

Pose Estimation for Vehicles Based on Binocular Stereo Vision in Urban Traffic

Pengyu Liu, Fei Wang^(✉), Yicong He, Hang Dong, Haiwei Yang,
and Yang Yang

Xi'an Jiaotong University, Xi'an, Shaanxi, China
{liu.pengyu, heyicong2013, yanghw.2005,
dhunter}@stu.xjtu.edu.cn,
{wfx, yyang}@mail.xjtu.edu.cn

Abstract. Extensive research has been carried out in the field of driver assistance systems in order to increase road safety and comfort. We propose a pose estimation algorithm based on binocular stereo vision for calculating the pose of on-road vehicles and providing reference for the decision of driving assistant system, which is useful for behavior prediction for vehicles and collision avoidance. Our algorithm is divided into three major stages. In the first part, the vehicle is detected and roughly located on the disparity map. In the second part, feature points on the vehicle are extracted by means of license plate detection algorithm. Finally, pose information including distance, direction and its variation is estimated. Experimental results prove the feasibility of the algorithm in complex traffic scenarios.

Keywords: Pose estimation · Behavior prediction · Binocular stereo vision

1 Introduction

The automotive industry is constantly evolving as there is a growing interest in the field of safety. In order to improve security in driving, several applications of Advanced Driving Assistance Systems (ADAS) were proposed. When a potential safety problem occurs, this system can timely alerts the driver. If the driver does not respond to warnings, measures will be taken by the system automatically so as to avoid the accident. This paper is aimed at the Forward Collision Avoidance System (FCAS), which is designed for assisting the driver to maintain a safe stopping distance related to the vehicle ahead in order to avoid collision when accident occurs or at least reduce the severity of the collision [1]. However, in many cases the safe distance is not maintained because of the driver's occasional distraction or other negligence. Besides, some emergencies, such as sudden changing of lanes, may cause accidents as the drivers don't have enough time for response (cf. Fig. 1). The algorithm we propose is based on binocular stereo vision algorithm which can detect vehicles and calculate its pose information including distance, direction and its variation. So that behaviors of the vehicles can be predicted at the same time. To date, RADAR-based system becomes a very popular solution for FCAS. However, most of them are used in the high class cars due to the exorbitant cost of laser equipment. One of the main goals of our research is

to provide cars with a not only cheap but also reliable forward collision avoidance system with speed, distance as well as pose information. Due to the fact that vision system can provide depth information as well as texture, it is still the first choice considering both cost and efficiency.



Fig. 1. Accident due to sudden changing of lanes

Our paper is organized as follows: the second part is the recent research work. The third part describes the algorithm in detail, including the coarse localization of the vehicles forward, vehicle feature extraction and pose estimation. Experimental results are displayed in the fourth part. The fifth part is the conclusion and prospect of our work.

2 Related Work

Many approaches to assist drivers in avoiding collisions have been developed based on different technologies such as RADAR [2], LASER scanner [3], ultrasounds [4] or vision systems [5–7]. As for drawbacks, the vision field of RADAR can be narrow, and mutual interference may occur when other vehicles on road are installed with same equipment due to radar's active sensing characteristic. Cheap LASERs are easily affected by fog and rain. Ultra-sounds easily fail because of lateral wind.

To date, many researches aims at vehicle detection. In [8], scene depth is deduced from texture gradients, defocus, color and haze using a multi-scale Markov Random Field (MRF). However, this algorithm may fail when the actual scene differs greatly with the training scene because of the algorithm's need for pre-training. In [9], the vehicle's distance measurement is based on the contact between the vehicle's tires and the road. This method may easily overestimate because of the difficulty of obtaining the exact tire-asphalt contact point which in addition does not correspond to the real back of the vehicle.

For number plate detection, there are methods based on Hough transform [10], color SVM classifiers to model character appearance variations [11] and methods based on top-hat transform [12]. The number plate location approach proposed adapts the widely employed morphological top-hat method to vehicles in motion where the position of the vehicle is unknown and therefore the dimensions of the number plate in the image are, in principle, unknown.

For pose estimation, the aim of the Perspective-n-Point problem (PnP) is to determine the position and orientation of a camera given its intrinsic parameters and a set of n correspondences between 3D points and their 2D projections. It is widely used in computer vision, Robotics, Augmented Reality and has received much attention in both the Photogrammetry [13] and Computer Vision [14] communities. A robust pose estimation algorithm is proposed by Yang [15]. He divides approximate coplanar points into coplanar points and non-coplanar points and uses non-coplanar points for final pose calculating.

In the field of behavior prediction, Toru Kumagai [16] uses Bayesian Dynamic Models as well as driving status to predict the parking intention at the intersection. Tesheng Hsiao [17] obtains parameters of turning model based on maximum a posterior estimation. However, the methods above require long time for modeling or training, which can't meet the real-time requirements essential for collision avoidance.

3 Algorithm Description

3.1 Method Overview

Our algorithm is divided into three parts: vehicle detection, vehicle feature extraction and pose estimation. The flowchart of our algorithm is shown in Fig. 2.

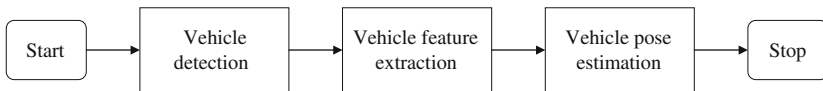


Fig. 2. Flowchart of our algorithm

In order to calculate the pose information of the vehicle, four feature points are needed according to the pose estimate algorithm. Unfortunately, different kinds of vehicles have different shapes, which makes it difficult to find fixed feature points. However, a common element on vehicle is the license plate which has strict regulations in every country. As a result, the vehicle's pose information can be calculated directly by localizing the front vehicle's license plate and establishing the relationship between the license plate's size in the image and the 3D coordinates in space in advance.

Nevertheless, as the front vehicle is far from the camera, the license plate occupies only a small proportion in the image, which makes it difficult to detect and locate. Thus coarse localization of the vehicles is needed previously. We calculate the bounding-box of the front vehicle in image and set it as ROI. License plate detection and pose estimation can be realized subsequently.

The license detection algorithm requires the plates being installed according to the regulations of China. Vehicles without plates or installed with twisted plates are not discussed in this paper.

3.2 Vehicle Detection

The vehicle detection procedure is based on two parts: the disparity map generation based on binocular stereo vision and front vehicle localization.

3.2.1 Scene Disparity Map Calculating

The model of binocular stereo vision observes the same scene from different viewpoints using two fixed cameras with given intrinsic parameters. Correspond pixels are calculated by stereo matching of the left and right images. In the structure of standard epipolar geometry, the epipolar of space point are parallel and aligned in left and right image planes. Suppose the corresponding points of p are p_1 and p_2 in left and right images, the base line is b , focus is f , and the distance from p to base line is z . (cf. Fig. 3).

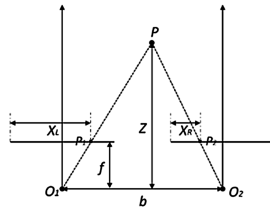


Fig. 3. Parallax theory based on epipolar geometry

According to geometry we have:

$$\frac{b}{Z} = \frac{(b + x_R) - x_L}{Z - f} \tag{1}$$

Then we get z from point p to base line:

$$z = \frac{bf}{x_R - x_L} = \frac{bf}{d} \tag{2}$$

where $d = x_R - x_L$

From the analysis above, the depth z of point p depends on the parallax d only. Thus the location of p in space can be uniquely identified by calculating the corresponding point p_2 of p_1 . We use block matching algorithm based on Graph Cuts.

The disparity map and pseudo-color image calculated by the model of binocular stereo vision are shown in Fig. 4, which is prepared for vehicle localization in the next section.

3.2.2 Vehicle Localization

After generating disparity map, more work is done to locate the vehicle in the image. The algorithm is described in [18] as follows.

1. Count u-disparity on map to detect horizontal line using Hough transform. The detected lines are the position information in x-axis of the vehicle.

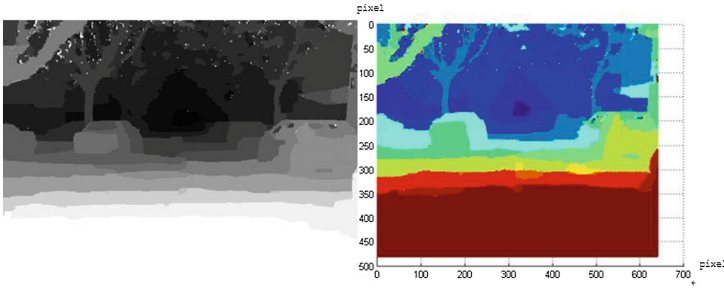


Fig. 4. Disparity map and pseudo-color image (Color figure online)

2. Count v-disparity on map to detect vertical line using Hough transform. The crossing points of road and vehicle are the position information in y-axis of the vehicle.
3. Sort the coordinates of vehicles by y-axis and match the vehicles after sorting. Then the (x, y) information of vehicles can be got.
4. Calculate the 3D position of vehicles according to the binocular stereo vision model.

The detected vehicle is marked with red box as Fig. 5 shows. After calculating the bounding-box of the vehicle forward using binocular stereo vision algorithm, we set the bounding-box area as ROI. We detect the license plate and calculate the coordinates of the feature points in ROI.

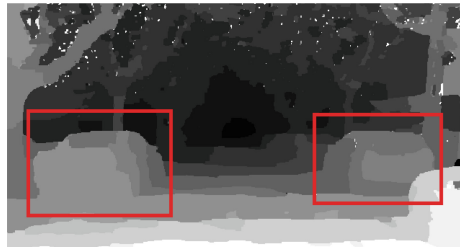


Fig. 5. Vehicle detection (Color figure online)

Considering the real condition, there may be several vehicles in the visual field at the same time. Thus several ROIs may be detected. We choose three nearest ones from the camera, if any, to calculate the pose of vehicles on current lane as well as the neighbor lanes.

3.3 Vehicle Feature Extraction

In order to calculate pose information, four feature points are required at least. However, different kinds of vehicles have unique shapes, which makes it difficult to look for

fixed feature points. Fortunately, the standard of license plate is fixed in every country. In China, for instance, according to regulations based on the public security industry standards of People's Republic of China, we summarize some characteristics of license plate designed for cars which are most common on roads. As Fig. 6 shows.

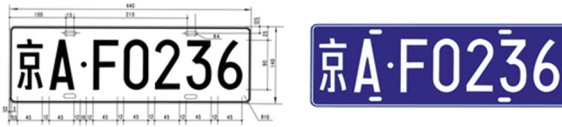


Fig. 6. Sample of license plate of China (Color figure online)

1. The background color of the license plate is quite different from the vehicle and characters.
2. The contour of the license plate is continues or interrupted due to abrasion.
3. The characters on the plate are on the same horizontal line. Much boundary information can be found on the plate.
4. The size of plate is fixed.

On the basis of the characteristics above, neighboring pixels on plate's boundary will vary frequently from 0 to 1 or 1 to 0, and the sum of changes will be larger than a threshold, which can be used for detecting the license plate.

The algorithm for license plate detection in this paper is based on grey level transformation of plate's region. Due to the contrast of gray level between the plate and the vehicle, binaryzation is achieved with a threshold calculated by local histogram. The small areas which are not plates are removed. Finally, in all connected domains, the one with max likelihood is marked as the plate.

The process of license plate detection is shown in Fig. 7. Figure 7-a is the RGB image of one ROI. Binary image with mark points of likelihood areas of license plate is shown in Fig. 7-b. The final plate is marked on Fig. 7-c and the four vertexes are used for pose estimation algorithm in next section.

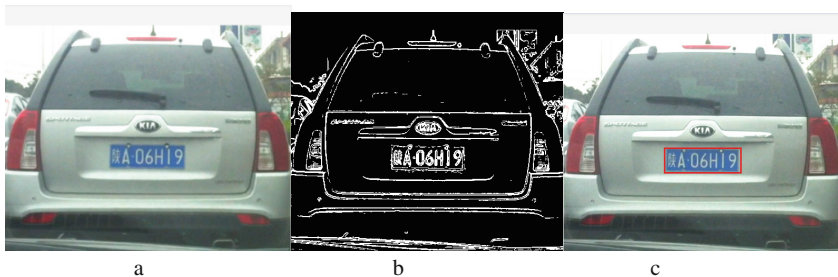


Fig. 7. Results for plate detection (Color figure online)

3.4 Vehicle Pose Estimation

PnP problem is the estimation of the pose of a calibrated camera from n 3D-to-2D point correspondences. The central idea of the algorithm is to express the n 3D points as a weighted sum of four virtual control points and estimate the coordinates of these control points in the camera referential. At last, the rotation matrix R and translation matrix T are calculated. The algorithm is described in detail as follows:

Suppose reference points of n points with 3D coordinates known in the world coordinate system be:

$$p_i, \quad i = 1, \dots, n \tag{3}$$

Similarly, let the 4 control points we use to express their world coordinates be

$$c_j, \quad j = 1, \dots, 4 \tag{4}$$

Then the references can be expressed by control points uniquely:

$$p_i^w = \sum_{j=1}^4 \alpha_{ij} c_j^w, \quad \text{with } \sum_{j=1}^4 \alpha_{ij} = 1 \tag{5}$$

Where α_{ij} are homogeneous barycentric coordinates.

The same relation holds in the camera coordinate system:

$$p_i^c = \sum_{j=1}^4 \alpha_{ij} c_j^c \tag{6}$$

Suppose A be the camera internal calibration matrix and $\{u_i\}_{i=1,\dots,n}$ are the projections of reference points expressed by $\{p_i\}_{i=1,\dots,n}$. We have:

$$\forall i, \quad w_i \begin{bmatrix} u_i \\ 1 \end{bmatrix} = A p_i^c = A \sum_{j=1}^4 \alpha_{ij} c_j^c \tag{7}$$

Where w_i is scalar projective parameter. Matrix A consists of the f_u, f_v focal length coefficients and the (u_c, v_c) principal point.

We now consider the specific 3D coordinates $[x_j^c, y_j^c, z_j^c]^T$ of each control point c_j^c , the 2D coordinates $[u_i, v_i]^T$ of the u_i projections. Then Eq. (7) becomes:

$$\forall i, \quad w_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^4 \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \tag{8}$$

Expand Eq. (8), we have:

$$Mx = 0 \quad (9)$$

Where $x = [c_1^{cT}, c_2^{cT}, c_3^{cT}, c_4^{cT}]^T$ is a 12-vector with the unknowns, and M is a $2n \times 12$ matrix. The solution can be expressed as:

$$x = \sum_{i=1}^N \beta_i v_i \quad (10)$$

Where the set v_i are the columns of the right-singular vectors of M .

Finally, we compute solutions for all four values of N and keep the one that yields the smallest re-projection error:

$$res = \sum_i dist^2(A[R|t] \begin{bmatrix} p_i^w \\ 1 \end{bmatrix}, u_i) \quad (11)$$

Where the rotation matrix R and the translation matrix T represent the direction and distance of the vehicle respectively.

4 Experimental Results

The experiments are implemented with a desktop PC with processor of i7-2.80 GHz, RAM 32 GB. The videos were recorded by two cameras made by Imagesource which are mounted in a car.

4.1 Pose Estimation of Vehicles Forward

In order to prove the robustness of our algorithm, we choose the image with a vehicle at corner. The results are shown in Fig. 8.

Figure 8-a shows the original image. Figure 8-b shows the coarse detection of vehicle according to disparity map. Figure 8-c shows the plate's vertexes detection result. Finally, as proposed in Sect. 3.4, we calculate the pose of vehicle forward and re-project the outlines of the vehicle using a virtual cube with axes for verification, as Fig. 8-d shows. The axes x , y and z are represented with green, yellow and blue respectively.

4.2 Pose Estimation of Vehicles While Changing Lanes

One of our goals is to predict vehicle's behavior, especially the intention of changing lanes, which may cause a collision. We calculate pose variations from consecutive frames. The re-projection results are shown in Fig. 9.

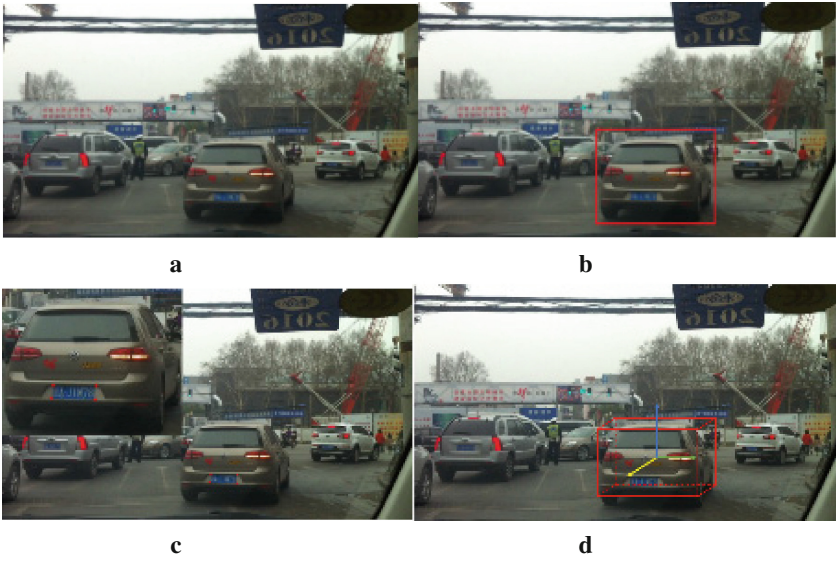


Fig. 8. Pose estimation of vehicle forward (Color figure online)



Fig. 9. Continuously pose estimation

The pose estimating result is expressed as:

$$pose = [x, y, z, \alpha, \beta, \gamma] \tag{12}$$

Where x, y, z represents translation in relevant axis direction and α, β, γ represents the rotation around x axis, y axis and z axis. The pose information in the 6 frames above are displayed in Table 1 and Fig. 10. The last column of Table 1 represents the error of pose estimation algorithm according to Eq. (11).

Table 1. Pose estimation results

Frame	x(cm)	y(cm)	z(cm)	$\alpha(^{\circ})$	$\beta(^{\circ})$	$\gamma(^{\circ})$	err
1	205.6	405.2	15.0	0.1	-0.2	8.6	0.2
2	180.2	320.6	14.2	-0.2	0.1	9.9	0.3
3	160.3	250.6	16.1	0.1	0	10.2	1.5
4	140.6	365.2	14.5	-0.3	0.3	9.6	2.2
5	80.8	1050.3	15.3	0	-0.2	7.5	1.6
6	30.5	1500.3	14.6	0.5	-0.3	4.3	1.2

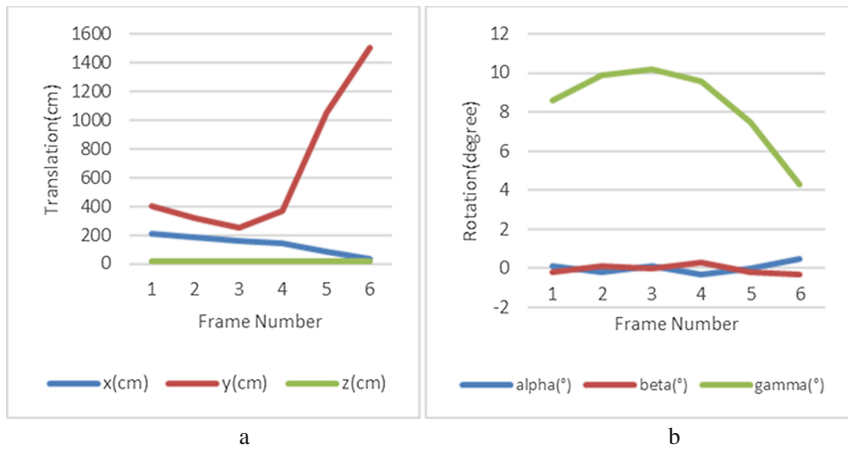


Fig. 10. Line chart of pose estimation results

From Fig. 10-a, the x value is continually decreasing when the vehicle is approaching to our car in horizontal direction. The y value increases first and then decreases as the vehicle comes close and then move away. The z value maintains around zero because the road is relatively flat.

From Fig. 10-b, the angle of gamma, which represents rotation around z axis, increases first and then decreases. It indicates the turning action of the vehicle. The angle of beta and gamma undulate around zero due to the flat road.

The results and analysis above exactly match the vehicle’s behavior in the picture, which proves that the pose we get can be used in vehicle behavior prediction as a reliable reference.

5 Conclusion and Forecast

This paper puts forward an innovative pose estimation method for vehicles' behavior prediction based on binocular stereo vision. Using the binocular camera, we get pose information of other vehicles including distance, speed and direction, which is useful for assisting drivers to keep a safe distance and predicting the behavior of other vehicles.

The method proposed in this paper uses two cameras as sensors, which cost much less than the traditional laser equipment. In addition, the vehicle detection based on disparity map gives coarse localization of vehicle and thus greatly reduces the searching space and increases the correct rate of license plate detection. We get vehicle's pose information with only four points based on the proposed pose estimation algorithm gets, which is accurate enough and provides sufficient reference to the ADAS. In road tests, pose information including distance, direction and its variation is estimated, and the analysis proves the feasibility of the algorithm in complex traffic scenarios.

Although the pose information can be calculated with four feature points in most cases using the algorithm we propose, more feature points are needed to reduce the error in pose estimation. In the future, more work are needed for feature point extraction and error analysis [19] to improve the accuracy.

Acknowledgment. This work is sponsored by National Natural Science Foundation of China (No.612173366) and National High Technology Research and Development Program of China (No.2013AA014601).

References

1. Bertolazzi, E., Biral, F., Da Lio, M., et al.: Supporting drivers in keeping safe speed and safe distance: the sasence subproject within the European framework programme 6 integrating project Prevent. *IEEE Trans. Int. Transp. Syst.* **11**(3), 525–538 (2010)
2. Skutek, M., Mekhaïel, M., Wanielik, G.: A Precrash system based on radar for automotive applications. In: *Proceedings of IEEE of Intelligent Vehicles Symposium, Columbus*, pp. 37–41 (2003)
3. Velupillai, S., Guvenc, L.: Laser scanners for driver-assistance systems in intelligent vehicles. *IEEE Control Syst. Mag.* **29**(2), 17–19 (2009)
4. Alonso, L., Pérez-Oria, J., Fernández, M., Rodríguez, C., Arce, J., Ibarra, M., Ordoñez, V.: Genetically tuned controller of an adaptive cruise control for urban traffic based on ultrasounds. In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds.) *ICANN 2010, Part II. LNCS*, vol. 6353, pp. 479–485. Springer, Heidelberg (2010)
5. Nedevschi, S., Danescu, R., Frentiu, D., et al.: High accuracy stereovision approach for obstacle detection on non-planar roads. In: *Proceedings of IEEE Intelligent Engineering Systems (INES), Cluj Napoca, Romania*, pp. 211–216 (2004)
6. Broggi, A., Bertozzi, M., Fascioli, A., Guarino, C., Piazzini, A.: Visual perception of obstacles and vehicles for platooning. *IEEE Trans. Intell. Transp. Syst.* **1**(3), 164–176 (2000)
7. Sun, Z., Bebis, G., Miller, R.: On-road vehicle detection: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(5), 694–711 (2006)

8. Michels, J., Saxena, A., Ng, A.Y.: High speed obstacle avoidance using monocular vision and reinforcement learning. In: 22nd International Conference on Machine Learning, Bonn, Germany, pp. 593–600 (2005)
9. Gat, I., Benady, M., Shashua, A.: A monocular vision advance warning system for the automotive aftermarket. SAE Technical Paper (2005)
10. Yanamura, Y., Goto, M., Nishiyama, D., Soga, M., Nakatani, H., Saji, H.: Extraction and tracking of the license plate using Hough transform and voted block matching. In: Proceedings of IEEE Intelligent Vehicles Symposium, pp. 243–246 (2003)
11. Chengwen, H., Yannan, Z., Jiabin, W., Zehong, Y.: An improved method for the character recognition based on SVM. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, pp. 457–461 (2006)
12. Hung, K.M., Hsieh, C.T.: A real-time mobile vehicle license plate detection and recognition. *J. Sci. Eng.* **13**, 433–442 (2010)
13. McGloves, C., Mikhail, E., Bethel, J.: *Manual of Photogrammetry*. American Society For Photogrammetry and Remote Sensing, New York (2004)
14. Hartley, R., Zisserman, A.: *Multiple View Geometry In Computer Vision*. Cambridge University Press, Cambridge (2003)
15. Yang, H., Wang, F., Chen, L., He, Y., He, Y.: Robust pose estimation algorithm for approximate coplanar targets. In: Huang, D.-S., Jo, K.-H., Wang, L. (eds.) ICIC 2014. LNCS, vol. 8589, pp. 350–361. Springer, Heidelberg (2014)
16. Kumagai, T., Sakaguchi, Y., Okuwa, M. et al.: Prediction of driving behavior through probabilistic inference. In: Proceedings of 8th International. Conference on Engineering Applications of Neural Networks, pp. 117–123 (2003)
17. Hsiao, T.: Time-varying system identification via maximum a posteriori estimation and its application to driver steering models. In: American Control Conference, pp. 684–689. IEEE (2008)
18. Hu, Z., Uchimura, K.: UV-disparity: an efficient algorithm for stereovision based scene analysis. In: Proceedings of IEEE on Intelligent Vehicles Symposium, pp. 48–54. IEEE (2005)
19. Song, Y., Wang, F., Gao, S., Yang, H., He, Y.: Error tracing and analysis of vision measurement system. In: Huang, D.-S., Bevilacqua, V., Premaratne, P. (eds.) ICIC 2014. LNCS, vol. 8588, pp. 729–740. Springer, Heidelberg (2014)