On-road Vehicle Detection Method based on Multi-scale Active Basis Model*

Yanjie Yao
The State Key Laboratory of Management and Control for Complex Systems
Beijing Engineering Research Center of Intelligent Systems and Technology
Institute of Automation, Chinese Academy of Sciences
Beijing, China

Gang Xiong
Dongguan Research Institute of CASIA
Cloud Computing Center, Chinese Academy of Sciences
Dongguan, China
The State Key Laboratory of Management and Control for Complex Systems
Beijing, China

***Abstract*—This paper introduces a vehicle detection method based on multi-scale active basis model in traffic surveillance systems. Due to the effects of perspective, vehicles which are close to the camera are larger and more detailed than the far ones on individual video frames. Using camera calibration, we get the multi-scale information of vehicles, and then we learn the multi-scale active basis model from the target training sample set by using the shared sketch algorithm. The multi-scale ABM can detect vehicles in various poses, shapes, and sizes in a whole video frame. The experiment results show that this proposed method can fit the changes of vehicle size in images.**

Keywords—ABM; multiscale; camera calibration; vehicle detection

I. INTRODUCTION

According to a report from Ward's Auto released in Aug. 15, 2011, the global number of vehicles surpassed 1 billion-unit mark in 2010, jumping from almost one thousand million the year before[1]. With the increasing number of vehicles，problems on traffic congestion and traffic safety pose great challenges on traffic management systems in most large and medium-sized cities. And at the same time, a great scale of traffic data set is emerged in the complex traffic systems[2], becoming the basis to model a traffic scene and analyze the traffic behavior. So how to get the on-road traffic data exactly is crucial in traffic information collection systems and the data-driven intelligent transportation systems (ITS).

Compared with traditional traffic information collection technology, such as geography induction coils, millimiter-wave radars, laser detector and so on, vision based devices have broadly been employed in traffic monitoring of ITS and has many advantages[3]: information to be understood easily, a wide range of information to be installed, operated, and maintained; a relatively higher price-to-performance ratio to be obtained. However, as a matter of fact, several factors make on-road vehicle detection using vision based devices very challenging[4][5]. Firstly, complex outdoor environment increases the difficulty in designing the vehicle detection and identification systems, such as light variations in different time of the day and changes in the weather. Then, changes in a vehicle's appearance is also a trouble in vehicle detection, that vehicles come into view with different speeds and may vary in shape, size, color, and pose. Also, there isn't a satisfactory solution to vehicle detection problems such as vehicle occlusion and moving shadow.

Various vehicle detection approaches have been reported in the traffic monitoring systems. A lot of researchers do vehicle detection using relative motion information. Optical-flow and background subtraction are the main methods to do motion-based localization in a continuous image sequence [6][7]. Unfortunately, when a car turns into stationary from moving or is parked for some time, these approaches are ineffective. Instead of the motion-based methods, many researchers have reported their progress in feather-base methods. They make use of vehicles' characteristic information such as color, texture, shadow, vehicle lights, corners and edges to do the detection and recognization.In[8], vehicle features (vehicle colors and local features) were extracted, and a dynamic Bayesian network was constructed to classify vehicles from aerial surveillance images. Wu et al.[9] use active basis model (ABM) which consists of a small number of Gabor wavelet elements at selected locations and orientations to cope with object detection problems. And currently, Li et al.[10] realized a vehicle detection method based on graphic structure on the basis of ABM.

A shared sketch algorithm in ABM was proposed by Wu et al. in [9]. The basic idea is to apply a gray-value local energy to find a common template together with its deformed versions from the training images. However, this algorithm is used only when objects appear with the same pose and at the same location under the same scale in the training images. But due to the effects of perspective, vehicles will represent a variety of scales in different locations, i.e. those which are close to the camera are larger and more detailed than the far ones on individual video frames. As a result, it will be hard to accurately detect a situation where objects will appear at different scales and unknown locations in an image.

To overcome the problem, we add the multi-scale information into the learned ABM from the target training sample set. In this paper, as a contribution, we take advantage of the camera calibration technology to learn the multi-scale active basis model by using the shared sketch algorithm. Because of the special characters of the traffic surveillance images, we use the camera model in [17] and [18] to do the

This work was supported in part by the National Natural Science Foundation of China under Grant 71232006, Grant 61233001, and Grant 61174172.

camera calibration task. The framework in [9] is adopted to do the learning, detection, and classification. Then, the template matching algorithm is used to recognize vehicles in different locations and at different scales. We have done a series of experiments to verify the performance of our method, and experimental results prove that our method can effectively detect vehicles in a variety of scales at different locations.

This rest of this paper is naturally organized as follows: Section II introduces some basic theories of the camera calibration, multi-scale information and active basis model (ABM). Section III presents the implementation of our method in detail. Section Ⅳ demonstrates and analyzes the results in experiments. Finally, Section Ⅴ concludes the paper.

II. BACKGROUND

Before introducing our method, some basic knowledge and background in this paper should be known in advance.

A. Camera Calibration

Camera imaging can be viewed as a perspective projection process. The relationship between the 3D geometrical position of a point in real-world coordinates and its corresponding point in the image is established by the camera model. And the camera parameters refer to the parameters of the geometric model, including intrinsic and extrinsic parameters. According to the extrinsic parameters, the correspondence relationship between the world coordinate system and the camera coordinate system can be obtained; adding the intrinsic parameters, the relationship between a 3D point and its image projection can also be obtained; so the 2D images can be mapped to the 3D world coordinate by exploiting camera parameters.

Camera calibration which is the first step in 3D computer vision in order to extract metric information from 2D images can produce a map from image pixel position to real-world coordinates. Many researchers have developed many well-known calibration methods. Usually, there are two types of techniques to calibrate a camera: model-based and priori-knowledge based. Model-based camera calibration technique is difficult to apply geometric methods in traffic scene [11] [12] because of the reference marks with known physical dimension are not readily available in most cases. In contrast, priori-knowledge based calibration methods make use of the prior information of parameters (lane width, typical vehicle width and vanishing points et al) avoiding the dependence of availability of suitable reference marks in images.

The pin-hole model is the simplest and most commonly used model to calibrate a camera which makes use of a homography matrix H to represent the relationship between the pixel coordinates and world coordinates, shown in (1):

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where s is an arbitrary scale factor, (X, Y, Z) is 3D point, and (u, v) is its image projection. H is a 3×4 projection matrix and contains the intrinsic parameters and extrinsic parameters of a camera, and can be defined as:

$$H = k[R|t] = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{2}$$

where k is the intrinsic parameters matrix while R and T are the extrinsic parameters matrix representing movements of rotation and translation, respectively, with (c_u , c_v) the coordinates of the principal point, f_u, f_v the focal length of the pixels.

Zhang's Camera Calibration Toolbox for Matlab [11][12] [13]is widely used to get the camera parameters, but it isn't quite suitable for the traffic monitoring scene because of a lack of so many corners in a flat road. In this paper, the lane markings are to do the calibration, and you can look for the Section Ⅲ B for detail information.

B. Multiscale Information

According to the information above, we can get the idea that the representations of vehicles differ in different scales in a 2D image, as a result of the perspective projection effect. For the images acquired by the video camera, multi-scale analysis of the images is similar to observing an object from different distances[14]: we can only see the outline of the distant one (large-scale) while we can get more details about the close one (small-scale), i.e. the representations of the closely located vehicle occupies a relatively large number of pixels, while some distant vehicles' representations are much smaller on the image plane. Information in different scales has different advantages and disadvantages. When in small-scale, there is a plenty of edge information in details, and it is easy to locate the edge, but the sensitivity to noise can't be ignored; when in large-scale, edges for objects detection are more stable, and own immunity to noise, but the positioning accuracy is poor.

By the camera calibration technology, the intrinsic parameters and the extrinsic parameters of a camera can be obtained. Thus, we can calculate continuous scales change of the learned model in different location of the input image by comparing the changes in the arbitrary scale factor between the real 3D points and their 2D correspondence points, and the way to calculate the changes ration in length is explained in Section Ⅲ B. As a result, there is a choice about multi-scale deformable templates base on different scales information.

C. Active Basis Model

With the following two characteristics: (1) Only requiring a small amount of training samples which are approximately aligned and consistent posture; (2) For different samples, the wavelets of ABM can make adaptive changes within a certain range for a better match to the local edge of the target, the Active Basis Model (ABM) is an excellent target expression model.

The ABM is a deformable model that learns the shapes of object categories by a shared sketch algorithm. In particular, the active basis refers to the locations and the orientations of the basis Gabor wavelet elements. According to the sparse coding theory [15], the image I can be represented by a linear

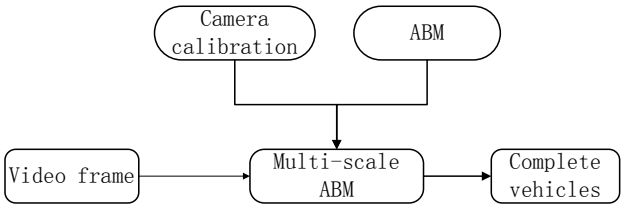


Fig.1. Block-diagram view of the proposed algorithm

combination of several Gabor wavelet elements which belong to an over-complete dictionary $\{G_{x,y,s,\theta}, (x,y), \theta \in \{i\pi/N, i = 0, \ldots, N-1\}\} \in D$, as follows:

$$I = \sum_{i=1}^{n} \alpha_i G_i + \gamma \quad (3)$$

Where n is the number of the selected elements, $G_i = G_{x_i,y_i,s,\theta_i}$ is the Gabor wavelet element to segment the edge in the domain D of image patch I, α_i corresponds to the coefficient of wavelet element, and γ is the unexplained residual image.

According to the theory of Wu et al, the Gabor wavelet element G_i can be transformed by dilation, rotation, and changing the aspect ratio to learn a deformable template. A sum-max maps computational architecture is used to match and recognize the deformable template from an image.

Before object detection, an ABM needs to be learned from a small set of training images. During the learning process, several Gabor wavelet elements are extracted to represent the vehicle object by the use of the shared sketch learning algorithm. For the detailed description of the shared sketch learning algorithm, you can look for [9] as reference.

III. METHOD

Our method includes three parts: (1) constructing an ABM to represent the on-road vehicle object category; (2) obtaining camera parameters by camera calibration algorithm; (3) generating multi-scale ABM for edge detection and candidates localization. Fig.1 shows the block-diagram illustrating a high level overview of the proposed method in this paper. In the following, we will introduce every part in sequence.

A. Learning ABM

At the beginning to learn the ABM, we prepare our training set. The images in training set are collected from real traffic scenes, which consist of 20 positive samples with front-view vehicles and negative samples. In this paper, we haven't considered buses model for its large deformation comparing to other vehicle types. The parameters and location of Gabor wavelets which make up an ABM can be learned by the shared sketch learning algorithm. We get an deformable model $T = \{G_i, x_i\}$,i=1,…,N. from the training images, where G_i is the i_{th} Gabor wavelet element , and x_i is the corresponding parameters. The core idea of the shared sketch learning algorithm is to select all the wavelets which share their locations and orientations in the sets. In this paper, the ABM consists of 30 Gabor wavelets.

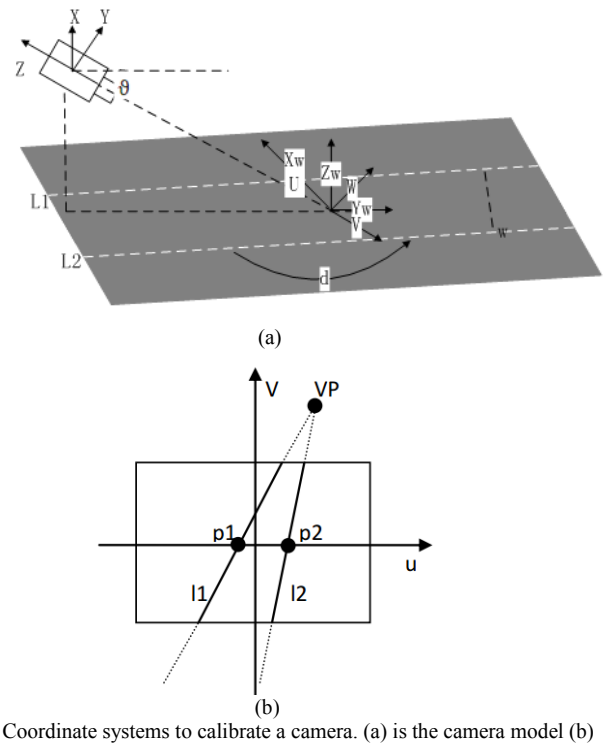
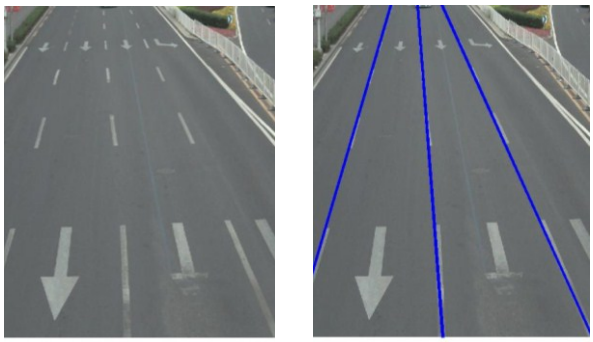


Fig.2. Coordinate systems to calibrate a camera. (a) is the camera model (b) is the road geometry in the image with the vanishing point from parallel markings.

B. Obtaining Camera Parameters by Camera Calibration Algorithm in Traffic Image

Considering the character in traffic scene, camera calibration technology usually utilizes lane markings, vehicle outlines, traffic flow and other objects on the road to accomplish the calibration system by manual approach [16]. It is easy to see the characters of these markings for calibration are parallel or perpendicular. In this paper, we refer the calibration model in [17] and [18], the parallel lane-marking lines to compute the parameters.

As we know, a pair of parallel lines will pass through the same vanishing point in an image because of the perspective projection effect. According to the work in [17] and [18], three coordinates will be used for calibration: world coordinate system, the camera coordinate and the camera shift coordinate, which denotes as O-$X_wY_wZ_w$, O-XYZ and O-UVW, respectively . Their geometry relationship can be seen in Fig.2, and (a) is the camera model, where $L_1$and L_2 are two parallel lane markings, w is the lane width between the parallel lanes, ϑ is the tilted angle that the camera installed above the ground plane, and θ is the pan angle between Y_w axis and the lane markings. And (b) is the road geometry in the image with the vanishing point from parallel markings, where P_1 and P_2 are intersections of lines L1 and L2 with the X_w axis, and $VP(u_0, v_0)$ is the vanishing point which is the real infinite far point's projection on the image plane. And the rectangle in Fig.2(b) represents the image plane. So, in this model, the camera parameters need to be calibrated can be written like the following with an acceptable error rate:



(a) (b)

Fig.3. The extracted background and lane markings. (a) the extracted background using GMM and(b) its calibration of parallel markings.

$$H = k[R|t] \approx f[\theta\ |\vartheta] \tag{4}$$

If a prior knowledge is given, namely, the width w between two parallel lane-markings, the coordinates p1 and p2, the length of a lane marking parallel d and its projection Δu, we can calculate the parameters f, θ and ϑ based on the following four equations:

$$u_0 = \lim_{Y\to\infty}\left(\frac{fX}{Y\cos\vartheta+D}\right) = f\tan\theta\cdot\sec\vartheta \tag{5-1}$$

$$v_0 = \lim_{Y\to\infty}\left(\frac{Y\sin\vartheta}{Y\cos\vartheta+D}\right) = f\tan\vartheta \tag{5-2}$$

$$\Delta u = \frac{fw\sec\theta}{D} \tag{5-3}$$

$$\sin 2\theta = \frac{2v_0 w}{d\Delta u}\left(\frac{u_i}{v_0-v_i}-\frac{u_j}{v_0-v_j}\right) \quad \text{if } v_i > v_j \tag{5-4}$$

where (u_i, v_i) and (u_j, v_j) are the endpoint of the known parallel, respectively. And (u_0, v_0) is the coordinate of vanishing point.

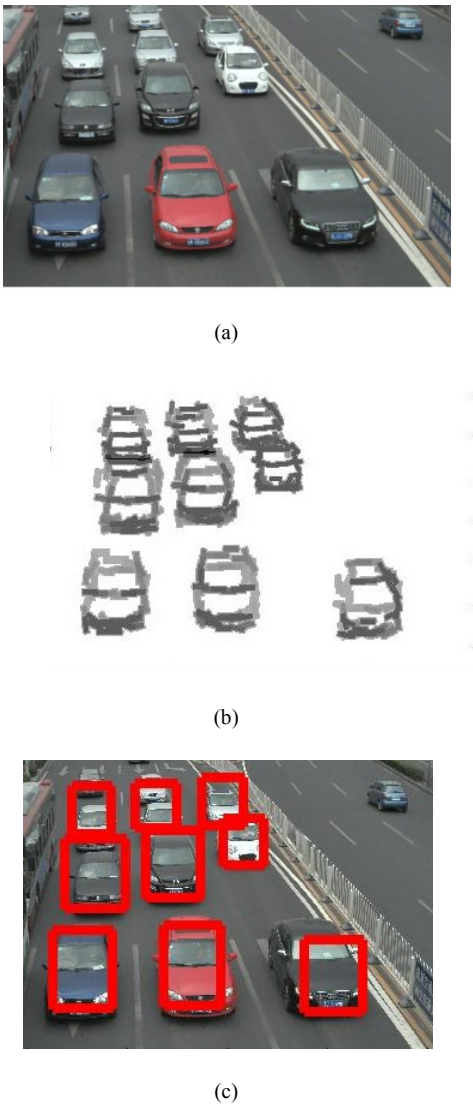
In this paper, the prior knowledge can be got by manual measuring. So here, w=3.35m, d=20m, and calculate the parameters of our camera.

Fig.3 gives an example of the detected land markings in a video image by background-subtraction using GMM. (a) shows the extracted background using GMM, and (b) is the example of its calibration of parallel markings.

Using the calibration results, we can easily get the relative changes between the world coordinate system and the camera shift coordinate, i.e. the length-changing ratio in the direction of Y axis. So we can change the size of our ABM to fit the vehicle's location in the video frame.

C. Edge Detection and Candidates Localization by Multi-scale ABM

In many research works, edge-based vehicle detection method is widely used, because it is often more effective than other background removal or threshold approaches for the reason that the edge information remains significant even in variations of ambient lighting. In this paper, we utilize Gabor wavelets to do the edge detection and candidates localization.



(a)

(b)

(c)

Fig.4. An example of the experiment result. (a) is the input image; (b) is the multi-scale ABM. result ; (c) is the detect result with bounding box.

For an image I_i, the matching condition between the model T and image I_i can be calculated as following:

$$P(I_i|T) = \log\left(\frac{p(I_i|T)}{q(I_i)}\right)$$

$$= \sum_{j=1}^{N}\left[x_j h\left(\left|\langle I_i, G_{i,j}\rangle\right|^2\right) - \log\varphi(x_j)\right] \tag{6}$$

where $T = \{G_i, x_i\}$, i=1,…,N, N is the number of Gabor wavelet elements of T, h(.) is Sigmoid transform.

We separate every single object detected in the ROI according to their locations obtained from the multi-scale Gabor filter. In the implementation, we don't need to change the size of the input images, but only change the scale of our ABM template according to the measurement of the image vehicle location. Finally, if the matching criteria are satisfied, we know that the candidate is a complete vehicle.

IV. EXPERIMENT RESULTS

In our experiment, the experiments have been done only in front-view vehicles which only include small size without considering heavy ones like buses or SUVs. To obtain length-changing ratio in the direction of Y axis, we use camera calibration technology to get the camera parameters f, θ and ϑ ,firstly. In this paper, the prior knowledge is w=3.35m, d=20m, and we also calculate the coordinates of P1,P2,and VP. And the parameters is $f = 550.1\text{pixels}$, $\theta = 109.3^{o}$ and $\vartheta = 38.7^{o}$ using the mean and stand deviation correction.

The size of images used in our experiment is 2592×1936 pixels, and we did not care about the changeable pose of each vehicle , such as the turning vehicles or the changing-lines ones, so here the vehicles are almost in the same pose aligned with parallel markings. And the ROI is set from 400 to 1900 pixels along the y-axis direction.

After calibration the camera, we get the length-changing ratio in the direction of Y axis by transforming the 3D points to 2D points. Then we can calculate the size changes of the vehicles. In the implementation, we don't need to change the size of the input images, but only change the scale of our ABM template according to the measurement of the image vehicle location. Fig.4 gives an example of the vehicles detected result using our multi-scale ABM. (a) is the input image; (b) is the detected result by multi-scale ABM; and (c) shows the last detection result labeled with bounding box. It illustrates that the multi-scale ABM can fit the changes of vehicle size in images.

V. CONCLUSION

In this paper, we propose an on-road vehicle detection method based on the multi-scale ABM. By taking advantage of the camera calibration technology, we learn the multi-scale active basis model by using the shared sketch algorithm. The improved ABM can detect vehicles which are different from poses, shapes, sizes, scales and locations. The experimental results illustrate our method can deal with multiple vehicles in an individual video frame to make up the lack of scales information in ABM. In our experiment, the experiments have been done only in front-view vehicles, and we didn't consider the occlusion and heavy traffic scene, as a result we will expand the detection of rear-view and side-view vehicles, and try to deal with the different pose and occlusion problem in the future.

REFERENCES

[1] http://wardsauto.com/ar/world_vehicle_population_110815.

[2] F.Y. Wang, "Parallel Control: A Method for Data-Driven and Computational Control," Acta Automatica Sinica, vol.39, no.4, pp. 293-302, 2013.

[3] J.P. Zhang, F.Y. Wang, K.F. Wang, et al, "Data-driven intelligent transportaion systems a survey," IEEE Transactions on Intelligent Transportation Systems, vol.12, no.4, pp. 1624-1639, 2011.

[4] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using Gabor filters and support vector machines," 14th International Conference on Digital Signal Processing, vol. 2, pp. 1019-1022, 2002.

[5] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.5, pp. 694-711, 2006.

[6] T. Naito, T. Ito, Y. Kaneda, "The obstacle detection method using optical flow estimation at the edge image," IEEE Intelligent Vehicle Symposium, pp. 817–822, 2007.

[7] B. Han, L. S. Davis, "Density-based multifeature background subtraction with support vector machine," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.34, no.5, pp. 1017-1023, 2012.

[8] H. Y. Cheng, C. C. Weng, and Y. Y. Chen, "Vehicle detection in aerial surveillance using dynamic Bayesian networks," IEEE Transactions on Image Processing, vol.21, no.4, pp. 2152-2159, 2012.

[9] Y. N. Wu, Z.Si, H.Gong, and S. C. Zhu, "Learning active basis model for object detection and recognition," International Journal of Computer Vision, vol.90, no.2, pp. 198-235, 2010.

[10] Y. Li, B. Li, B. Tian, and Q.M. Yao, "Vehicle Detection Based on the And-Or Graph for Congested Traffic Conditions," accepted, DOI: 10.1109 /TITS.2013.2250501, 2013.

[11] http://www.vision.caltech.edu/bouguetj/calib_doc/

[12] 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.

[13] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," The Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, pp. 666-673, 1999.

[14] E. P.Simoncelli, W. T. Freeman, E. H. Adelson, D. J. Heeger, "Shiftable multiscale transforms," IEEE Transactions on Information Theory, vol.38, no.2, pp. 587-607, 1992.

[15] B. A. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature, vol.381, no.6583, pp. 607-609, 1996.

[16] J. Tan, J. Li, X. An, and H. He, "An interactive method for extrinsic parameter calibration of onboard camera," In Intelligent Vehicles Symposium (IV), 2011 IEEE, pp. 236-24, 2011.

[17] Y.T. Li, F.H. Zhu, Y.F. Ai, F.Y. Wang, "On automatic and dynamic camera calibration based on traffic visual surveillance," Intelligent Vehicles Symposium, 2007 IEEE, pp. 358-363, 2007.

[18] K. Wang, H. Huang, Y. Li, and F.Y. Wang, "Research on Lane-Marking line based camera calibration," IEEE International Conference on Vehicular Electronics and Safety, pp. 1-6, 2007.