Adaptive Forward Vehicle Collision Warning Based on Driving Behavior

Yuan Yuan[a], Yuwei Lu[a], Qi Wang[a,*]

[a]*School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China*

Abstract

Forward Vehicle Collision Warning(FCW) is one of the most important functions for the Advanced Driver Assistance System (ADAS). In this procedure, vehicle detection and distance measurement are core components, requiring accurate localization and estimation. In this paper, we propose a simple but efficient forward vehicle collision warning framework by aggregating monocular distance measurement and precise vehicle detection. In order to obtain forward vehicle distance, a quick camera calibration method which only needs three physical points to calibrate related camera parameters is utilized. As for the forward vehicle detection, a multi-scale detection algorithm that regards the result of calibration as distance prior is proposed to improve the precision. What's more, traditional deterministic FCW approaches cannot be personalized for different drivers, which will lead to false warnings when drivers are in diverse driving status. Therefore, abnormal driver behaviors are introduced to make FCW adaptive. Specifically, the proposed adaptive FCW generates warnings by considering the different behaviors of the driver. Intensive experiments are conducted in our established real scene dataset and the results have demonstrated the effectiveness of the proposed framework.

Keywords: Advanced driver assistance system (ADAS), adaptive forward vehicle collision warning, abnormal driver behavior, multi-scale detection

*Corresponding author

Email address: crabwq@gmail.com (Qi Wang)

1. Introduction

Over 10 million people are injured yearly worldwide in road accidents. Among these accidents, rear-end collision is a serious safety problem, accounting for almost 30% of all crashes [1]. Significant effort has been made on the safety of road vehicles in recent decades. Advanced driver assistance system(ADAS) plays a significant role in increasing the safety of passengers and of vehicles. Forward vehicle collision warning (FCW) as one of the fundamental techniques in ADAS is the core function to mitigate rear-end collisions.

A series of devices mounted on the vehicle could provide the solution to FCW [2, 3]. The traditional systems available today are typically based on radar sensors [4]. However, the narrow field of view and the poor lateral resolution limit the performance of these systems. From a technological point of view, fusion of radar and vision information seems to be an attractive way. In such systems [5, 6] the radar provides accurate distance and velocity, while vision obtains exact locations of the forward vehicle. Unfortunately, expensive and complex course of fusing radar and vision degrades its practicability. Given these practical difficulties, a simple but efficient forward vehicle collision warning framework is proposed using only vision information in this paper. The proposed framework includes two stages. The camera calibration stage gets distance from forward vehicles to the camera, while the vehicle detection stage based on the distance provides exact locations of forward vehicles.

On the other hand, in the conventional designs of FCW the warnings are triggered with a deterministic model whenever a potential collision is detected [7, 8]. However, utilizing the deterministic warning model makes these approaches unable to adapt the different driving behaviors of each individual driver and prevents them from making decisions for different driving behaviors. Therefore, adaptive algorithms have been proposed to tackle this issue and achieve enhanced versions of the FCW systems [9, 10, 11].

Every driver has its own driving style, which seriously affects his decisions and reactions in different driving situations. These personal mental and physical characteristics, which are summarized as driving behaviors in the literature [12], can be studied to generate adaptive FCW systems. Different behaviors could be employed in order to build adaptive FCW algorithms. For instance, braking and steering styles are two descriptive indicators which have been extensively used to make up an adaptive model [9, 13].

Under normal driving conditions, the driver is assumed to be fully focused on his driving and pay enough attention to road conditions. However, the driver's attention is often dispersed in some situations such as talking with passengers and fatigued driving. In this paper, we call these distracting driving as abnormal driver behaviors. These abnormal behaviors which can be considered as one of the driver behavioral modes need to be modeled carefully, since it may lead to dangerous. For example, 19% of United States total fatalities 2016 are due to alcohol impaired driving, which is one of the most important abnormal driver behaviors, as reported by National Highway Traffic Safety Administration (NHTSA)[14]. To this end, ADAS applications should be appropriately designed to generate warnings adaptive to the abnormal driver behavior. This adaptive design requires a reliable abnormal driving detection mechanism.

In this paper, the abnormal driver behaviors are introduced to build FCW via an in-vehicle camera to generate collision warnings adaptively. Different from monitoring vehicle signals, such as acceleration, braking, etc., we propose an abnormal driver behavior detection method by directly detecting the driver's face. To the best of our knowledge, adaptive FCW based on abnormal driver behavior has not been proposed in the literature.

The main contributions of our framework are as follows: First, a simple but effective framework is proposed for forward vehicle collision warning. Since it is based on vision information, the framework is inexpensive and easy to setup. Second, distance information is applied to improve the performance of forward vehicle detection. Third, an abnormal driver behavior detection method is proposed to make the proposed FCW adaptive via an in-vehicle camera.

This work is an extension of our earlier conference paper [15]. The more detailed method description and the further experimental analysis are shown in this version.

The rest of this paper is organized as follows. Section 2 introduces the related work and Section 3 describes the proposed framework. Experimental results are demonstrated in Section 4 while conclusion is presented in Section 5.

2. Related Work

In this section, we will first review the camera calibration algorithms. Then, some detection methods that are used in FCW will be introduced. Finally, current adaptive FCW approaches will be discussed.

2.1. Camera Calibration

Camera calibration has been studied extensively in computer vision and photogrammetry. According to the dimension of the reference, calibrating methods can be roughly classified into four categories as follows:

3D object-based calibration. Techniques in this category are required to observe a calibration object whose geometry in 3D space is know with very good precision. Calibration can be done efficiently [16]. Since the calibration object always consists of two or three planes orthogonal to each other. Sometimes a plane undergoing a precisely known translation is also used [17]. These approaches require an expensive calibration apparatus, and a complex setup.

2D object-based calibration. Camera calibration is performed by observing a planar pattern shown at a few different orientations [18, 19]. Different from 3D-based calibration methods[17], the knowledge of the plane motion is not necessary. Since such a calibration pattern is easy to be made, the setup becomes more uncomplicated.

1D object-based calibration. *Zhang* proposes one-dimensional object based calibration in 2004[20]. This method considers 1D objects composed of a set of collinear points. It uses poorer knowledge of the observation compared to 2D and 3D object-based calibration methods. However, since the observed object is too sample, the accuracy of calibration is relatively poor.

Self-calibration. Techniques in this category don't utilize any calibration objects [21]. By moving a camera in a static scene, the internal parameters of the camera will be estimated with image information alone. Though no calibration objects are necessary, a large number of parameters still need to be estimated. Computational complexity will be greatly increased.

2.2. Vehicle Detection

Recent years many deep learning methods have been proposed for computer vision. Wang *et al.* put forward the attention model [22, 23]. Dong *et al.*[24] perform a quadruplet network. Wang *et al.*[25] propose a dynamic fully convolutional network. A deep Q-learning model is introduced by Shen *et al.*[26]. [27] formulate triplet loss in Siamese network. However, considering the application scenarios of forward vehicle collision warning, traditional methods will be mainly reviewed in this section. Vehicle detection approaches are divided into two types : template-based and appearance-based.

Template-based methods. Methods in this category apply predefined patterns from the vehicle class and perform correlation between the image and the template. Li *et al.* [28] propose an And-Or model that integrates context and occlusion for detecting vehicles. Felzenszwalb *et al.* [29] propose deformable part models(DPM) to structure template model. Each model is composed of parts with different viewpoints. They detect vehicles by comparing the similarity of each hypothesis and the DPM models. Leon *et al.* [30] put forward a template-based approach using mixture of deformable parts models. They expand the original DPM [29] to adapt to crowded scenes. Wang *et al.* [31] also propose a probabilistic inference framework based on part models for improving detection performance. Since these methods detect vehicles by matching template, they are time consuming.

Appearance-based methods. Appearance-based methods learn the features of vehicles from a set of training images which should capture the variability in vehicle appearance. Usually, appearance models treat a two-class pattern classification problem: vehicle and nonvehicle. Wu and Zhang [32] apply standard Principal Components Analysis (PCA) for extracting global features to detect vehicles. Owing to small training data set, it is difficult to draw any meaningful conclusions. Li *et al.* [33] employ segmentation and neural network classifier for distinguishing vehicles from background. Khammari *et al.* [34] add depth image to set up their appearance models. Apart from the observed features, Zheng *et al.* [35] design image strip features based on the vehicle structure for vehicle detection. Since features come from the side view of the vehicle, this detector is sensitive to the viewpoint. Dollar *et al.* [36] propose aggregate channel features (ACF) and Yuan *et al.* [37] improve the features for detection.

2.3. Adaptive FCW

Many variations of deterministic FCW systems have been proposed in literature and most of them are evaluated in NHTSA reports [7]. Due to deterministic FCW cannot adapt different driving behaviors, only adaptive FCW methods will be reviewed in this section.

Nakaoka *et al.* [38] regard the variable road friction coefficient as a parameter to slightly adapt the warning generation criteria of FCW. However, there are no specific driver characteristics in the proposed adaptive framework. Chang *et al.* [39] employ a fuzzy logic based algorithm to differentiate various bus drivers' behaviors in terms of some driving parameters, such as perception-reaction time and braking deceleration, to improve the accuracy of warning in the bus rear-end collision scenarios. Wang *et al.* [6] propose a model for driver's risk perception to individualize collision warning generation process. In this model, the accuracy of the generated warning is tuned according to drivers braking data.

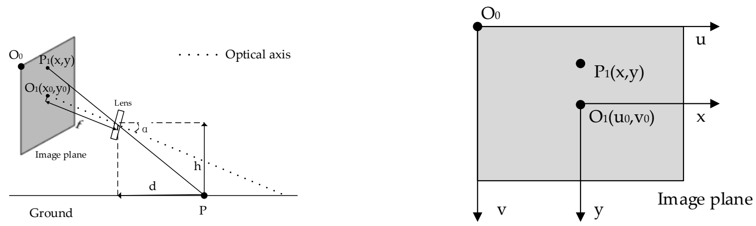
Although the concept of adaptive generating collision warning has been proposed, it seems that the abnormal driver behaviors have not been considered as an adaptive factor so far. Since abnormal driver behaviors greatly affects drivers' crash avoidance reactions, taking it into consideration in designing adaptive collision warning systems will improve the performance.

Abnormals driver behaviors have been considered as a special driver situation and different modeling methods have been proposed in the literature to identify its aspects. Different parameters such as unusual increasing velocity, increasing distance from the leading vehicle, abrupt steering wheel movements, reduction of control on lateral movements, slow reactions to the brake action of leading vehicle and changes in the driver's normal glance pattern have been considered as the different distraction indicators in the literature [40, 41].

Iranmanesh *et al.* [42] firstly utilize driver distraction which is one of the abnormal driving styles to build an adaptive FCW system. They target a combinational design of adaptive safety systems with driver distraction detection in order to reduce annoying false warnings while preserving the required ones. However, they indirectly utilize driver distraction to build adaptive FCW by braking data. As a result, the abnormal driving behavior that they find are limited. Therefore, in order to take full advantage of abnormal driver behaviors, we detect them by analyzing the facial expression of the driver in this paper. Owing to directly using facial information, more abnormal behaviors of the driver can be found and the performance of FCW will also be better.

3. Our Method

As mentioned before, the proposed framework will be introduced with two parts: 1) vision based FCW with camera calibration and multi-scale vehicle detection. 2) adaptive FCW with abnormal driver behaviors. To simplify the calibration course, a point-based calibration method [43] is employed to get camera parameters and to calculate distance from the forward car. During the detection course, we expand original ACF detector [36] into a distance-based multiple scale detector. The distance is not only used for forward collision warning, but also employed for improving vehicle detection. When generating collision warnings, an in-vehicle is used to collect driver's facial information. A modified version of DSOD [44] is implemented to detect abnormal driver behaviors.



(a) (b)

Figure 1: The pinhole imaging model of forward point P. (a) is the projection model and Eq 1 is derived from it; (b) shows the relation of idealized image coordinate system xO_1y to camera's pixel location coordinate system uO_0v.

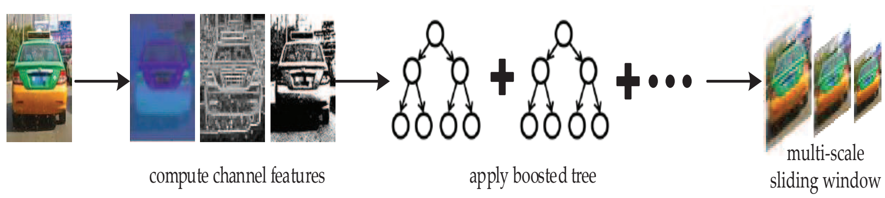


Figure 2: Overview of our detection framework.

3.1. Point-based calibration

Camera calibration is a necessary step in distance measuring with monocular vision. In engineering practice, the object distances are usually considerably larger than the focal length of camera. Hence, the pinhole camera model can be used to measure the distance. The geometry relationship of actual point P on the ground and its projection point on the image plane P_1 is shown in Fig 1(a). According to [45], the distance from point P to camera is:

$$d = \frac{h}{\tan(\alpha + \arctan[|(y_0 - y)/f|])}, \tag{1}$$

Here, α is the pitch angle of the camera; h is the height of the camera from the ground; (x_0, y_0) is the cross point of optical axis of the camera and the image plane; and y is the vertical coordinate of P_1. In order to simplify the calibration process, let dx, dy denote the physical dimension of one pixel along the x-axis and the y-axis separately. Then the coordinates of point P_1 in the image physical coordinate plane xO_1y and its position in the image pixel reference frame uO_0v are related by the transformation equation:

$$u = \frac{x}{dx} + u_0, v = \frac{y}{dy} + v_0, \tag{2}$$

In theory, as the corresponding pixel location of (x_0, y_0), (u_0, v_0) usually locates in the center of image. But in fact, there might be slight departure due to fabrication. In that case, u_0 and v_0 need to be measured. So, Eq. 1 can be expressed as

$$d = \frac{h}{\tan(\alpha + \arctan[|(v_0 - v)/f_y|])}. \tag{3}$$

Here, $f_y = f/dy$. Hence, we can get the distance d by solving the ratio f_y rather than calculating the optical length and pixel physical dimension separately.

In practice, the height of camera h can be measured after the camera is mounted on the car. Therefore, the distance from forward point to camera is determined by the camera parameters f_y, v_0, α, and the vertical coordinate of point in the pixel coordinate system v. Supposed that we have already got three calibration points. Their distances from camera are (d_1, d_2, d_3), and locations $(u_1, v_1), (u_2, v_2), (u_3, v_3)$ in pixel coordinate system. The height of camera from the ground h is measured. Then we can get the camera parameters by solving equations as below:

$$\begin{cases} d_1 = h/\tan(\alpha + \arctan[(v_0 - v_1)/f_y]) \\ d_2 = h/\tan(\alpha + \arctan[(v_0 - v_2)/f_y]) \\ d_3 = h/\tan(\alpha + \arctan[(v_0 - v_3)/f_y]) \end{cases} \tag{4}$$

Obviously, Eq. 4 is hard to solve as a nonlinear equation system. To simplify the calculation, we established two linear equation systems by variable substitution. Firstly, Eq. 3 can be written as

$$d = h\frac{f_y - B(v - v_0)}{Bf_y + (v - v_0)}, \tag{5}$$

where $B = \tan\alpha$. Let

$$C = d \cdot v, \tag{6}$$

Eq. 5 can be expressed as

$$d(v_0 - Bf_y) + h(f_y + Bv_0) - vBh = C. \tag{7}$$

Next, let

$$\begin{cases} x_1 = v_0 - Bf_y \\ x_2 = f_y + Bv_0 \\ x_3 = B \end{cases} \tag{8}$$

Eq 7 can be transformed into

$$dx_1 + hx_2 - vhx_3 = C. \tag{9}$$

Eq. 8 means that if x_1, x_2, x_3 are known, f_y, v_0, B can be calculated. Then the values of f_y, v_0, α will be obtained by the definition of $B = \tan\alpha$. In order to solve x_1, x_2, x_3, we write Eq. 9 into its matrix form as follows

$$\mathbf{x} = \mathbf{M}^{-1}\mathbf{c}. \tag{10}$$

Here,

$$\mathbf{M} = \begin{bmatrix} d_1, & h, & -v_1h \\ d_2, & h, & -v_2h \\ d_3, & h, & -v_3h \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}, \tag{11}$$

Camera height h, distance from points to the camera (d_1, d_2, d_3) and corresponding coordinates (v_1, v_2, v_3) are known. What's more, C_1, C_2, C_3 can be calculated according to Eq. 6. Put all of them into Eq.10, x_1, x_2, x_3 can be solved and the results of Eq. 9 are

$$\begin{cases} B = x_3 \\ f_y = x_2 - x_3(x_1 + x_2x_3)/(x_3^2 + 1). \\ v_0 = (x_1 + x_2x_3)/(x_3^2 + 1) \end{cases} \tag{12}$$

According to $B = \tan\alpha$, camera parameters f_y, v_0, α are calculated. We can obtain distance by Eq. 3. Up till now, we have estimated the relevant parameters for measuring distance from forward vehicle to the camera.

This algorithm needs only three fixed points to complete the calibration, which greatly reduces the computational complexity. Different from traditional ones in Section 2.1, this method estimates less parameters (only f_y, v_0, α and h) with the purpose of measuring distance. Estimating less parameters makes the calibration course easy to setup.

3.2. Multi-scale detection

When detecting forward vehicles, one of the greatest challenge is that vehicles have various scales at different distances. Multi-scale and multi-aspect ratio make this problem difficult. Due to perspective principle of the camera, the features of vehicles will change with different size. The structural feature is significant when the forward vehicle is near. However, when forward vehicles are far, they are made up of a few pixels in the image plane. We can hardly get structural features in this situation. Therefore, we apply color based features [36] to detect forward vehicles and distance information is employed to tackle multi-scale problem.

The proposed detection framework is exhibited in Fig. 2. Given an input image I, we compute its channel features. Then the boosting is used to train and combine decision trees over these channel features to distinguish object from background. Next, a distance based multi-scale sliding window approach is employed to detect vehicles. Fig.3 illuminates the major differences between original ACF detector and the proposed detector. Windows with various scales will slide the whole image in the ACF detector, while the proposed detector uses several windows with certain scale and aspect ratio to slide part of the image. Due to applying diverse windows in different vertical coordinates, the proposed method will be less time consuming.

The scale of sliding windows is related to distances between cars and the camera. Eq. 3 can be changed into the following form:

$$v = v_0 - f_y \tan(\arctan \frac{h}{d} - \alpha). \tag{13}$$

Eq. 13 is the foundation of multi-scale detection with distance prior. It indicates

(a) original ACF detector (b) the proposed detector

Figure 3: The difference between original ACF detector and our detector.

that if v_0, f_y and α are estimated, the vertical coordinate can be obtained by giving the real distance d. Therefore, we build a mapping from forward distance to the vertical coordinate in the image plane.

The distance prior can be calculated with a calibrated camera according to Eq. 13. When camera calibration is completed, we can not only obtain distance from for ward vehicles to the camera, but also get locations in the image plane according to the distance conversely. Table 1 demonstrates some scales of sliding windows in different distance.

Table 1: The change of aspect ratio at different distances

distance / m	**scale**	**vertical coordinate / pixel**
5	400 × 275	482
10	110 × 95	325
20	50 × 45	260

On account of the mapping from forward distance to the vertical coordinate, we don't need to slide various size of windows in the whole image. According to the distance prior, we can use multiple scale sliding windows in different vertical coordinates on the image. The size of sliding window can be determined by statistics. During the statistical process of window size, we discover that not only the scale but also the aspect ratio of forward vehicles will change as the distance varies. When the vehicle is far, its scale is small and the aspect ratio will be approximate to 1:1. However, the aspect ratio of vehicles will change into nearly 1.5:1 when they are close to the camera, e.g. 5 meters.

The reason for the change of aspect ratio is the extension distortion caused by wide-angle camera. Since drive recorders always utilize the wide-angle camera, the change of aspect ratio does exist in practice. Although extension distortion can be calibrated and corrected, we don't calibrate it in practice. Calibrating more parameters will make calibration course more difficult to setup. However, when measuring the distance of a forward vehicle, extension distortion will not affect this course. For the reasons above, extension distortion is ignored in the calibration course.

According to the distance prior, our multi-scale detection could search vehicles in different distance with a certain scale. The main advantages of our multi-scale detection are as follows: First, we relieve the multi-scale problem in forward vehicle detection. Then, due to sliding window with a certain scale in different locations in an image, the proposed method speeds up the detection course. Hence, our multi-scale detector can be faster than the original one and reach 50fps on CPU.

3.3. Abnormal driver behavior detection

In a real traffic scene, drivers do not always focus on the front. They will be affected by some abnormal behaviors such as fatigue, distraction, phone, and so on. Therefore, reaction time of the driver will keep changing. In this section, an abnormal driver behavior detection method is proposed to make the FCW system adaptive.

The most important thing in abnormal driver behavior detection is how to define abnormal behaviors. An in-vehicle camera shoots the driver's face continuously. The inappropriate definition will lead to unsatisfying false warnings. As a result, only those activities that affect driving safety is defined as abnormal behaviors. Specifically, the following behaviors are considered as abnormal: yawn, sleep, phone, head down, glance right and left.

Table 2: The architecture of FDSOD

Layer		FDSOD
Stem	Convolution	3×3 *conv*, *stride* 2
	Convolution	3×3 *conv*, *stride* 1
	Convolution	3×3 *conv*, *stride* 1
	Pooling	2×2 *max pool*, *stride* 2
Dense Block (1)		$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 6$
Transition Layer		1×1 *conv*
		2×2 *max pool*, *stride* 2
Dense Block (2)		$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 8$
Transition w/o Pooling Layer		1×1 *conv*

Different facial states is applied to judge whether the driver is abnormal or not. However, when detecting abnormal behaviors, face components detection methods will not be utilized. Due to the diversity of face states, components based methods, such as blinking detection, cannot cover all these abnormal behaviors. Approaches that monitor the whole face state are supposed to be used. On the other hand, the hand-crafted features such as HOG, Haar-like do work in face detection, however, they will fail in abnormal detection with facial states. Traditional features are not sufficient to distinguish changes in facial states. Therefore, a deep learning based method is proposed in this section to detect abnormal behaviors.

We propose an improved version of DSOD [44] that is called FDSOD (Fast Deeply Supervised Object Detectors) to detect abnormal behaviors. The proposed method is a multi-scale proposal-free detection framework. The network structure can be divided into two parts: the backbone sub-network for feature extraction and the front-end sub-network for prediction over multi-scale response maps. The backbone sub-network is composed of a stem block, two dense blocks, one transition layers and one transition w/o pooling layer. The front-end sub-network for prediction which is the same as DSOD fuses multi-scale prediction responses with an elabrated dense structure.

Table 2 shows the architecture of FDSOD. The stem block which consists of three 3×3 convolution layers followed by a 2×2 max pooling layer can reduce the information loss from raw input images. The reward of stem block have proven to be significant for detection performance in [44]. Each transition layer contains a pooling operation to down-sample the feature maps. A dense block connects all preceding layers to the current layer. This structure can reduce the number of parameters. Fewer parameters make the model faster. Due to dense connection, information loss among layers is also decreased. In original design of DenseNet [46] the number of dense blocks is fixed. However, introducing transition w/o pooling layer [44] eliminates this restriction of the number of dense blocks. According to this design thought, we propose the FDSOD.

Compared with DSOD, the proposed structure remove two dense blocks, one transition layer and one transition w/o pooling layer. Though we remove some layers of the network, the performance of FDSOD doesn't decrease too much. The main reason that some layers are removed is the speed limitation of DSOD. As we all know, the FCW systems need to handle various situations in real time. Thence, abnormal driver behavior detection is also supposed to be real time. Besides removing these layers, semi-precision optimization is applied for Caffe layer to reduce computing time. The proposed FDSOD can reach 25fps in NVIDIA TX1 while DSOD is 8fps under the same conditions. For the same reason, some recent deep learning techniques such as [23, 27] will not be applied. These networks can improve the performance of driver behavior detection, but slow down the whole FCW system.

Since the driver needs some time to recover from an abnormal state, we test people's responses time to alerts in different abnormal behaviors. The results showed that two seconds are enough for the driver to respond to the warning information. Therefore the warning of FCW will be start two seconds in advance when abnormal behaviors are detected.

4. Experiment

To demonstrate the capabilities of the presented adaptive FCW framework, extensive experiments are conducted and evaluated. In this section, we will introduce the experiment from the following three aspects: camera calibration, multi-scale vehicle detection and abnormal driver behavior detection.

Table 3: Experimental results of camera calibration

Car No.	d / m	d' / m	$e*$ / m	e_r / %
1	5.00	5.00	0.00	0.00
2	7.00	6.98	0.02	0.29
3	9.00	9.08	0.08	0.89
4	11.00	11.11	0.11	1.00
5	13.00	13.23	0.23	1.77
6	15.00	15.26	0.26	1.73
7	17.00	17.31	0.31	1.82

4.1. Validation of camera calibration

In the proposed adaptive FCW framework, the in-vehicle camera which detects abnormal driver behaviors does not need to be calibrated. Only the camera that measures distance requires to be calibrated.

In our experiment, the images comes from the camera of ordinary driving recorder and its size is 1280×720. The height of the camera is 122.5cm, and three fixed points used for the calibration are 4m, 5m and 7m away from the camera. Their vertical coordinate are 461, 428, 383. Following the calibration steps mentioned in Section 3.1, we obtain camera parameters for measuring distance. The calibration results are $\alpha = 0.1194rad, f_y = 1094.313$ and $v_0 = 363.331$. Then a set of test cars are substituted into the algorithm to detect its measurement error. The estimated distance is denoted by d'. The absolute error and relative error can be expressed separately as $e^* = |d - d'|$ and $e_r = e^*/d$. The measuring results are demonstrated in Table 3.

As illustrated in Table 3, this algorithm performs well when the points are near, and relative errors increase with the distance becomes far. This is because that along with the object getting farther, one pixel on the image covers longer distance. In other words, if one pixel represents several centimeters in the near, it may represent several meters in the far distance. It is an inherent defect of monocular vision.

4.2. Validation of multi-scale vehicle detection

In order to illustrate the performance of the proposed detection method, comparisons are made between our detector and [47, 28, 48, 36]. All of these methods are trained by KITTI car detection dataset [49] and tested on 5400 images of real scene collected by ourselves. The test images come from 30 different driving videos taken by the same recorder. These videos cover urban road, highway, night, rainy and other situations. Each video is 3 minutes with 30 fps, and test images are selected every one second. Considering the limitation of computing resource in the practical application, deep learning methods will not be compared in this section.

Table 4 shows the comparisons of detection rate and FPPI (false positive per image). Benefiting from distance prior, our detector has the knowledge of vehicle size in different vertical coordinates. FPPI decreases obviously, which means less false detection occurs during our framework. Because we have certain scales in different vertical coordinates, our detector performs better. Besides,

Table 4: Comparison of various detection methods

	Detection rate	FPPI	Time(s)/frame
DPM [47]	91.23%	0.098	4.0
And-Or [28]	89.08 %	0.133	3.0
SubCat [48]	92.70 %	0.087	0.7
ACF [36]	94.02 %	0.065	0.04
Ours	96.61%	0.046	0.02

Figure 4: Overview of the data set.

certain scales also decrease the number of sliding windows. It also makes the proposed detector faster than others.

Our multi-scale detection framework can achieve 50 frames per second on Intel i5 quad core CPU with 3.20GHz, which can meet the requirements of other automatic driving and assistance driving applications besides forward vehicle collision warning in the future.

Table 5: Experimental results of abnormal behavior detection

Method	fps	mAP	Down	Left	Right	Sleep	Phone	Yawn	Normal
DSOD [44]	8	95.6%	99.0%	98.1%	98.7%	96.3%	94.5%	88.3%	94.6%
Ours	25	91.6%	96.5%	95.7%	95.1%	92.2%	90.2%	82.6%	88.7%
Baseline [50]	3	83.1%	88.2%	86.4%	86.1%	83.7%	81.8%	74.9%	80.3%

4.3. Validation of abnormal driver behavior detection

We build our own data set to evaluate the performance of abnormal driver behavior detection. The data set has 17500 images and six facial states: *yawn*, *sleep*, *phone*, *head down*, *glance right*, *glance left* and *normal*. Each facial state has 2500 samples. We use 2000 samples to be train set and 500 samples to be test set. Since some behaviors, such as phone and sleep, are dangerous in real driving situations, we obtain such dangerous behaviors when parking. Figure 4 illustrates some examples of the data set. Images in our data set cover different illumination, daytime and night. After arranging, the data set will be released soon.

From a practical point of view, abnormal driver detection is performed on NVIDIA TX1 with 56 core Pascal GPU. Table 5 shows the experiment results. In the experiment, we regard Faster RCNN [50] as the baseline. After pruning the network structure of DSOD, the performance of the proposed FDSOD doesn't decrease a lot. However, the processing time of our method is nearly three times than DSOD. And the speed of 25fps is also achieve the real-time requirements.

5. Conclusion

In conclusion, we propose a vision based adaptive forward vehicle collision warning framework. Easy and efficient calibration method makes our framework convenient to build. Multi-scale detector improves detection accuracy and decrease time consumption. Innovatively introducing abnormal driver behaviors detection via an in-vehicle camera makes FCW to generate warnings adaptively. The entire FCW framework can run in real time, which makes our work highly practical.

6. Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant U1864204 and 61773316, State Key Program of National Natural Science Foundation of China under Grant 61632018, and Project of Special Zone for National Defense Science and Technology Innovation.

References

[1] N. C. for Statistics, Analysis, Summary of motor vehicle crashes : 2015 data.

[2] G. R. Widmann, W. A. Bauson, S. W. All, Development of collision avoidance systems at delphi automotive systems, in: IEEE International Conference on Intelligent Vehicles, Vol. 2, 1998.

[3] Y. Yuan, J. Fang, Q. Wang, Incrementally perceiving hazards in driving, Neurocomputing 282 (2018) 202 – 217.

[4] S. K. Kenue, Selection of range and azimuth angle parameters for a forward looking collision warning radar sensor, in: Proceedings of the Intelligent Vehicles '95 Symposium, 1995, pp. 494–499.

[5] N. Srinivasa, Y. Chen, C. Daniell, A fusion system for real-time forward collision warning in automobiles, in: Proceedings of the IEEE International Conference on Intelligent Transportation Systems, Vol. 1, 2003, pp. 457–462.

[6] J. Wang, C. Yu, S. Li, L. Wang, A forward collision warning algorithm with adaptation to driver behaviors, IEEE Transactions on Intelligent Transportation Systems 17 (4) (2016) 1157–1167.

[7] R. J. Kiefer, M. T. Cassar, C. A. Flannagan, D. J. Leblanc, M. D. Palmer, R. K. Deering, M. A. Shulman, Forward collision warning requirements project: Refining the camp crash alert timing approach by examining "last-second" braking and lane change maneuvers under various kinematic conditions.

[8] S. M. Iranmanesh, E. Moradi-Pari, Y. P. Fallah, S. Das, M. Rizwan, Robustness of cooperative forward collision warning systems to communication uncertainty, in: Systems Conference, 2016, pp. 1–7.

[9] J. Wang, C. Yu, S. E. Li, L. Wang, A forward collision warning algorithm with adaptation to driver behaviors, IEEE Transactions on Intelligent Transportation Systems 17 (4) (2016) 1157–1167.

[10] L. Zhang, L. Zhang, D. Zhang, K. Li, An adaptive longitudinal driving assistance system based on driver characteristics, IEEE Transactions on Intelligent Transportation Systems 14 (1) (2013) 1–12.

[11] F. Muehlfeld, I. Doric, R. Ertlmeier, T. Brandmeier, Statistical behavior modeling for driver-adaptive precrash systems, IEEE Transactions on Intelligent Transportation Systems 14 (4) (2013) 1764–1772.

[12] A. Sathyanarayana, P. Boyraz, J. H. L. Hansen, Driver behavior analysis and route recognition by hidden markov models, in: IEEE International Conference on Vehicular Electronics and Safety, 2015, pp. 276–281.

[13] S. Sekizawa, S. Inagaki, T. Suzuki, S. Hayakawa, N. Tsuchida, T. Tsuda, H. Fujinami, Modeling and recognition of driving behavior based on stochastic switched arx model, IEEE Transactions on Intelligent Transportation Systems 8 (4) (2007) 593–606.

[14] N. C. for Statistics, Analysis, Alcohol impaired driving: 2016 data.

[15] Y. Lu, Y. Yuan, Q. Wang, Forward vehicle collision warning based on quick camera calibration, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2018, pp. 2586–2590.

[16] P. K. Ghosh, S. P. Mudur, Three-dimensional computer vision: A geometric viewpoint, The Computer Journal 12 (1) (1995) 475–475.

[17] R. Tsai, A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses, Robotics and Automation IEEE Journal of 3 (4) (1987) 323–344.

[18] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (11) (2000) 1330–1334.

[19] P. F. Sturm, S. J. Maybank, On plane-based camera calibration: A general algorithm, singularities, applications, in: IEEE Conference on Computer Vision and Pattern Recognition, 1999, pp. 1432–1437.

[20] Z. Zhang, Camera calibration with one-dimensional objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (7) (2004) 892–9.

[21] M. Pollefeys, Multiple view geometry, Encyclopedia of Biometrics 2 (9-10) (2000) 181–186.

[22] W. Wang, J. Shen, H. Ling, A deep network solution for attention and aesthetics aware photo cropping, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (7) (2019) 1531–1544.

[23] W. Wang, J. Shen, Deep visual attention prediction, IEEE Transactions on Image Processing 27 (5) (2018) 2368–2378.

[24] X. Dong, J. Shen, D. Wu, K. Guo, X. Jin, F. Porikli, Quadruplet network with one-shot learning for fast visual object tracking, IEEE Transactions on Image Processing 28 (7) (2019) 3516–3527.

[25] W. Wang, J. Shen, L. Shao, Video salient object detection via fully convolutional networks, IEEE Transactions on Image Processing 27 (1) (2018) 38–49.

[26] X. Dong, J. Shen, W. Wang, Y. Liu, L. Shao, F. Porikli, Hyperparameter optimization for tracking with continuous deep q-learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 518–527.

[27] X. Dong, J. Shen, Triplet loss in siamese network for object tracking, in: European Conference on Computer Vision, 2018, pp. 472–488.

[28] B. Li, T. Wu, S. Zhu, Integrating context and occlusion for car detection by hierarchical and-or model, in: European Conference on Computer Vision, 2014, pp. 652–667.

[29] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (9) (2010) 1627–1645.

[30] L. C. Leon, R. H. Jr, Vehicle detection using mixture of deformable parts models: Static and dynamic camera, in: Sibgrapi Conference on Graphics, Patterns and Images, 2012, pp. 237–244.

[31] C. Wang, Y. Fang, H. Zhao, C. Guo, Probabilistic inference for occluded and multiview on-road vehicle detection, IEEE Transactions on Intelligent Transportation Systems 17 (1) (2016) 215–229.

[32] J. Wu, X. Zhang, A pca classifier and its application in vehicle detection, in: International Joint Conference on Neural Networks, 2001, pp. 600–604.

[33] X. Li, X. Yao, Y. L. Murphey, R. Karlsen, G. Gerhart, A real-time vehicle detection and tracking system in outdoor traffic scenes, in: International Conference on Pattern Recognition, 2004, pp. 761–764.

[34] A. Khammari, F. Nashashibi, Y. Abramson, C. Laurgeau, Vehicle detection combining gradient analysis and adaboost classification, in: IEEE Transactions on Intelligent Transportation Systems, 2005, pp. 66–71.

[35] W. Zheng, L. Liang, Fast car detection using image strip features, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2703–2710.

[36] P. Dollár, R. Appel, S. J. Belongie, P. Perona, Fast feature pyramids for object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (8) (2014) 1532–1545.

[37] Y. Yuan, Z. Xiong, Q. Wang, An incremental framework for video-based traffic sign detection, tracking, and recognition, IEEE Transactions on Intelligent Transportation Systems 18 (7) (2017) 1918–1929.

[38] M. Nakaoka, P. Raksincharoensak, M. Nagai, Study on forward collision warning system adapted to driver characteristics and road environment, in: International Conference on Control, Automation and Systems, 2008, pp. 2890–2895.

[39] C. Y. Chang, Y. R. Chou, Development of fuzzy-based bus rear-end collision warning threshols using a driving simulator, IEEE Transactions on Intelligent Transportation Systems 10 (2) (2009) 360–365.

[40] D. J. Simons, Examining cognitive interference and adaptive safety behaviours in tactical vehicle control, Ergonomics 50 (8) (2007) 1340–1350.

[41] B. Donmez, L. N. Boyle, J. D. Lee, Safety implications of providing real-time feedback to distracted drivers., Accident Analysis and Prevention 39 (3) (2007) 581–590.

[42] S. M. Iranmanesh, H. N. Mahjoub, H. Kazemi, Y. P. Fallah, An adaptive forward collision warning framework design based on driver distraction, IEEE Transactions on Intelligent Transportation Systems PP (99) (2018) 1–10.

[43] C. Lin, F. Su, H. Wang, J. Gao, A camera calibration method for obstacle distance measurement based on monocular vision, in: International Conference on Communication Systems and Network Technologies, 2014, pp. 1148–1151.

[44] Z. Shen, Z. Liu, J. Li, Y. Jiang, Y. Chen, X. Xue, DSOD: learning deeply supervised object detectors from scratch, in: IEEE International Conference on Computer Vision, 2017, pp. 1937–1945.

[45] R. Wang, L. I. Bin, J. Chu, J. I. Shouwen, Study on the method of measuring the leading vehicle distance based on the on-board monocular camera, Journal of Highway and Transportation Reseach Andk Development.

[46] G. Huang, Z. Liu, L. V. D. Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2261–2269.

[47] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (9) (2010) 1627–1645.

[48] E. Ohn-Bar, M. M. Trivedi, Learning to detect vehicles by clustering appearance patterns, IEEE Intelligent Transportation Systems 16 (5) (2015) 2511–2521.

[49] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[50] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) (2017) 1137–1149.