FORWARD VEHICLE COLLISION WARNING BASED ON QUICK CAMERA CALIBRATION

Yuwei Lu[1]   *Yuan Yuan*[1]   *Qi Wang*[1,2*]

[1]School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL),
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
[2]Unmanned System Research Institute (USRI),
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China

ABSTRACT

Forward Vehicle Collision Warning (FCW) is one of the most important functions for autonomous vehicles. 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 priori is proposed to improve the precision. Intensive experiments are conducted in our established real scene dataset and the results have demonstrated the effectiveness of the proposed framework.

***Index Terms*—** Forward vehicle collision warning, vehicle detection, distance measurement, camera calibration

1. INTRODUCTION

Significant effort has been made on the safety of road vehicles in recent decades. Over 10 million people are injured yearly worldwide in road accidents. Automatic driving and assistance driving are the research directions to increase the safety of passengers and of vehicles. Forward vehicle collision warning is the fundamental function in both automatic driving and assistance driving.

A range of devices mounted on the vehicle could provide the solution to this problem [1, 2]. The traditional systems available today are typically based on radar sensors [3]. 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 is an attractive approach. In such a system [4, 5] the radar provides accurate distance and velocity, while vision gets exact locations of forward vehicles. However, this solution is expensive and complex in practical applications.

Considering 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 obtains distance from forward vehicles to the camera, while the vehicle detection stage based on the distance gets exact locations of forward vehicles. 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. Second, distance information is utilized to improve the performance of detecting forward vehicles.

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

2.1. Camera Calibration

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

object-based calibration. Techniques in this category are required to observe a calibration object [6, 7, 8, 9, 10]. Some of these methods [6, 7] require the geometry information of object in 3D space with very good precision. These approaches always need an expensive calibration apparatus, and a complex setup. Therefore, some methods calibrate camera by observing a planar pattern shown at a few different orientations [8, 9]. Since such a calibration pattern is easy to be made, the setup becomes easier. In order to make calibration easier, Zhang [10] proposes one-dimensional object based calibration. It uses less knowledge of the observation

compared to 2D and 3D object-based calibration methods.

Self-calibration. Techniques in this category don't utilize any calibration objects [11]. 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

Considering the practical applications, deep learning methods will not be reviewed in this section. Vehicle detection approaches are mainly 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.* [12] propose an And-Or model that integrates context and occlusion for detecting vehicles. Felzenszwalb *et al.* [13] propose deformable part models to structure template model. Each model is composed of parts with different viewpoints. They detect vehicles by integrating various parts of vehicles. Wang *et al.* [14] 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. Zheng *et al.* [15] 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.* [16] propose aggregate channel features (ACF) and Yuan *et al.* [17] improve the features for detection. ACF utilizes color information of the objects to improve the performance.

3. OUR METHOD

As mentioned before, the proposed framework includes two stages: camera calibration and vehicle detection. To simplify the calibration course, a point-based calibration method [18] is employed to get camera parameters and to calculate distance from the forward car. During the detection course, we expand original ACF detector [16] into a distance-based multiple scale detector. The distance is not only used for forward collision warning, but also employed for improving vehicle detection.
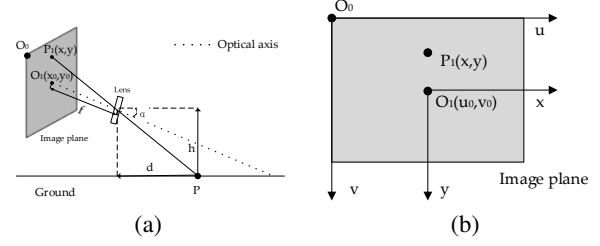

(a)   (b)

Fig. 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 cameras pixel location coordinate system uO_0v.

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 [19], the distance from point P to camera is:

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

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:

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

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

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

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. Only f_y, v_0, α need to be calibrated. Based on Eq. 3,
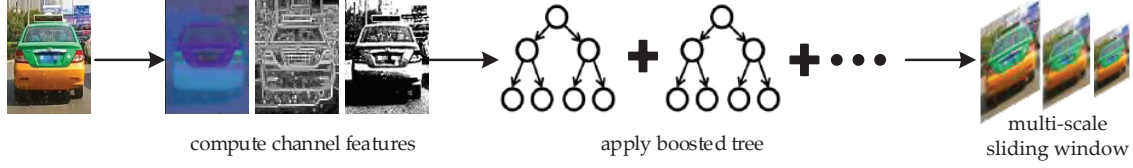
Fig. 2. Overview of our detection framework.

three fixed points are utilized to estimate camera parameters in [18]. With the input of height h, distances information of three calibration points (d_1, d_2, d_3) and corresponding coordinates $(u_1, v_1), (u_2, v_2), (u_3, v_3), f_y, v_0, \alpha$ will be solved.

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 [16] 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:

$$v = v_0 - f_y \tan(\arctan \frac{h}{d} - \alpha). \qquad (4)$$

Eq. 4 is the foundation of multi-scale detection with distance prior. It indicates that if v_0, f_y and α are estimated, the verti-

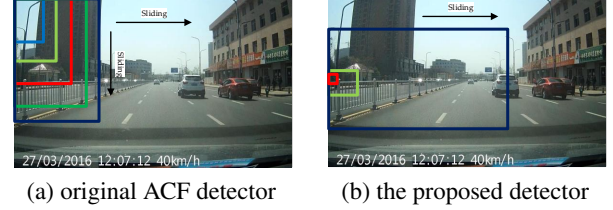(a) original ACF detector   (b) the proposed detector

Fig. 3. The difference between original ACF detector and our detector.

cal 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. 4. 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 detection can be faster than the original one and reach 50fps on CPU.

4. EXPERIMENT

To demonstrate the capabilities of the presented approach, extensive experiments are conducted and evaluated in an Intel i5 quad core CPU with 3.20GHz. In this section, we will introduce the experiment from the following two aspects: calibration and detection.

Table 2. 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	15.00	15.26	0.26	1.73
6	17.00	17.31	0.31	1.82

4.1. Validation of calibration method

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 2.

As illustrated in Table 2, 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 distance based detection

In order to illustrate the performance of the proposed detection method, comparisons are made between our detector and [20, 12, 21, 16]. All of these methods are trained by KITTI car detection dataset [22] 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 3 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, 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 CPU, which can meet the requirements of other automatic driving and assistance driving applications besides forward vehicle collision warning in the future.

Table 3. Comparison of various detection methods

	Detection rate	FPPI	Time(s)/frame
DPM [20]	91.23%	0.098	4.0
And-Or [12]	89.08 %	0.133	3.0
SubCat [21]	92.70 %	0.087	0.7
ACF [16]	94.02 %	0.065	0.04
Ours	96.61%	0.046	0.02

5. CONCLUSION

In conclusion, we propose a simple but efficient forward vehicle collision warning framework by aggregating monocular distance measurement and precise vehicle detection. Point based calibrating algorithm greatly reduces the computational complexity and can be easily achieved. Multi-scale detection has taken excellent advantage of distance prior to improves accuracy and decrease time consumption. The processing speed of the proposed framework can achieve 50 fps on CPU and the experiments have exhibited its outstanding performance.

6. REFERENCES

[1] Glenn R. Widmann, William A. Bauson, and Steven W. All, "Development of collision avoidance systems at delphi automotive systems," in *IEEE International Conference on Intelligent Vehicles*, 1998, vol. 2.

[2] Y. Yuan, J. Fang, and Q. Wang, "Incrementally perceiving hazards in driving," *Neurocomputing*, vol. 282, pp. 202 – 217, 2018.

[3] 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.

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

[5] J. Wang, C. Yu, S. Li, and L. Wang, "A forward collision warning algorithm with adaptation to driver behaviors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1157–1167, 2016.

[6] P. K. Ghosh and S. P. Mudur, "Three-dimensional computer vision: A geometric viewpoint," *The Computer Journal*, vol. 12, no. 1, pp. 475–475, 1995.

[7] 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*, vol. 3, no. 4, pp. 323–344, 1987.

[8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[9] Peter F. Sturm and Stephen 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.

[10] Z. Zhang, "Camera calibration with one-dimensional objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 892–9, 2004.

[11] Marc Pollefeys, "Multiple view geometry," *Encyclopedia of Biometrics*, vol. 2, no. 9-10, pp. 181–186, 2000.

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

[13] Pedro F. Felzenszwalb, Ross B. Girshick, David A. McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[14] C. Wang, Y. Fang, H. Zhao, and C. Guo, "Probabilistic inference for occluded and multiview on-road vehicle detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 215–229, 2016.

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

[16] Piotr Dollár, Ron Appel, Serge J. Belongie, and Pietro Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532–1545, 2014.

[17] Y. Yuan, Z. Xiong, and Q. Wang, "An incremental framework for video-based traffic sign detection, tracking, and recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1918–1929, 2017.

[18] C. Lin, F. Su, H. Wang, and 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.

[19] R. Wang, L. I. Bin, J. Chu, and 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*, 2001.

[20] Pedro F. Felzenszwalb, Ross B. Girshick, David A. McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[21] Eshed Ohn-Bar and Mohan Manubhai Trivedi, "Learning to detect vehicles by clustering appearance patterns," *IEEE Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2511–2521, 2015.

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