perceptual hashing tracking," *IEEE Trans. Intell. Transp. Syst.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7891988/

[37] H. Li, K. Fu, M. Yan, X. Sun, H. Sun, and W. Diao, "Vehicle detection in remote sensing images using denoizing-based convolutional neural networks," *Remote Sens. Lett.*, vol. 8, no. 3, pp. 262–270, Jan. 2017.

[38] Z. Wang, H. Liu, and Z. Huo, "Scale-invariant feature matching based on pairs of feature points," *IET Comput. Vis.*, vol. 9, no. 6, pp. 789–796, Feb. 2015.

[39] P. Hanhoon and M. Kwang-Seok, "Fast feature matching by coarse-to-fine comparison of rearranged SURF descriptors," *IEICE Trans. Inf. Syst.*, vol. 98, no. 1, pp. 210–213, Oct. 2015.

[40] Q. Liu, Y. Yang, Y. Gao, and R. Hong, "Texture-adaptive hole-filling algorithm in raster-order for three-dimensional video applications," *Neurocomputing*, vol. 111, pp. 154–160, Dec. 2013.

[41] F. de Morsier, D. Tuia, M. Borgeaud, V. Gass, and J.-P. Thiran, "Semi-supervised novelty detection using SVM entire solution path," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1939–1950, Apr. 2013.

[42] B. Yang and S. Chen, "A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image," *Neurocomputing*, vol. 120, no. 23, pp. 365–379, Nov. 2013.

[43] Y. Wei *et al.*, "Cross-modal retrieval with CNN visual features: A new baseline," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 449–460, Feb. 2017.

[44] A. L. Buczak and E. E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2015.

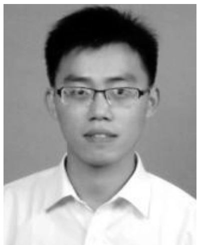
[45] M. V. Afonso, J. C. Nascimento, and J. S. Marques, "Automatic estimation of multiple motion fields from video sequences using a region matching based approach," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 1–14, Dec. 2014.

[46] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. ICAIL*, 2011, pp. 215–223.

[47] J. Qi, S. Dong, F. Huang, and H. Lu, "Saliency detection via joint modeling global shape and local consistency," *Neurocomputing*, vol. 222, no. 26, pp. 81–90, Jan. 2017.

**Weidong Min** received the B.E., M.E., and Ph.D. degrees in computer application from Tsinghua University, China, in 1989, 1991 and 1995, respectively, on the research subjects of computer graphics, image processing, and computer aided geometric design. He was an Assistant Professor with Tsinghua University from 1994 to 1995. From 1995 to 1997, he was a Post-Doctoral Researcher with the University of Alberta, Canada. From 1998 to 2014, he was a Senior Researcher and a Senior Project Manager with Corel and other companies in Canada. In recent years, he cooperated with the School of Computer Science and Software Engineering, Tianjin Polytechnic University, China. Since 2015, he has been a Professor with the School of Information Engineering, Nanchang University, China. He is a member of The Recruitment Program of Global Expert of Chinese government. He is an Executive Director of the China Society of Image and Graphics. His current research interests include computer graphics, image and video processing, distributed system, software engineering, and network management.

**Mengdan Fan** received the B.E. degree in computer application from Jiangxi Agricultural University, China, in 2015. She is currently pursuing the master's degree with Nanchang University, China, on the research subject of abnormal behavior detection in video surveillance.



**Xiaoguang Guo** received the B.E. degree in computer application from Tianjin Polytechnic University, China, in 2015, where he is currently pursuing the master's degree on the research subject of abnormal behavior detection in video surveillance.

**Qing Han** received the B.E. and M.E. degrees in computer application from Tianjin Polytechnic University, China, in 1997 and 2006, respectively. She is currently an Associate Professor with the School of Information Engineering, Nanchang University, China. Her research interests include image and video processing, network management.