IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 15, 2019, accepted June 6, 2019, date of publication June 14, 2019, date of current version July 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923199

Detection and Classification of Moving Vehicle From Video Using Multiple Spatio-Temporal Features

YU WANG[1,2], XIAOJUAN BAN[1], HUAN WANG[3], DI WU[4], HAO WANG[5], SHOUQING YANG[2], SINUO LIU[1], AND JINHUI LAI[2]

[1]Beijing Advanced Innovation Center for Materials Genome Engineering, School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]National University of Defense Technology, Changsha 410009, China
[3]School of Information Science Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China
[4]Department of ICT and Natural Science, Norwegian University of Science and Technology, 6009 Ålesund, Norway
[5]Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Corresponding authors: Xiaojuan Ban (banxj@ustb.edu.cn) and Huan Wang (wanghuan_emma@163.com)

This research was funded by The National Key Research and Development Program of China (under Grant No.2016YFB0700502), it support for image processing and object detection of this article; National Natural Science Foundation of China (under Grant Nos. 61873299), it support for machine learning and target classification of this article.

ABSTRACT The information acquisition and automatic processing technology based on visual surveillance sensors in intelligent transportation system (ITS) has become an important application field of computer vision technology. The first step of a visual traffic surveillance system usually needs to correctly detect objects from videos and classify them into different categories. In this paper, the improved spatiotemporal sample consistency algorithm (STSC) is proposed, to enhance the robustness of background subtraction in complex scenes. To address this challenge of classifying acquired from visual traffic surveillance sensors in a particular area in China, improved spatiotemporal sample consistency algorithm is proposed, which consists of two main stages. In the first stage, the robustness of moving object detection is further provided by the method we proposed based spatiotemporal sample consistency; in the second stage, we propose the target classification method based prior knowledge, in addition correcting in tracking progress. The experiments on the CDnet 2014, MIO-TCD, and BIT-Vehicle show that the method we proposed successfully overcomes the adverse effects in the complex environment with different shooting angle and resolution taken by single fixed cameras, besides effectively reduces the false alarm rate of classification.

INDEX TERMS Moving object detection, vehicle type classification, spatio-temporal, monitoring video.

I. INTRODUCTION

The development of artificial intelligence technology has improved traditional magnetic detector, radar speed measurement, gravity sensors. The traffic information acquisition equipment has also updated to GPS and video surveillance system based on computer vision. As a result, traffic congestion evaluation method based on floating vehicle technology is getting more sophisticated. The video surveillance system can collect a wealth of information to detect, locate, track and identify targets in the traffic flow. At the same time, information such as various traffic parameters can also be deduced from this system. To process traffic videos is the key question in modern intelligent transportation.

In regular traffic system, congestion severity and vehicle speed are negatively correlated. Current research in traffic operation evaluation for urban roads mainly considers road grade and vehicle speed. However, in practice, there are obvious difference for different types of moving vehicles (such as lorry and racing car). Therefore, in order to ensure the credibility of the traffic congestion evaluation results, it is necessary to accurately distinguish the target types in traffic monitoring video, so as to minimize the false alarm rate of vehicle target classification. We need an effective classification method to traffic congestion evaluation. Another critical issue for this method is to comply with local standard for automobile classification.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan.

There are two categories of methods to classify vehicles in the traffic video. The first category is the image segmentation methods based on features. These methods usually require for heavy computation but are weak to differentiate motions and stationary targets. Most of them have low recognition rate and high false alarm rate in complex environment. The second category is deep learning methods. They require high training sample quantity and high quality computing environment. However, their main disadvantage is efficiency. They are difficult to adjust to near-realtime traffic systems. The Spatio-Temporal Sample Consensus (STSC) has been developed for good effect on moving target detection. In this paper, we propose an Improved Spatio-Temporal Sample Consensus (ISTSC) method. It contains two steps to implement detection and classification of moving vehicle from traffic videos collected by fixed cameras. First is to detect and extract the moving targets in the video by background subtraction. Second is to identify and classify the foreground moving targets based prior knowledge.

Moving target detection is the premise of high-level processing tasks such as model detection, license plate recognition, vehicle tracking and behavior analysis. Motion target detection methods can be divided into three types: optical flow method [1], inter-frame difference method [2] and background subtraction method [3]. The background subtraction method has complete detection results, simple principle and good real-time performance. It is one of the most commonly applied target detection method at present. The key of this method is to construct a robust background model which can adapt to the situation challenges in the video sequences with complex application environments [4]. Many of these difficulties are related to fixed cameras, such as moving target in the initial frame, illumination variation, dynamic background, moving object appearance changes, camera jitter, complex background, noise and shadow. Considering the challenges above, this paper propose a Spatio-Temporal Sample Consensus (STSC) method, which focuses on robust background modeling, foreground threshold segmentation strategy and model updating, illumination variation solution, target post-processing and shadow remove, in order to improve the system robustness. This paper further build a model to classify the moving target types by the multi-feature fusion, such as vehicle symmetry, plate number, vehicle sharp features and other prior knowledge, as well as building cascading classifier, besides correcting target types in tracking process.

II. RELATED WORK

For traffic monitoring video analysis technology, the literature [5], [6] introduced the automatic detection method of traffic events based on trajectory detection. They realized traffic events detection through background extraction, motion trajectory extraction and so on. Cucchiara proposed a vehicle detection algorithm based on rule reasoning [7]. Maurin proposed a multi-layer model algorithm, which used the Kalman Filter Tracking to realize traffic congestion detection [8]. The Advanced Research Development Activity (ARDA) Institute of the United States began a Video Analysis and Content Extraction (VACE) project in the autumn of 2003 to detect, identify and track objects by image processing in intelligent traffic monitoring.

Since last century, the background subtraction method has attracted many attentions because of its advantages in simple principle and complete detection results. Meanwhile several representative classical algorithms have been put forward. For example, Wren [9] proposed the Tangos background modeling algorithm based on the statistical characteristics of sequence image pixels in time. Stauffer and Grimson [10] proposed the hybrid Gaussian modeling algorithm, compared to the Tangos model, which could describe the multi-peak state of the background, and was suitable for modeling under the complex background such as tree swing and illumination gradient, but it had some problems, such as complex models and slow updating speed; Zivkovic [11] proposed an improved hybrid Gaussian background modeling method by introducing prior knowledge, which the number of modes was adaptive, and the learning rate was adjusted by recursive method online; in order to avoid the influence of parameter estimation on the background in modeling, Elgammal [12]–[15] proposed the background subtraction algorithm of no parameter model based on kernel function density estimation, which did not make any hypothesis, used the previous historical information to estimate the future state, eliminated the influence of parameter selection on the background model, and improved the stability of the model, but the disadvantage was that the calculation amount of the algorithm is large; Kim [16], [17] proposed the Codebook model by introducing the clustering idea into the background modeling method, through the calculation of color distortion and brightness fluctuation range to cluster codeword, the distinction between the target and the background was achieved, meanwhile the algorithm had a good real-time performance. In addition, some common algorithms in other fields, such as Markov model [18], random condition field [19], image texture features [20], [21], artificial neural network [19], [22] and so on, have also been well applied in the field of background modeling.

In addition to the above target detection methods, the common detection methods include the image segmentation method, the point detection method and the target recognition method. The detection method based on image segmentation has active contour [23]–[25], edge operator [26], graph cut [27], mean shift [28], region growth [29] and so on. The advantage of the point detection is that the feature points do not change with the illumination intensity and the camera perspective. Typical these methods include Moravec point detection operator [30], Harris point detection [31], SIFT [32] and SURF [33] invariant feature point detection. The target recognition method is a detection method that trains classifiers to automatically obtain the features of specific targets, including face detection [34], pedestrian detection [35], vehicle detection [36] and so on. This type of method needs to

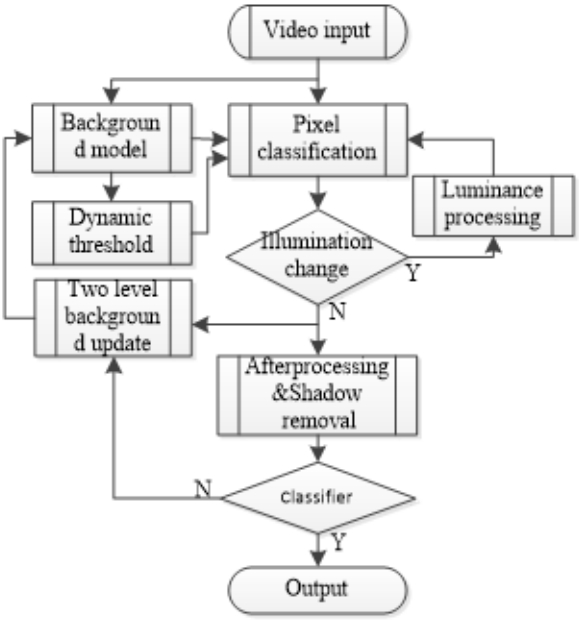

FIGURE 1. Improved STSC algorithm flow chart.

calibrate many training samples manually in advance and carry out progressive scanning of the image in the form of sliding window. What's more, it requires a lot of prior knowledge and heavy computation.

Most of methods of vehicle type classification fall into two categories: model-based methods and appearance-based methods. Petrovic et al. [9] extracted a great many of features, such as Sobel edge response, edge orientation, direct normalized gradients, locally normalized gradients, and Harris corner response from vehicle frontal view images to classify vehicle types. Negri et al. [37] presented a voting algorithm for a multiclass vehicle type recognition system based on oriented contour points. Psyllos et al. [38] used SIFT features to recognize the logo, manufacture and model of a vehicle. Peng et al. [39] represented a vehicle by license plate color, vehicle front width, and type probabilities for vehicle type classification. Zhen et al. [40] propose an appearance-based vehicle type classification method from vehicle frontal view images. Besides other appearance-based methods using vehicle side view images [41]–[44].

III. IMPROVED SPATIO-TEMPORAL SAMPLE CONSENSUS

We develop a Spatio-Temporal Sample Consensus algorithm, considering the actual application situation of highway network and the need of detecting the moving vehicles. We, classifying the moving object by the multi-feature fusion, and further improves the robustness of illumination variation, and eliminates the interference of the shadow. The proposed motion vehicle detection and type classification algorithm of improved spatio-temporal sample consensus show in Fig.1.

A. BACKGROUND SUBTRACTION

1) BACKGROUND MODELING AND INITIALIZATION

Our algorithm is inspired by the CS-STLTP algorithm [45] and ViBe algorithm [46]. The Background modeling process based on Spatio-temporal information fusion is shown in Fig.2. The $v(x)$ is used to represent the pixel value of image midpoint x. The background modeling is as follows:

Step 1: Establish a $1 \times n$ dimension temporal dimension sample set for pixel x;

$$T(x) = v^0(x), v^1(x), \cdots, v^{n-2}(x), v^{n-1}(x) \tag{1}$$

Step 2: Solve $V^{n-1}(x)$, which is the moving average of $T(x)$, and getting the moving average background model of the first n frames:

$$V^{i+1}(x) = (1-a)V^i(x) + av'(x) \tag{2}$$

where a is the updating rate.

Step 3: Establish a $1 \times N$ dimension background sample $B(x)$;

Step 4: Initialize the background model randomly extracte $k(k < n)$ sample pixels from $T(x)$, and add them to $B(x)$. Then randomly extracte the remaining $N - k$ sample pixels from the 8-neighborhood space of $V^{n-1}(x)$ to the background model.

The time domain information effectively solves the phenomenon of "ghost shadow"; the calculating of moving average $V^{n-1}(x)$ can adapt the noise effectively, which also can mitigate the adverse effects of dynamic background; the moving average spatial information can suppress the interference of camera jitter effectively.

2) DYNAMIC THRESHOLD CLASSIFY

In this paper, we use dynamic threshold method to classify the truth foreground pixels. Shannon's information entropy theory [47] is a concept aiming at measure information, which can be used to describe the complexity of an area of an image, and the difference between pixels and entropy value is positively correlated. The entropy value of dynamic background area or object edge is high, and the entropy value of coherent region such as sky and pavement is low. We classify the dynamic region and the static region of image according to the concept of background sample entropy as follows.

$$H_x = -\sum_{u=1}^{l} p_u \log p_u$$
$$p_u = \left(\sum_{i=1}^{M} \delta\left[m\left(b_i\right) - u\right] \right) / M \tag{3}$$

$B(x) = \{b_0, b_1, \cdots, b_{N-1}\}$ is the background sample set of the pixel x, $N_G(x)$represents the set of pixels in the 3×3 neighborhood of x, $p_u = \{p_u\}_{u=1,2,\ldots k}$ refers to the grayscale probability density of all samples of $N_G(x)$, l represents the quantization level; the pixel b_i corresponds to a characteristic level of $m(b_i)$; Dirac function $\delta\left[m(b_i) - u\right]$ is

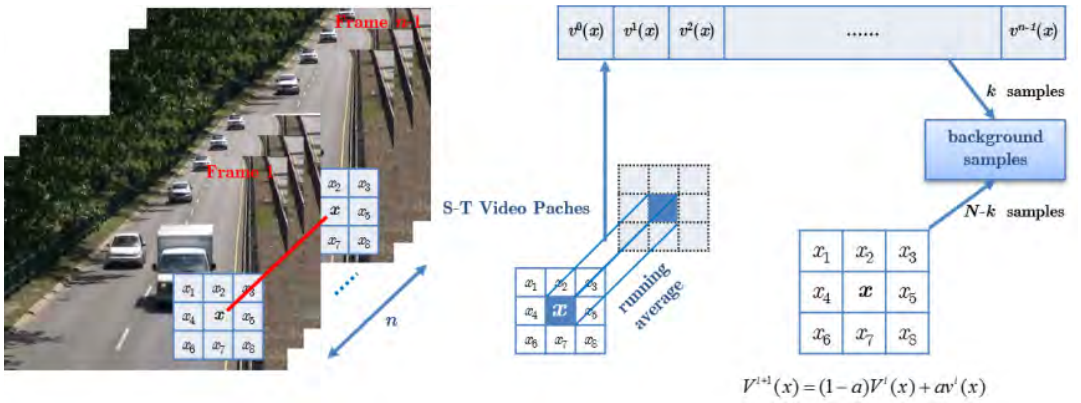

FIGURE 2. Background model building sketch.

used to determine whether the color value of pixel b_i in the target image belongs to the u of b_{in} (equal to 1, otherwise 0); $M = 3 \times 3 \times N$ is the sum of all sample pixels in spatial neighborhood set $N_G(x)$ of pixel x. represents the entropy value of background pixel x, by using the maximum inter-class variance method [48], the image can be classified into complex regions and simple regions.

After classifying the area by background sample entropy, set a label $Label(x)$ each pixel x which marks whether the neighborhood of x is a complex background. Define a threshold $Th^t(x)$ of x at time as:

$$Th^t(x) = \begin{cases} Th + \eta\sigma, & Label(x) = 1 \\ Th - \eta\sigma, & Label(x) = 0 \end{cases} \tag{4}$$

where Th is global constant, σ is the standard deviation for all background sample points in the x neighborhood. To maintain an appropriate local thresholds, we set the threshold updating period to L, while set the maximum and minimum thresholds $Th_{\max}$ and $Th_{\min}$.

$B(x) = \{b_0, b_1, \ldots\ldots b_{N\text{-}1}\}$ represents background set of pixel x, $V^t(x)$ represents the pixel value of x at time t, taking the absolute value of the difference of $V^t(x)$ and $B(x)$, the cumulative difference is less than the number of $Th^t(x)$, if the absolute is greater than the threshold Γ, then defining it as the background, otherwise foreground pixel, $M^t(x)$ represents the pixel classification after the mask value, defined as:

$$\Phi_i(x,t) = \begin{cases} 1, & if \ \left|V'(x) - b_i(x)\right| \le Th^t(x) \\ 0, & \text{otherwise} \end{cases}$$
$$M^t(x) = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} \Phi i(x,t) < \Gamma \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

We can get a binarization detected foreground by:

$$x = \begin{cases} \text{foreground}, & \text{if } M^t(x) = 1 \\ \text{background}, & \text{if } M^t(x) = 0 \end{cases} \tag{6}$$

3) BACKGROUND UPDATING POLICY

In the actual traffic application scenario, background always changes, so the background model needs to be updated over time. This paper combines the stochastic updating method in Vibe algorithm and the overall updating method in SACON algorithm, then proposes a two-level updating policy, including background level updating policy and foreground level updating policy.

1)Background Updating Policy: When pixel $V^t(x)$ is judged as a background pixel, the older pixel is replaced to update the temporal sample set $T(x)$ by timing. In order to suppress the interference of caused by the short-time parking. We improvement on the basis of ViBe algorithm, which randomly replaces a pixel b_i in $B(x)$ with the average value of n pixel contained in $T(x)$. At the same time, in order to maintain the consistency of pixel neighborhood, the same method is used to update the background sample in $G_N(x)$.

2)Foreground Updating Policy: In order to adapt to slight movement or local action of stationary vehicles, the detected foreground is divided into *Blob*the simply connected domain. Comparing the nearest *Blob* in two frames, if its centroid or number of pixel does not change, the *Blob* is considered as a static blob, when designing the number *Map* for each pixel, if a *Blob* is stationary, the *Map* value of all pixels in the *Blob* adda 1. If a *Blob* is in moving, the *Map* value is set to Zero. When the *Map*, which is the value of the object, is higher than the threshold β, all pixels contained in the object are updated to the background pixels.

$$Map_t(y|y \in Blob) = \begin{cases} \text{Map}_t(y|y \in Blob) + 1, & \text{if Blob is static} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

4) ILLUMINATION VARIATION

The illumination variation can be divided into two categories, the one whose global illumination mutation follows the Bayesian division, and the one which is the caused by the illumination facility switch. First, we divide the image into blocks with size of $W \times H$. The global illumination variation

judgment as:

$$\begin{cases} \phi = 1, & pix_f / pix_b \geq r_1 \\ \phi = 0, & \text{otherwise} \end{cases} \tag{8}$$

where $\phi = 1$ indicated the global illumination variation happened, $\phi = 0$ indicated not happened, pix_f represents number of pixels in foreground, pix_f represents number of pixels in background.

The illumination variation in block ($1 \leq i \leq N$) judgment as:

$$\begin{cases} \phi_i = 1, & Sm_i / St_i \geq r_3, \quad pix_i^f / pix_i^b \geq r_2 \\ \phi_i = 0, & \text{otherwise} \end{cases} \tag{9}$$

where Sm_i represents number of matching feature SIFT between previous frame in block i, Sm_i represents total number of feature SIFT,r_1, r_2, r_3 are thresholds.

5) SHADOW REMOVAL

It is necessary for us to do same job in shadow removing for improving the accuracy of classification. In this paper, the method combining color features and contour features is used to detect the target shadow region. Firstly, open and close operation is carried out successively for the results of moving target extraction. The connected domain is marked to eliminate the isolated region with small area and obtain several complete foreground target area. The contour of the moving target is extracted, and the direction of the shadow is judged according to the contour trend; secondly, the shadow seed point is obtained according to the contour direction, as showing in Fig. 3, in which the black dot is used to identify the significant ''inflection point'' detected; then the possible shadow seed points are screened and the ''false shadow points'' are excluded according to the spatial characteristics of the shadow RGB; finally, the exact shadow area is determined by clustering method as the center of various subpoints, and the shadow removal of the object is completed.

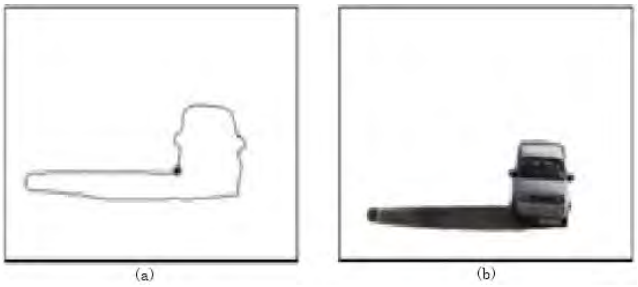

FIGURE 3. Object contour extraction results, (a) target detection result, (b) contour extraction result.

B. CLASSIFICATION BASED MULTI-FEATURE FUSION

According to the evaluation index system of traffic congestion in ITS, the collected data only belongs to two kinds of moving targets: large vehicles and cars. According to the definition of medium-sized vehicles in National Standards of the People's Republic of China, the vehicles that the model yard is more than 6000mm (including) are defined as large vehicles. A large variety of moving targets are detected by traffic video surveillance, especially under the urban traffic conditions.

In the past studies, there are several methods to classified targets into different categories. Chen [17] classified the objects into four main vehicle categories, such as car, van, bus and motorcycle, Mithun et al. [49] classified them into seven types, such as motorbike, rickshaw, autorickshaw, car, jeep, covered-van and bus. In order to evaluate parameters more accurately in China traffic scenes, we divide them into the following three types:

Type I (large vehicles): trucks, coaches, buses and trailers.

Type II (cars): Sedan, SUV, MPV and pickup below medium size.

Type III (others): two-wheeled mopeds, three-wheeled motorcycles, electric bikes and bicycles, pedestrians, livestock.

There are obvious differences between the three types of targets such as plate number, shape characteristic, symmetry, face features and area features. In order to improve the classification accuracy, we intend to integrate five features to form a strong classifier through a number of weak classifier cascading. In addition correcting target types in tracking process, the Figure 4 shows the flow of method we proposed to classify the target into three types.

1) LICENSE PLATE AND SHAPE CHARACTERISTICS

If camera resolution is high enough, without occlusion, we can get the plate number for accurate vehicle information by license plate recognition.

According to the National Standards of the People's Republic of China ''GB1589-2004 Limits of Dimensions, Axle Load And Masses for Road Vehicles'', the width of license plate of China's road vehicles is unified to 440mm, and the width of general passenger cars is 1400mm to 2000mm; the width of license plate of motorcycles is 220mm, the width of general two-wheeled motorcycles is 300mm to 1000mm, the maximum width of passenger vehicles and cargoes is 2500mm.

On the basis of making full use of the prior knowledge of license plate, which is mentioned in literature [50], a license plate location method based on HSV color space and mathematical morphology was put forward to obtain the position of vehicle license plate in the foreground target. The accurate type information of the vehicle can be obtained by retrieving the license plate information easily. However, in view of the failure of license plate character recognition caused by interference factors such as local alteration of license plate, local occlusion of license plate or Illumination variation, it is better to classify the target vehicle according to the ratio of license plate and target width, as well as the ratio of length and width of the target.

The vehicle is classified by calculating the ratio of the width w_{lp} of license plate image to the width w_{veh} of vehicle

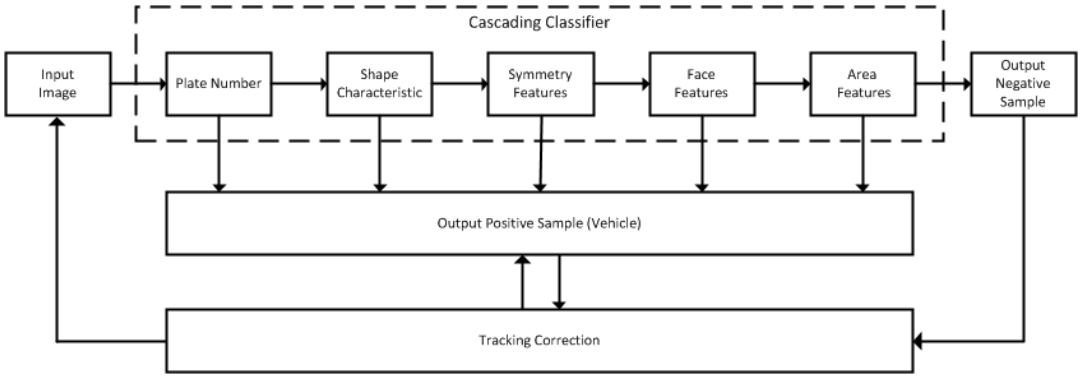

FIGURE 4. Flow chart of method to classify targets.

target image. It can be judged by calculating the ratio of length l_{veh} to width w_{veh} of vehicle in the image. Set the moving target license plate parameter as LP. When the license plate area is detected in the target pixel area, $LP = 1$, otherwise $LP = 0$; the target type parameter is T_{veh}, that is, when the moving target x is a large vehicle or car, then $T_{veh} = 1$; when the moving target x is a riding target or pedestrian target, then $T_{veh} = -1$, when the target type is uncertain, then $T_{veh} = 0$; set the type parameter as S_r, and under the condition that the moving target is confirmed to be effective, that is $T_{veh} = 1$, when the target vehicle is a large vehicle, then $S_r = 1$, when the target vehicle is a car, then $S_r = -1$, when the target type is unknown, then $S_r = 0$. The classify methods are as follows:

$$T_{veh} = \begin{cases} 1, & \frac{w_{lp}}{w_{veh}} < \frac{1}{2.5}, \quad LP = 1 \\ 0, & \text{else} \\ -1, & 1 > \frac{w_{lp}}{w_{veh}} > \frac{1}{2}, \quad LP = 1 \end{cases}$$

$$S_r = \begin{cases} 1, & \frac{w_{\text{veh}}}{l_{\text{veh}}} \leq \frac{1}{6}, \quad T_{\text{veh}} = 1 \\ 0, & \text{else} \\ -1, & \frac{w_{\text{veh}}}{l_{\text{veh}}} > \frac{2.5}{6}, \quad T_{\text{veh}} = 1 \end{cases} \tag{10}$$

where w_{lp} is the license plate width,. is the vehicle target image width, l_{veh} is the body image length, w_{veh} is the body image width, is the license plate parameter, T_{veh} is the target effective parameter, S_r is the target type parameter.

We extract video in the road environment with single camera in this paper. Images usually have large distortion caused by both projective transform and the lens distortion. We using geometrical invariability of line to perform camera self-calibration from a single image which was first proposed by Fu [51]

2) LICENSE PLATE AND SHAPE CHARACTERISTICS

In the process of automobile design and manufacture, due to many factors such as manufacturing cost, aesthetics, vehicle breeding and aerodynamic balance, vehicles are basically designed with symmetry, except the special operating vehicles.

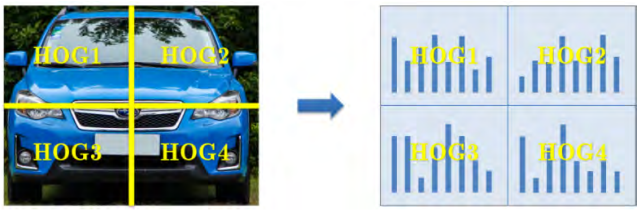

FIGURE 5. symmetry of HOG feature in an image.

Using the characteristics of contour and texture symmetry of the vehicle, this paper adopts the method of literature [13] to divide the vehicle image into two sets of symmetric regions (Fig. 5) by using the Hog feature.

The shape difference of the lower part of most vehicles is significant, and in order to avoid the adverse effects on the detection caused by glass reflection and the characteristics of the objects in the vehicle, the symmetric regions of HOG3 and HOG4 are selected as the research objects in this paper. S_3 and S_4 are eigenvectors of the HOG3 and HOG4 regions respectively. Each HOG feature is obtained from 8 bin, then the definition of symmetric vector S_3 and S_4 as follows [13]:

$$S_3 = [s_{31}, s_{32}, S_{33}, S_{34}, S_{35}, S_{36}, S_{37}, S_{38}] \tag{11}$$

$$S_3 j = \begin{cases} \frac{h_{li}}{\sum_{k=1}^{8} h_{lk}} / \frac{h_{rj}}{\sum_{k=1}^{8} h_{rk}}, & \frac{h_{li}}{\sum_{k=1}^{8} h_{lk}} / \frac{h_{rj}}{\sum_{k=1}^{8} h_{rk}} \\ \frac{h_{rj}}{\sum_{k=1}^{8} h_{rk}} / \frac{h_{li}}{\sum_{k=1}^{8} h_{lk}}, & \text{other} \end{cases} \tag{12}$$

S_{3j} represents the symmetrical relationship between left and right pixel h_{lj} and $\bar{h}_{rj}$ in the image. The symmetric vectors of symmetric features can be calculated by combining S_3 and S_4 vectors. The eigenvectors of the left and right sides in Fig.5 are $H_L = [h_{l1}, h_{l2}, h_{l3}, h_{l4}, h_{l5}, h_{l6}, h_{l7}, h_{l8}]$ and $H_R = [h_{r1}, h_{r2}, h_{r3}, h_{r4}, h_{r5}, h_{r6}, h_{r7}, h_{r8}]$, H_L and $\bar{H}_R$ are symmetrical.

$$\begin{aligned} \overline{H}_R &= [h_{r1}, h_{r2}, h_{r3}, h_{r4}, h_{r5}, h_{r6}, h_{r7}, h_{r8}]^T \\ &= [\bar{h}_{r1}, \bar{h}_{r2}, \bar{h}_{r3}, \bar{h}_{r4}, \bar{h}_{r5}, \bar{h}_{r6}, \bar{h}_{r7}, \bar{h}_{r8}] \end{aligned} \tag{13}$$

FIGURE 6. Examples of classification by face features. (a) Example of type I; (b) Example of type III.

Through the comparison of the HOG eigenvectors above, the detection target can be identified as whether a vehicle or not. In order to further adapt to the responsible environment, as well as improve the operation speed and accuracy, the HOG features are layered, and the principal component analysis (PCA) is used to extract the feature and reduce the eigenvector dimension. The hierarchical methods are: (1) Divide the grayscale image area into 9 blocks, (2) Dividing each block into 4 cells, (3) Divide each cell into 8 bins.

3) FACE FEATURES

For the actual road targets, the human face is one of the main characteristics when the target is oriented to the video acquisition equipment. The classification method-based features are as follows:

$$T_{veh} = \begin{cases} -1, & \dfrac{w_{veh}}{w_{face}} < 2.5 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where w_{face} is the width of the face image detected in the foreground area.

The examples of classification by face features in type I and type III were presented by Fig.6.

The size of the projection area of an object in a image is inversely proportional to the distance between object and the lens. So, on the basis of some of the target types that have been identified, we can classify the target types through the comparison of areas with them. is the targets set have been extracted by the camera, the target image $a, b \in R$, classification is as follows:

$$T^b_{veh} = \begin{cases} 1, & L_b \geq L_a, \quad A_b \geq A_a, \ T^a_{veh} = 1 \\ 0, & \text{else} \\ -1, & L_b \geq L_a, \quad A_b \geq A_a, \ T^a_{veh} = -1 \end{cases}$$
$$L_r = \sqrt{x_r^2 + y_r^2}$$
$$A = \tau n \tag{15}$$

where l_{veh} is the body image length, w_{veh} is the body image width, T_{veh} refers to the target effective parameter, S_r represents the target type parameter, the distance between the moving target centroid and the camera is L_r, x, y are the moving target centroid coordinates, A is the target image pixel area, τ refers to a single pixel area, and n is the target image number of pixels.

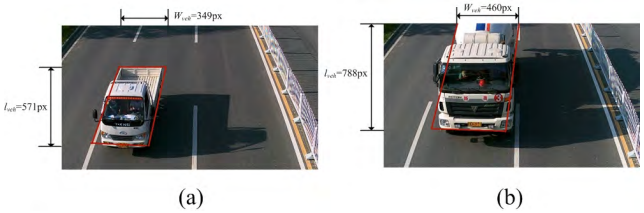

FIGURE 7. Example of an area comparison method in type I.(a) Target a;(b) Target b.

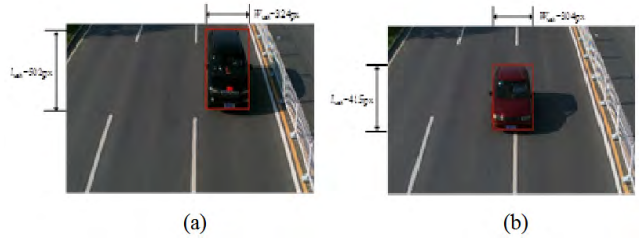

FIGURE 8. Example of an area comparison method in type II. (a) Target a;(b) Target b.

The examples of area comparison method in type I and type II were presented by Fig.7 and Fig.8 respectively.

4) CONSTRUCTION OF CASCADING CLASSIFIER

Traditional classification has a high false positive rate, and there are higher requirements for training sample through the deep learning. In order to deal with the influence of the complex environment in the actual situations, the prior knowledge of the vehicle appearance is used to gradually filter positive sample targets, which has a small amount of computation. By investing more time in the region of the target which is difficult to identify, the detection speed is effectively improved while the false alarm rate is significantly reduced. In practical application situation, several weak classifiers are generally combined into one classifier.

5) TRACKING CORRECTION

In this paper, we establish a tracking target database, which was used to correct target types in tracking process. $Z_i^k = (x, y, v_x, v_y, w, h, lpn, ID, Sta, T_{veh}, S_r)$ represents the information of the historical target i in the frame k, where x, y represent the location of the target, (v_x, v_y)represent the speed in the horizontal and vertical directions, w, h represent the width and height of the tracking frame, lpn represents the plate number, Sta is state identifier.

IV. EXPERIMENTS

A. EVALUATION DATASET

1) CDnet2014 DATASET

For moving object detection performance evaluation, we used the CDnet2014 benchmark data set proposed in [56], CDnet2014 dataset contains 11 video categories with 4-6 videos sequences in each category. The scenarios used to evaluate different methods contain bad weather, baseline,

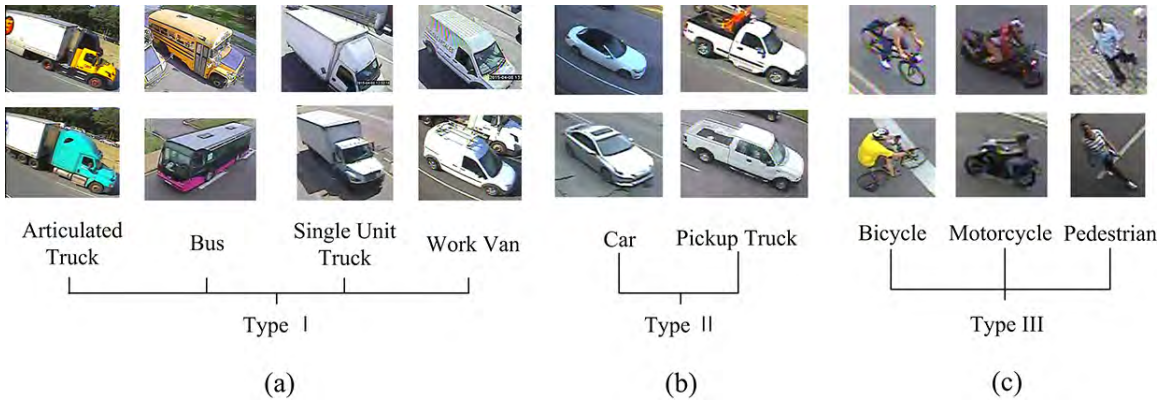

FIGURE 9. Reclassification of MIO-TCD dataset. (a) Type I includes articulated truck, bus, single unit truck and work wan; (b) Type II includes car, pickup truck; (c) Type III includes bicycle, motorcycle and pedestrian.

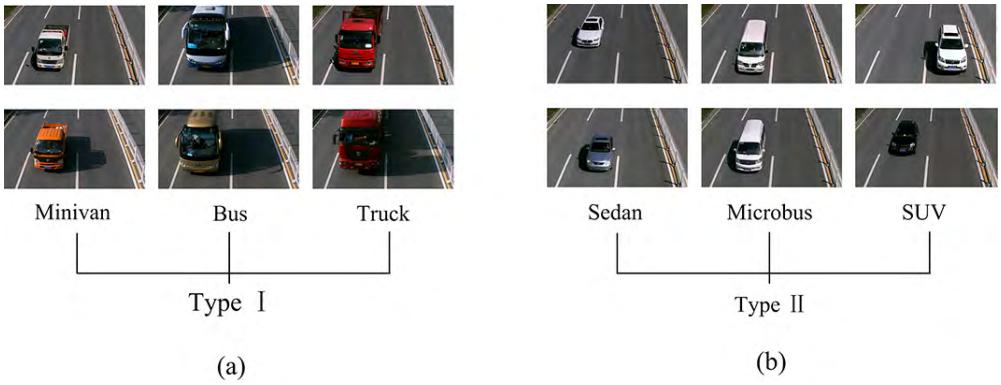

FIGURE 10. Reclassification of BIT-Vehicle dataset. (a) Type I includes minivan, bus and truck; (b) Type II includes sedan, microbus suv.

camera jitter, dynamic background, intermittent motion object and shadow.

2) MIO-TCD DATASET

MIO-TCD [57] is the largest dataset for motorized traffic analysis to date, which includes 11 traffic object classes such as cars, trucks, buses, motorcycles, bicycles, pedestrians. It contains 786,702 annotated images acquired at different times of the day and different periods of the year by hundreds of traffic surveillance cameras deployed across Canada and the United States. The dataset consists of two parts: a ''localization dataset'', containing 137,743 full video frames with bounding boxes around traffic objects, and a ''classification dataset'', containing 648,959 crops of traffic objects from the 11 classes, We reclassified the dataset into three categories as show in Fig.9.

3) BIT-VEHICLE DATASET

The BIT-Vehicle dataset [40] was used to test our method. The BIT-Vehicle dataset contains 9,850 vehicle images. There are images with sizes of 1600*1200 and 1920*1080 captured from two cameras at different time and places in the dataset. The images contain changes in the illumination condition, the scale, the surface color of vehicles, and the viewpoint. The top or bottom parts of some vehicles are not included in the images due to the capturing delay and the size of the vehicle. The dataset can also be used for evaluating the performance of vehicle detection. All vehicles in the dataset are divided into six categories: Bus, Microbus, Minivan, Sedan, SUV, and Truck. The number of vehicles per vehicle type are 558, 883, 476, 5,922, 1,392, and 822, respectively.6 vehicle types, 9,850 images. We reclassified the dataset into two categories as show in Fig.10.

B. PERFORMANCE MEASURE

There are seven metrics used in above datasets, the definitions of those metrics are as below [15], [56], [57]:

$$F-\text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision recall}},$$

$$PWC = 100 \times \frac{FN + FP}{TP + FN + FP + TN},$$

$$\text{specificity} = \frac{TN}{FP + TN},$$

$$\text{Precision} = \frac{TP}{TP + FP},$$

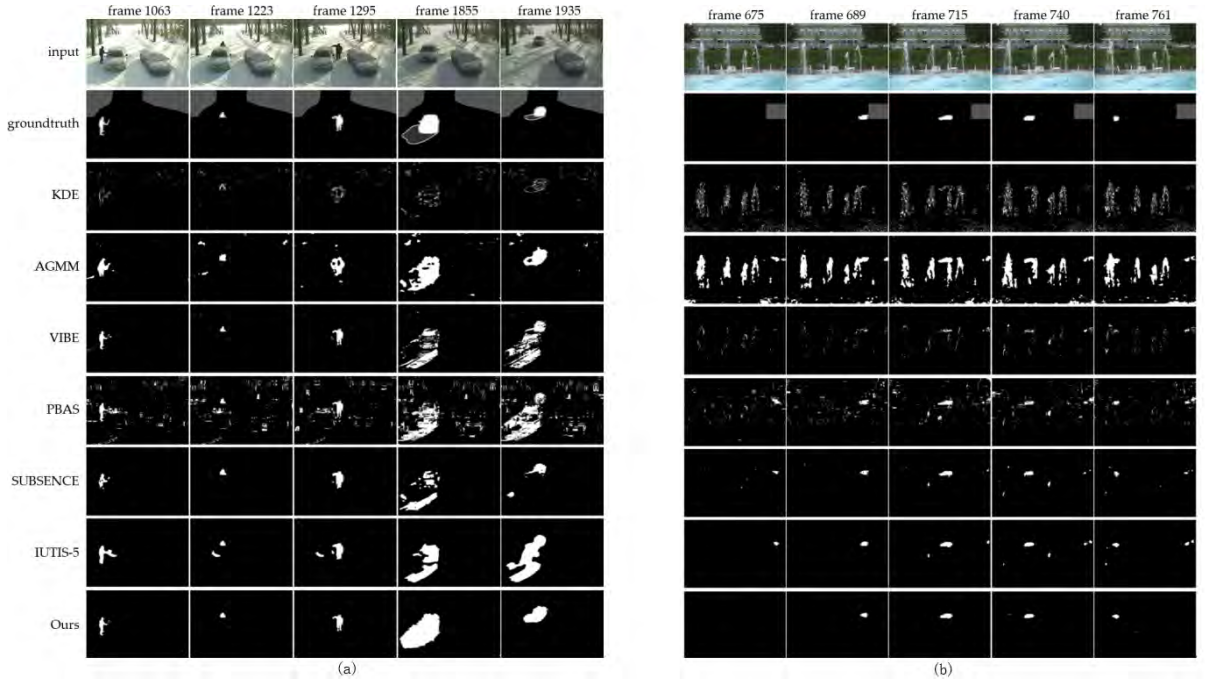

FIGURE 11. Comparison of the segmentation outputs, (a) The scenes winter driveway (intermittent object motion); (b) The scenes fountain (dynamic back ground). The first column is the input image, the second column is its ground truth image, the third column represents the output of the KDE [12]. The fourth column is the output of the AGMM [52], the fifth column represents the output of the Vibe [46], the sixth column represents the output of the PBAS [53], the seventh column represents the output of the SUBSENCE [54], the eighth column represents the output of the IUTIS-5 [55] and the last column shows the output of ours. Gray pixels in the ground truth segmentation indicate pixels which are not of interest.

TABLE 1. Average evaluation metrics of seven methods on CDnet 2014 dataset.

Method	Recall	Specificity	FPR	FNR	PWC	F-measure	Precision
KDE	0.7375	0.9519	0.0481	0.2625	5.6262	0.5688	0.5811
AGMM	0.5603	0.9719	0.0274	0.4397	2.4064	0.5389	0.6342
VIBE	0.4651	0.9868	0.0132	0.5349	3.041	0.4718	0.653
PBAS	0.5079	0.992	0.008	0.4921	2.411	0.5505	0.709
SUBSENCE	0.8124	0.9904	0.0096	0.1876	1.678	0.7408	0.7509
IUTIS-5	0.7849	0.9948	0.0052	0.2151	1.1986	0.7717	0.8087
Ours	0.7647	0.9902	0.0058	0.2353	0.7164	0.7786	0.8373

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN}, \\ FNR &= \frac{FN}{FN+TP}, \\ FPR &= \frac{FP}{FP+TN} \end{aligned} \tag{16}$$

C. RESULTS OF DETECTION

1) QUALITATIVE

We choose six representative scenes from each category of CDnet2014, they are highway, bad weather, dynamic back ground, camera Jitter, intermittent object motion and shadow. We visually compare them with each other and against the ground truth. As show in Fig.11 (a), our method has better robustness in winter driveway scenes (intermittent object motion), which has challenges such as illumination variation, moving object appearance changes, complex background and shadow, Fig.11 (b) shows the effect of our method in fountain scenes (dynamic back ground). The results of six methods mentioned above are from CDnet2014 and they are presented in Fig.12 with ours.

2) QUANTITATIVE

Seven metrics on above six scenes of CDnet 2014 dataset were calculated under all methods, and the results are presented in Table.1. We notice that our method achieves a promising result compared with the other six methods, especially obtains the best in terms of ''PWC'', ''F-measure'' and ''Precision''. The details of each metrics of different methods in six senses are presented in Fig.13

D. EVALUATION DATASET

Since our classification method is based on the prior knowledge of vehicle features, there is no need for training. The classification method is based on the national standards of

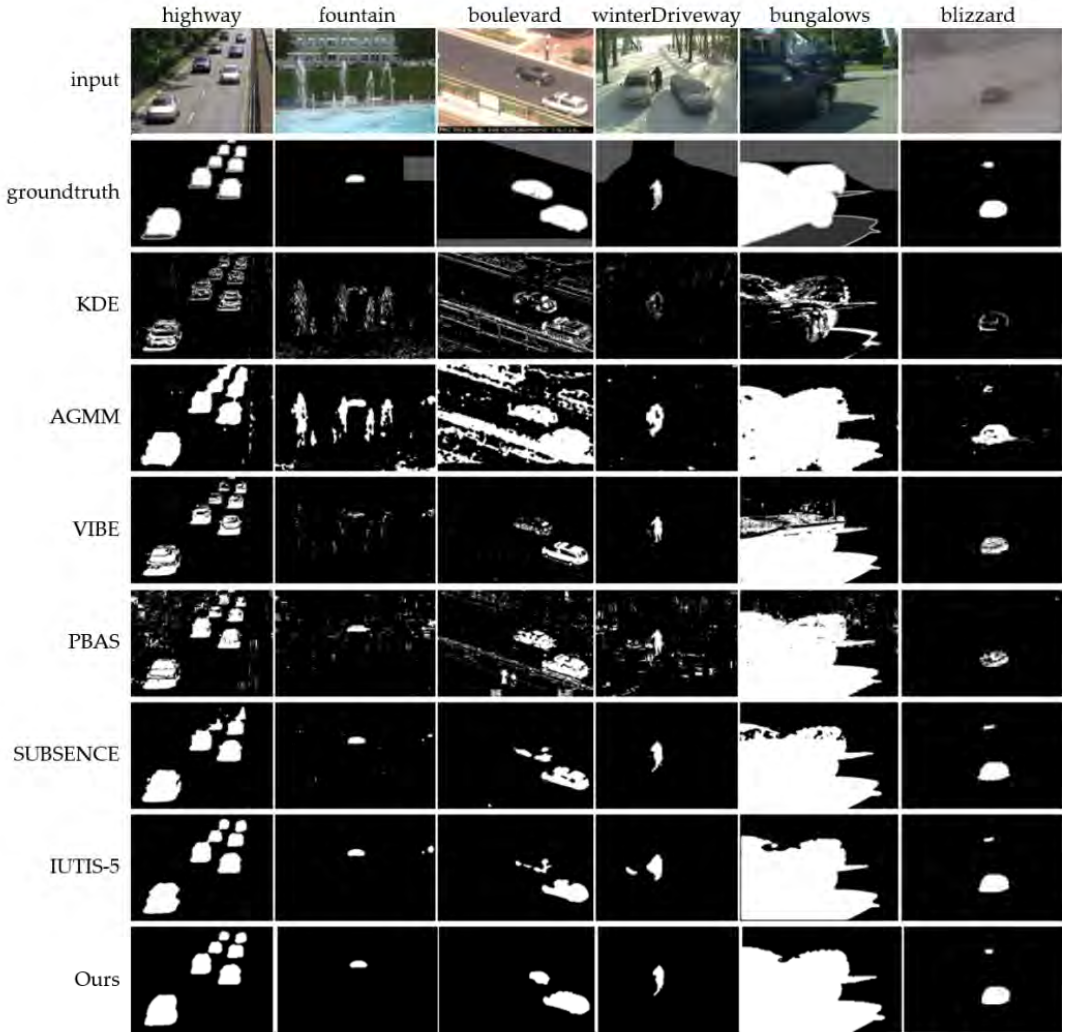

FIGURE 12. Results of six classical methods compared with ours. From first column to sixth column are the scenes: Highway (baseline), fountain (dynamic back ground), boulevard (camera Jitter), winter driveway (intermittent object motion), bungalows (shadow) and blizzard (bad weather); from first row to last row are input, ground truth, the results of KDE, AGMM, Vibe, PBAS, SUBSENSE, IUTIS-5 and ours.

TABLE 2. Classification confusion matrix on the MIO-TCD classification challenge dataset.

	Type I	Type II	Type III
Type I	8711	157	0
Type II	678	76172	1008
Type III	0	105	2526

TABLE 3. The overall results on the MIO-TCD dataset.

Method	Accuracy	FRP(type I)	FRP (typeII)	FRP (typeIII)
Googlenet[62]	0.9327	0.0541	0.0578	0.0417
Resnet[63]	0.9576	0.0415	0.0394	0.0359
EDLM[64]	0.9784	0.0228	0.0199	0.0303
EDeN[65]	0.9795	0.0221	0.0191	0.0211
Ours	0.9782	0.0177	0.0217	0.0399

vehicle and license plate in China. The BIT-Vehicle dataset was selected to validate our method, however, the BIT-Vehicle dataset does not contain type targets, so the performance of the MIO-TCD dataset was calculated and common verify the effective of our method.

1) RESULTS ON MIO-TCD DATASET

The MIO-TCD test dataset contains 89357 images can be divided into three types: typeI: 8868, typeII: 77858, typeIII: 2631, TABLE2 present the classification confusion matrix of the proposed method on the MIO-TCD classification challenge dataset.

The performance comparison with other models are shown in TABLE 3, with regard to the overall performance, we got 0.9782 classification accuracy, 0.0177 FRP of type I, and 0.0217 FRP of type II. Since the randomness of MIO-TCD test dataset contains image shooting angle and low resolution,

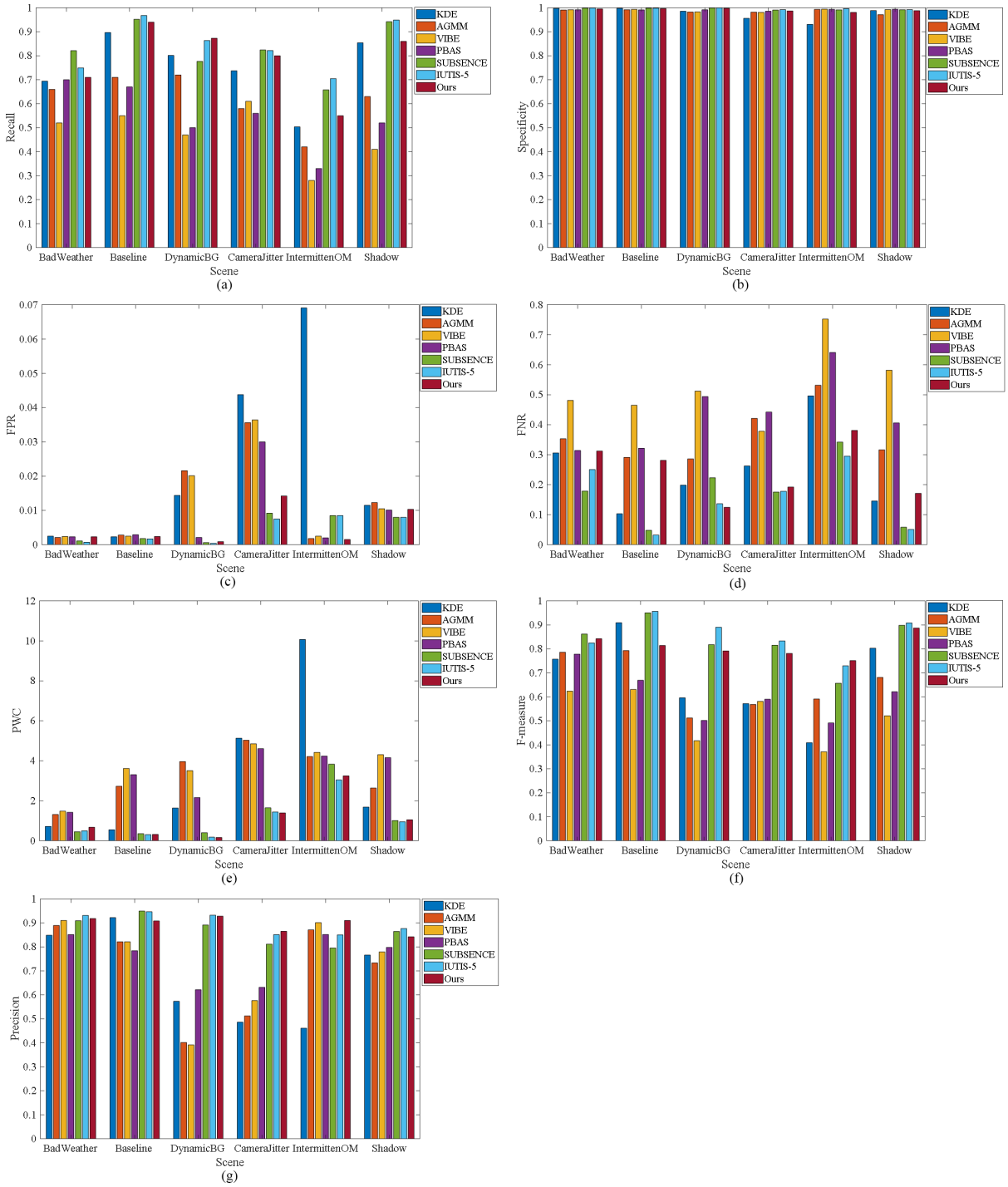

FIGURE 13. Each metrics comparison of different methods for six scenes. (a) The recall of different methods; (b) The Specificity of different methods; (c) The FPR of different methods; (d) The FNR of different methods; (e) The PWC of different methods; (f) The F-measure of different methods; (g) The Precision of different methods.

besides image were taken in North America. We did not get good effect in accuracy, but done well in FRP.

2) RESULTS ON BIT-VEHICLE DATASET

The BIT-Vehicle test dataset contains 89357 images which can be divided into three types: type I 1856, type II 8197. TABLE 4 present the classification confusion matrix of the proposed method on the BIT-Vehicle dataset. The performance comparison with other models are shown in TABLE 5, with regard to the overall performance, we got a classification accuracy of 0.9922, 0.0075 FRP of type I, and 0.0078 FRP of type II.

TABLE 4. Classification confusion matrix on the BIT-Vehicle dataset.

	Type I	Type II
Type I	1842	14
Type II	64	8133

TABLE 5. The overall results on the BIT-Vehicle dataset.

Method	Accuracy	FRP(type I)	FRP (typeII)
Dong's2013[66]	0.9692	0.0496	0.0374
Dong's 2014[41]	0.9816	0.0226	0.0152
Dong's 2015[67]	0.9888	0.0197	0.0093
Ali's 2018[68]	0.9907	0.0153	0.0085
Ours	0.9922	0.0075	0.0078

Due to the high definition of the vehicle frontal view images contained in the BIT-Vehicle dataset, additional images are captured from two cameras in China. Our method performs better on BIT-Vehicle dataset than MIO-TCD test dataset, both of accuracy and FRP.

V. CONCLUSIONS

In complex environment, there are many interference factors in the detection and classification of vehicles in traffic monitoring video streaming. In order to reduce the false alarm rate, this paper proposes a motion vehicle detection and type classification algorithm named improved spatio-temporal sample consensus (ISTSC). It first takes the spatio-temporal sample consensus algorithm to detect moving objects, from the interference of illumination variation and shadow on vehicle identification. Secondly, it classify the objects through the feature fusion methods considering vehicle symmetry, plate number, area, sharp and face features. Experimental results on three prevalent data sets showed the good effectiveness of our method on vehicle detection and type classification.

Through the experimental analysis of the actual road traffic monitoring video data, the method proposed in this paper has obvious advantages in false alarm rate and real-time performance. It can fulfill the accuracy requirements of road traffic condition evaluation based on the video speed measurement. In the future, the focus of the study will be the detection of vehicles and the further segmentation of confusing vehicle models, under the situation of low Illumination at night.

REFERENCES

[1] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation: An error analysis of gradient-based methods with local optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 2, pp. 229–244, Mar. 1987.

[2] J. Fan, R. Wang, L. Zhang, D. Xing, and F. Gan, "Image sequence segmentation based on 2D temporal entropic thresholding," *Pattern Recognit. Lett.*, vol. 17, no. 10, pp. 1101–1107, Sep. 1996.

[3] B. Shoushtarian and H. E. Bez, "A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking," *Pattern Recognit. Lett.*, vol. 26, no. 1, pp. 5–26, 2005.

[4] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1937–1944.

[5] Y.-K. Jung, K.-W. Lee, and Y.-S. Ho, "Content-based event retrieval using semantic scene interpretation for automated traffic surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 2, no. 3, pp. 151–163, Sep. 2001.

[6] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event detection and analysis from video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 873–889, Aug. 2001.

[7] R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 2, pp. 119–130, Jun. 2000.

[8] B. Maurin, O. Masoud, and N. Papanikolopoulos, "Monitoring crowded traffic scenes," in *Proc. IEEE 5th Int. Conf. Intell. Transp. Syst.*, Sep. 2002, pp. 19–24.

[9] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1999, pp. 246–252.

[11] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit.*, Aug. 2004, pp. 28–31.

[12] A. M. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 751–767.

[13] Q. T. Geng, Y. U. Fan-Hua, Y. T. Wang, and Q. K. Gao, "New algorithm for vehicle type detection based on feature fusion," J. Jilin Univ., Tech. Rep., 2018.

[14] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.

[15] X. Zeng, G. Xu, X. Zheng, Y. Xiang, and W. Zhou, "E-AUA: An efficient anonymous user authentication protocol for mobile IoT," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1506–1519, Apr. 2019.

[16] 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.

[17] Z. Chen and T. Ellis, "Multi-shape descriptor vehicle classification for urban traffic," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Dec. 2011, pp. 456–461.

[18] B. Stenger, V. Ramesh, N. Paragios, F. Coetzee, and J. M. Buhmann, "Topology free hidden Markov models: Application to background modeling," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, Jul. 2001, pp. 294–301.

[19] Y. Wang, K.-F. Loe, and J.-K. Wu, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289,Feb. 2006.

[20] M. Heikkila and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.

[21] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1301–1306.

[22] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 179–186, 2010.

[23] D. Chung, W. J. Maclean, and S. Dickinson, "Integrating region and boundary information for improved spatial coherencein object tracking," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun./Jul. 2004, p. 3.

[24] H.-G. Kang and D. Kim, "Real-time multiple people tracking using competitive condensation," *Pattern Recognit.*, vol. 38, no. 7, pp. 1045–1058, Jul. 2005.

[25] D. Cremers and S. Soatto, "Motion competition: A variational approach to piecewise parametric motion segmentation," *Int. J. Comput. Vis.*, vol. 62, no. 3, pp. 249–265, 2005.

[26] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[27] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[28] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[29] S. J. Pundlik and S. T. Birchfield, "Motion segmentation at any speed," in *Proc. BMVC*, Sep. 2006, pp. 427–436.

[30] H. P. Moravec, "Visual mapping by a robot rover," in *Proc. IJCAI*, 1979, pp. 598–600.

[31] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[33] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 3951. May 2006, pp. 404–417.

[34] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2001, pp. 511–518.

[35] J. Wu, C. Geyer, and J. M. Rehg, "Real-time human detection using contour cues," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2011, pp. 860–867.

[36] Z. Zhu, Y. Zhao, and H. Lu, "Sequential architecture for efficient car detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[37] P. Negri, X. Clady, M. Milgram, and R. Poulenard, "An oriented-contour point based voting algorithm for vehicle type classification," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2006, pp. 574–577.

[38] A. Psyllos, C.-N. Anagnostopoulos, and E. Kayafas, "Vehicle model recognition from frontal view image measurements," *Comput. Standards Interfaces*, vol. 33, no. 2, pp. 142–151, Feb. 2011.

[39] Y. Peng *et al.*, "Vehicle type classification using data mining techniques," in *The Era of Interactive Media*. New York, NY, USA: Springer, 2013.

[40] Z. Dong, M. Pei, Y. He, T. Liu, Y. Dong, and Y. Jia, "Vehicle type classification using unsupervised convolutional neural network," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 172–177.

[41] X. Ma and W. E. L. Grimson, "Edge-based rich representation for vehicle classification," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2005, pp. 1185–1192.

[42] C. Zhang, X. Chen, and W.-B. Chen, "A PCA-based vehicle classification framework," in *Proc. 22nd Int. Conf. Data Eng. Workshops (ICDEW)*, Apr. 2006, p. 17.

[43] P. Ji, L. Jin, and X. Li, "Vision-based vehicle type classification using partial Gabor filter bank," in *Proc. IEEE Int. Conf. Automat. Logistics*, Aug. 2007, pp. 1037–1040.

[44] Y. Shan, H. S. Sawhney, and R. Kumar, "Unsupervised learning of discriminative edge measures for vehicle matching between non-overlapping cameras," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 894–901.

[45] L. Lin, Y. Xu, X. Liang, and J. Lai, "Complex background subtraction by pursuing dynamic spatio-temporal models," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3191–3202, Jul. 2014.

[46] Q. Jiang, M. K. Khan, X. Lu, J. Ma, and D. He, "A privacy preserving three-factor authentication protocol for e-health clouds," *J. Supercomput.*, vol. 72, no. 10, pp. 3826–3849, 2016.

[47] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

[48] J. Yao and J.-M. Odobez, "Multi-layer background subtraction based on color and texture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[49] N. C. Mithun, N. U. Rashid, and S. M. M. Rahman, "Detection and classification of vehicles from video using multiple time-spatial images," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1215–1225, Feb. 2012.

[50] Y. Wang, X. Ban, J. Chen, B. Hu, and X. Yang, "License plate recognition based on SIFT feature," *Optik*, vol. 126, no. 21, pp. 2895–2901, Nov. 2015.

[51] F. Dan, "A method of camera calibration using geometrical invariability of line," *J. Image Graph.*, 2009.

[52] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit.*, Aug. 2004, pp. 28–31.

[53] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on cooccurrence of image variations," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 65–72.

[54] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 38–43.

[55] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[56] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 393–400.

[57] Z. Luo, F. Branchaud-Charron, C. Lemaire, J. Konrad, S. Li, A. Mishra, A. Achkar, J. Eichel, and P.-M. Jodoin, "MIO-TCD: A new benchmark dataset for vehicle classification and localization," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5129–5141, Oct. 2018.

[58] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[60] W. Liu, M. Zhang, Z. Luo, and Y. Cai, "An ensemble deep learning method for vehicle type classification on visual traffic surveillance sensors," *IEEE Access*, vol. 5, pp. 24417–24425, 2017.

[61] R. Theagarajan, F. Pala, and B. Bhanu, "EDeN: Ensemble of deep networks for vehicle classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017.

[62] Z. Dong and Y. Jia, "Vehicle type classification using distributions of structural and appearance-based features," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4321–4324.

[63] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015.

[64] A. Şentaş, I. Tashiev, F. Küçükayvaz, S. Kul, S. Eken, A. Sayar, and Y. Becerikli, "Performance evaluation of support vector machine and convolutional neural network algorithms in real-time vehicle type and color classification," in *Evolutionary Intelligence*. 2018, pp. 1–9.