IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON INNOVATION AND APPLICATION OF INTERNET OF THINGS AND EMERGING TECHNOLOGIES IN SMART SENSING

Received October 22, 2019, accepted November 10, 2019, date of publication November 19, 2019,
date of current version December 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954341

Vehicle Ego-Localization Based on the Fusion of Optical Flow and Feature Points Matching

CHENG XIN[1], ZHOU JINGMEI[2], ZHAO XIANGMO[1], (Member, IEEE), WANG HONGFEI[1], AND CHANG HUI[1]

[1]School of Information Engineering, Chang'an University, Xi'an 710064, China
[2]School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China

Corresponding authors: Cheng Xin (xincheng@chd.edu.cn) and Zhou Jingmei (jmzhou@chd.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1600600, in part by the National Natural Science Funds of China under Grant 51278058, in part by the 111 Project on Information of Vehicle-Infrastructure Sensing and Intelligent Transportation System (ITS) under Grant B14043, in part by the Shaanxi Natural Science Basic Research Program under Grant 2019NY-163, and in part by the Special Fund for Basic Scientific Research of Central Colleges, Chang'an University, China, under Grant 300102329101 and Grant 300102249101.

ABSTRACT To meet the requirement of vehicle real-time and precise ego-localization on the flat road of city, a vehicle ego-localization method based on the fusion of optical flow and feature points matching is proposed. A novel FAST algorithm with self-adaptive threshold is applied to detect feature points. Based on the assumption of flat plane, the improved Lucas-Kanade algorithm is carried out to track feature points, and then the custom LARSAE is used to amend vehicle offsets. Meanwhile, Hu moments are used as the feature descriptor to complete image matching, realizing vehicle motion estimation. These two methods are fused by the discrete kalman filter to update and optimize vehicle position. Experimental results show that the fusion algorithm overcomes the shortcomings of poor positioning accuracy of optical flow and the low processing speed of feature matching, and is able to provide more accurate real-time positioning output, having a certain robustness for circumstances such as illumination change and low pavement texture.

INDEX TERMS Vehicle ego-localization, image matching, optical flow, optimized FAST, Kalman filter.

I. INTRODUCTION

High-precision vehicle position information is of great significance for vehicle behavior safety analysis. In recent years, machine vision based vehicle positioning has gradually become the research hotspot of autonomous navigation [1], [2]. Compared with the traditional positioning methods (GPS, INS, Odometer, etc.), visual odometer does not have signal blind area, and is not affected by wheel slip, having the advantages of independence, high positioning accuracy, and not susceptible to interference [3], [4]. Uchiyama *et al.* [5] put forward an ego-localization using streetscape image sequences from in-vehicle cameras. All-dimensional street view images were shot in advance to store in a database. In the process of vehicle moving, two cameras were applied to take street scenes from different angles in real-time, then they were matched with pre-existing street images, and the triangulation method was used to calculate the precise location of vehicle. This method had high positioning accuracy, but required to establish the image database beforehand. So it wasted time and energy, and was not applicable to vehicle real-time positioning without capturing panoramic images. Wu and Ranganathan [6] proposed the vehicle localization using road markings. According to the actual coordinates of the detected road signs in the database and its relative distance to vehicle, the vehicle position could be estimated precisely. However, establishing the database also wasted time and required to make a regular update. In the literatures [7], [8], according to the homography matrix calculated by local feature points matching, vehicle trajectory was calculated to realize ego-localization. The method had the accurate positioning information in a short time, but it had low efficiency and was affected by the accumulated error to make the trajectory drifted over time. Zhang *et al.* [9] introduced a method of visual odometry based on random finite set statistics in urban environment. Targets were

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

optimally selected and dynamically evaluated by a recursive filtering algorithm in the presence of clutter and high association uncertainty, to make sure them more accurate. Meanwhile, he presented that vehicle trajectory could be calculated based on random finite set statistics using a single camera in complex urban environments [10], and a sensor fusion approach with cumulative error elimination was used to achieve vehicle precise positioning [11]. Pink *et al.* [12] proposed the visual features for vehicle localization and ego-motion estimation, however the speed was relatively slow. Vehicle positioning based on image matching had good accuracy, but the time efficiency was too low. After nearly 30 years of development, the optical flow method could meet the requirements of certain accuracy and stability for vehicle motion estimation, and because of the fast computational speed, visual odometry based on optical flow method had been able to reach the level of practical application [13]. Vision-based mobile robot navigation based on optical flow proposed by Ban [14] could be analogically applied in vehicle, which studied the optical flow effects of images captured by panoramic camera, fish-eye lens and spherical panoramic stereo imaging system, to accomplish obstacle avoidance and depth estimation. Steven et al. put forward the accurate visual odometry from a rear parking camera [2], [15], which aligned images to calculate relative offsets by Efficient Second-order Minimisation (ESM) algorithm, and fused GPS to complete accurate positioning. Monocular vision odometry based on the fusion of optical flow and feature points matching was put forward by Zheng *et al.* [16] and Zheng [17]. The fusion of optical flow and feature points matching was accomplished by using kalman filter, to realize ego-localization on the flat surface by reducing error on the basis of considering real time. An optical flow-based integrated navigation system inspired by insect vision was presented by Pan *et al.* [18], which also used the kalman filter to correct the cumulative error.

In order to ensure real time performance with better positioning accuracy, this paper proposes a vehicle ego-motion estimation method based on the fusion of optical flow and feature points matching. A novel FAST algorithm with self-adaptive threshold is applied to detect feature points, and vehicle offsets are calculated by using the modified Lucas-Kanade algorithm. Meanwhile, Hu moments describe feature points to complete image matching, completing vehicle motion estimation. Finally, these two methods are fused by using kalman filter to update and optimize vehicle position. Through the campus environment test, experimental results show that the method can obtain more accurate and smoother vehicle trajectory, and provide more complete location information.

The remaining part of this paper is organized as follows: The implementation procedure of the proposed method is introduced in the next section, together with the algorithm principle. Some experiments using real data are conducted and the analyses are shown in Section III. Finally, conclusions are summarized in the last section.

II. VEHICLE EGO-LOCALIZATION IMPLEMENTATION PROCEDURE

A. FAST WITH SELF-ADAPTIVE THRESHOLD

Feature point extraction is the basis of vehicle ego-localization with a lot of existing methods such as Harris and Stephens [19], SIFT [20], SURF [21] etc. FAST (Features from Accelerated Segment Test) [22] is widely used. It is a simple and rapid, but has the following disadvantages: (1) Because of the difference in contrast and noise for images obtained under different conditions, FAST with the fixed threshold has poor robustness; (2) There is a high similarity between the real feature points and their nearby pixels, so the feature block can be produced, leading to high agglutination rate and reducing the application performance. In order to improve these limitations, dynamic global threshold and local threshold are combined to enhance the stability and adaptive ability of the algorithm. At the same time, NMS (Non-Maximum Suppression) is used to inhibit the formation of multiple feature point blocks.

1) FEATURE EXTRACTION BASED ON DYNAMIC GLOBAL THRESHOLD

The FAST threshold represents the minimum contrast of detected feature points and the maximum tolerance of resisted noise. The lager the threshold value is, the less the number of detected points is, otherwise the more. Although using fixed threshold is simple in calculation, it cannot meet the requirement of feature point extraction for different images. So the KSW entropy method is used to dynamically set a global threshold T_1 in this paper.

(1) N sample images are selected, which don't exist significant interference from light, shadow, etc.

(2) The KSW entropy method is used to calculate the global threshold T_{1i}, $i = 1, 2, \cdots, N$.

A threshold value t is set to divide gray range [0, L-1] into two parts of s_1 and s_2, which represent respectively pixel frequency distributions of [0, t] and [t+1, L-1]. That is $s_1 = \{p_0, p_1, p_2, \ldots, p_t\}$, $s_2 = \{p_{t+1}, p_{t+2}, p_{t+3}, \ldots, p_{L-1}\}$, where p_i is the frequency of (i)th level gray scale.

Setting $P_t = \sum_{i=0}^{t} p_i$, the entropies S_1 and S_2 of s_1 and s_2 are relatively defined as

$$S_1 = -\sum_{i=0}^{t} \frac{p_i}{P_t} \ln \frac{p_i}{P_t} \tag{1}$$

$$S_2 = -\sum_{i=t+1}^{L-1} \frac{p_i}{1-P_t} \ln \frac{p_i}{1-P_t} \tag{2}$$

Therefore, the image entropy S is the sum of S_1 and S_2, namely $S = S_1 + S_2$. According to the above description and inference, when traversing each t in gray scale range, two parts of entropies are calculated. Finally, we can get the gray-scale level $T_{\max}$ corresponding to the maximum sum of entropies, and $T_{\min}$ corresponding to the minimum. So the

best global threshold of the (i)th image is given by

$$T_{1i} = k \times |T_{\max} - T_{\min}| \tag{3}$$

where, k is the proportionality coefficient, and k is equal to 0.2 in the experiment.

(3) The average of sample images is calculated by equation (4), and it is the dynamic global threshold T_1 for such images.

$$T_1 = \sum_{i=1}^{N} T_{1i} \Big/ N \tag{4}$$

The feature extraction algorithm based on dynamic global threshold is to calculate a fixed optimal global threshold for each type of image. It overcomes the poor robustness of the FAST algorithm. However, comparing with the algorithm that calculates an optimal global threshold for each image, its time efficiency has greatly improved.

2) FEATURE OPTIMIZATION BASED ON DYNAMIC LOCAL THRESHOLD

The global threshold T_1 meets the requirements of different images on thresholds, but its correlation with global gray distribution cannot adapt to local gray change. So the study further screens through setting dynamic local threshold T_2 for feature points obtained by dynamic global threshold T_1.

Analyzing the contrast of image gray value, the adaptive local threshold value T_2 is calculated under the different regions of the image. (x_0, y_0) is assumed as the candidate feature point, for which chooses a square area with length of side L. In this area, the dynamic local threshold T_2 is calculated by

$$T_2 = k \times \frac{\frac{1}{n} \times \left[\sum_{i=1}^{n} I_{i_{\max}} - \sum_{i=1}^{n} I_{i_{\min}} \right]}{I_{i_{aver}}} \tag{5}$$

where, k is the proportionality coefficient, $I_{i_{\max}}$ and $I_{i_{\min}} (i = 1, 2, \ldots, n)$ represent respectively n grays values of the maximum and minimum in the square area, $I_{i_{aver}}$ is the average of gray values in the square area.

Setting $\Delta = \frac{1}{n} \times \left[\sum_{i=1}^{n} I_{i_{\max}} - \sum_{i=1}^{n} I_{i_{\min}} \right]$, it is the contrast. Because the nature of the FAST algorithm is a measure of contrast for adjacent pixels, so the local threshold is related to the local contrast of image (a proportional relation).

Through the local threshold T_2 further screening, the candidate feature points are made more stable. Then in order to eliminate the gathered phenomenon of feature point block, the NMS is used to get finally the effective FAST feature points.

B. VEHICLE MOTION ESTIMATION USING LUCAS-KANADE

Bouguet [23] proposed an improved Lucas-Kanade algorithm aiming at affine transform based on pyramid layering, which had fast computing speed and high stability. Its principle is that motion vector and transformation matrix of pixels are selected to make the gray difference of images minimum. Assuming a pixel $u = [u_x, u_y]^T$ of the previous frame I, it is matched with a pixel $v = [u_x + d_x, u_y + d_y]^T$ of the next image J, that is the gray difference of two pixels is the least. So $d = [d_x, d_y]^T$ is the optical flow of the pixel u in I.

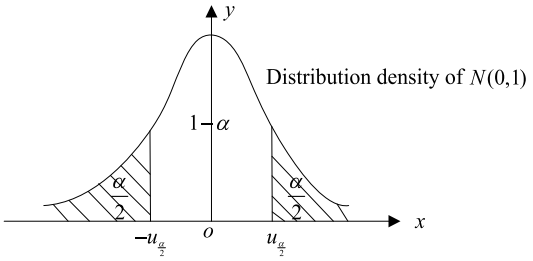


FIGURE 1. Standard normal distribution diagram.

The transformation matrix of two images is defined to

$$A = \begin{bmatrix} 1 + d_{xx} & d_{xy} \\ d_{yx} & 1 + d_{yy} \end{bmatrix} \tag{6}$$

where $d_{xx}, d_{xy}, d_{yx}, d_{yy}$ are parameters of affine transform, so the gray difference is expressed to

$$\begin{aligned} \varepsilon(d, A) &= (d_x, d_y, d_{xx}, d_{xy}, d_{yx}, d_{yy}) \\ &= \sum_{x=-w_x}^{w_x} \sum_{y=-w_y}^{w_y} (I(x+u) - J(Ax + d + u))^2 \end{aligned} \tag{7}$$

where, w_x and w_y are used to set up the rectangle area of image whose size is $(2w_x + 1) \times (2w_y + 1)$, and their values are 7, 8, 10 or 20 in general.

The improved Lucas-Kanade algorithm is used as follows. Firstly, two consecutive frames are respectively built pyramids by sampling, all improved FAST feature points in the previous frame are traced from the top layer, computing optical flow d and transformation matrix A, then gray difference ε in this layer of improved FAST feature points between two consecutive frames after using d and A is made minimum by constant iteration. And the result of this layer is transferred to the next layer, the optical flow d and transformation matrix A are recalculated, then continue to transfer until the last layer (the layer of original image), the d and A of the layer are the final result. The advantage of the algorithm is that it has a good tracking effect for small inter-frame displacement motion, and can increase the number of layers to meet the large-scale displacement tracking in building pyramids.

In order to effectively and quickly retain more accurate optical flow, the LARge SAmple Estimation (LARSAE) is proposed to eliminate error. Assuming that the image optical flow in the pixel coordinate system is matrix X, a large number of samples $(X_1, X_2, \cdots, X_n)$ is obtained by the above method, and the confidence interval of u (the mean of the matrix) can be estimated by $\bar{X}$(the sample mean). According to the central limit theorem, when n is very large, $\bar{X}$ approximately obeys the normal distribution, therefore U approximately obeys the standard normal distribution as shown in Fig.1,

where S is standard deviation.

$$U = \frac{\bar{X} - u}{\frac{S}{\sqrt{n}}} \tag{8}$$

$1-\alpha$ is known, $u_{\frac{\alpha}{2}}$ (Upper quantile about $\alpha/2$ of the standard normal distribution) can be found to make Equation (9) right.

$$P\{|U| < u_{\frac{\alpha}{2}}\} = P\left\{\frac{|\bar{X} - u|}{\frac{S}{\sqrt{n}}} < u_{\frac{\alpha}{2}}\right\} \approx 1 - \alpha \tag{9}$$

So the confidence interval of u is $\left(\bar{X} - u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$.

Regardless of the rotation, affine and projection, the confidence interval of u can be estimated. Feature points whose optical flows are not included in the interval are eliminated as error data, the feature points within the confidence interval are as effective feature points to calculate vehicle trajectory, where the confidence probability $1-\alpha$ is 95%.

The initial position of vehicle is assumed to one point in the world coordinates, and the ground is a plane, so the height in Z direction is always set to 0. The movement of camera and vehicle is rigid, and the projection relationship between camera coordinate system and image coordinate system is obtained by using the calibration algorithm [24]. Therefore, the optical flow d and transformation matrix A obtained by effective feature points in image coordinate system can be transformed to the new coordinates in three-dimensional space, then the vehicle position is estimated and vehicle trajectory can be drawn.

C. VEHICLE MOTION ESTIMATION BASED ON FEATURE POINTS MATCHING

The main idea of this section is as follows. Hu moments are selected as feature descriptors to describe feature points obtained by the optimized FAST algorithm. They are the statistical features with invariants of translation, rotation and scale transformation. The maximum and minimum distance method is taken as the matching criterion. Then 2-D matching points are obtained and further corrected by the literature [7]. Finally, the coordinate transforming algorithm above (Similar to the section 2.2) is used to estimate vehicle trajectory.

$(p+q)th$ order moment is defined by equation (10), its corresponding center moment is calculated by equation (11).

$$m_{pq} = \sum_x \sum_y x^p y^p f(x, y) \tag{10}$$

$$u_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q f(x, y) \tag{11}$$

where, (x, y) is for the image coordinates, $f(x, y)$ is the image gray value, $\overline{x} = \frac{m_{10}}{m_{00}}, \overline{y} = \frac{m_{01}}{m_{00}}$.

The u_{pq} is normalized to get

$$\eta_{pq} = \frac{u_{pq}}{u_{00}^r} \tag{12}$$

where, $r = (p+q)/2 + 1, p+q = 2, 3, \cdots$.

Hu moments have the following seven orthogonal invariants of an image [25]:

$$\begin{aligned}
\phi_1 &= \eta_{20}+\eta_{02} \\
\phi_2 &= (\eta_{20}-\eta_{02})^2+4\eta_{11}^2 \\
\phi_3 &= (\eta_{30}-3\eta_{12})^2+(3\eta_{21}-\eta_{03})^2 \\
\phi_4 &= (\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2 \\
\phi_5 &= (\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{12}+\eta_{03})^2] \\
&\quad +(3\eta_{21}-\eta_{03})(\eta_{03}+\eta_{21})[3(\eta_{12}+\eta_{30})^2-(\eta_{03}+\eta_{21})^2] \\
\phi_6 &= (\eta_{20}-\eta_{02})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2] \\
&\quad +4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{03}+\eta_{21}) \\
\phi_7 &= (3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2] \\
&\quad +(3\eta_{12}-\eta_{30})(\eta_{03}+\eta_{21})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]
\end{aligned} \tag{13}$$

When generating the feature descriptors of FAST feature points, each feature point is as the center, the radius is $L/2$ to select a circular area, and then Hu moments of this area are calculated as descriptors. Schematic diagram is shown in Fig.2.

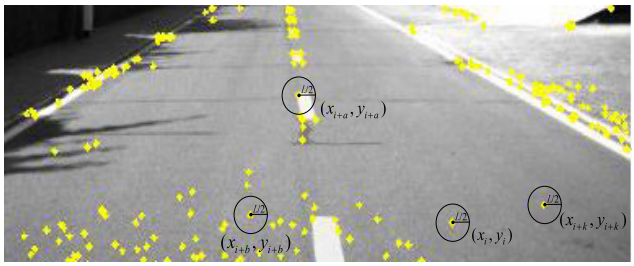


FIGURE 2. Schematic diagram of calculating feature descriptors.

The gray mean value F and the standard deviation σ of the circular region are extracted as gray feature descriptors, which have translational and rotational invariance as shown by equations (14) and (15).

$$F = \sum_{(x-a)^2+(y-b)^2=r^2} f(x, y) \Big/ T \tag{14}$$

$$\sigma = \sqrt{\frac{\sum_{(x-a)^2+(y-b)^2=r^2} (f(x, y) - F)^2}{T}} \tag{15}$$

where, (a, b) is the position coordinate of one feature point, T is the total number of pixels in the circular area, and $f(x, y)$ is the gray value of the position (x, y) in the circular area.

The maximum and minimum distance method is used to calculate the similarity between feature points, which is defined by equation (16). A judgment standard is that the difference between the maximum distance and the second maximum distance is greater than the threshold T, which is

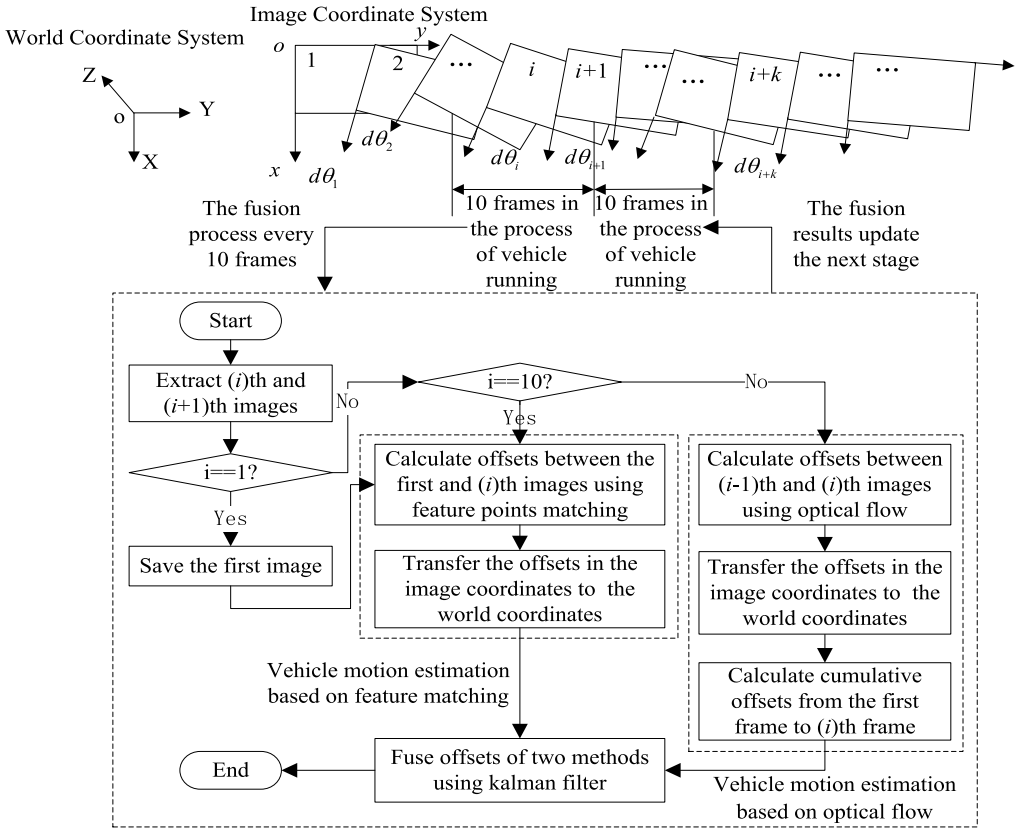


FIGURE 3. Flow chart of the fusion method using Kalman filter.

used to screen the effective matching points.

$$r_{ij} = \frac{\sum_{k=1}^{M} \min(f_{ik}, f_{jk})}{\sum_{k=1}^{M} \max(f_{ik}, f_{jk})} \quad (i, j = 1, 2, \cdots, N+1) \quad (16)$$

where, i and j are tags of feature points, M is the total number of elements in a feature vector, N is the total number of feature vectors, f_{ik} represents feature value of (i, k), and (i, k) is the $(k)th$ element of the $(i)th$ feature vector.

The correct matching points are obtained, and then the vehicle trajectory estimation is accomplished by the literature [7].

D. KALMAN FILTER-BASED FUSION

The vehicle location estimation based on optical flow method has fast computing speed and high positioning frequency, so it is suitable for vehicle real-time positioning with small displacement between frames. However, the cumulative error will be increased with time. The positioning method based on feature points matching has higher precision for existing large displacement between frames, but its positioning frequency is low and run time is long. To get the best of both approaches, these two kinds of vehicle location estimation methods are fused by using the discrete kalman filter, whose process is shown in Fig.3.

As seen in Fig. 3, in one stage, the offsets are mainly calculated based on optical flow for two consecutive frames, and continue to accumulate from the first image. When the number of accumulated images is 10, the first frame and the $10th$ frame are matched based on feature matching method to calculate offsets. Then the results of two methods (their offsets in the world coordinates) are fused by kalman filter to get a more accurate result. Because feature points matching need a longer time, after a completion, a few frames have been computed by optical flow. Therefore, the fusion result will be as the initial value to update the next stage of computing. Finally, this fusion process is repeated until all images are calculated.

In the fusion process, time update equation o of the discrete kalman filter is given by

$$\begin{aligned} \hat{x}_{k,k-1} &= A\hat{x}_{k-1} + Bu_{k-1} \\ P_{k,k-1} &= AP_{k-1}A^T + Q \end{aligned} \quad (17)$$

Observation equation is defined as

$$z_k = Hx_k + v_k \quad (18)$$

where, $\hat{x}_{k-1}$ is the cumulative offsets of calculating 10 frames based on optical flow; z_k is the offsets using feature points matching; H is unit matrix; u_{k-1} represents the position control function; P_{k-1} is the covariance estimation of the previous frame; Q represents the process noise covariance matrix; B is the ride gain, which is set to a constant; A is a constant; v_k represents observation noise, which is white noise obeying normal distribution, $p(v) \sim N(0, R)$.

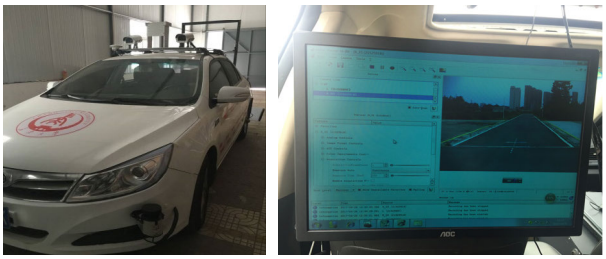
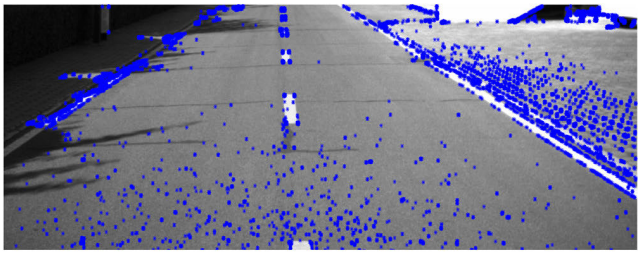
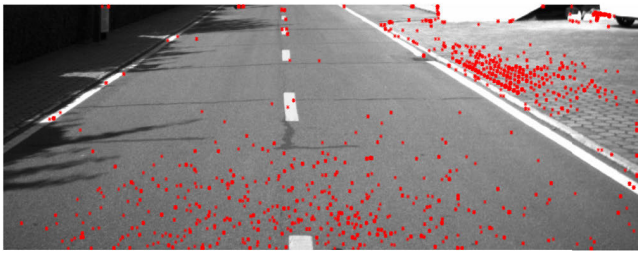


FIGURE 4. Experiment equipment.

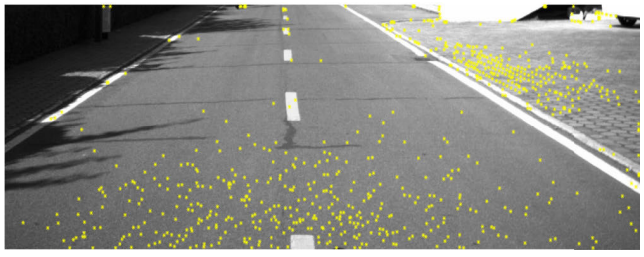
(a)



(b)



(c)



(d)

FIGURE 5. Effects of the improved FAST algorithm: (a) original image; (b) T_1; (c) $T_1 + T_2$; (d) $T_1 + T_2$+ NMS.

Measurement equation for update process is calculated by

$$\begin{aligned} K_k &= P_{k,k-1}H^T(HP_{k,k-1}H^T + R)^{-1} \\ \hat{x}_k &= \hat{x}_{k,k-1} + K(z_k - H\hat{x}_{k,k-1}) \\ P_k &= (I - K_kH)P_{k,k-1} \end{aligned} \tag{19}$$

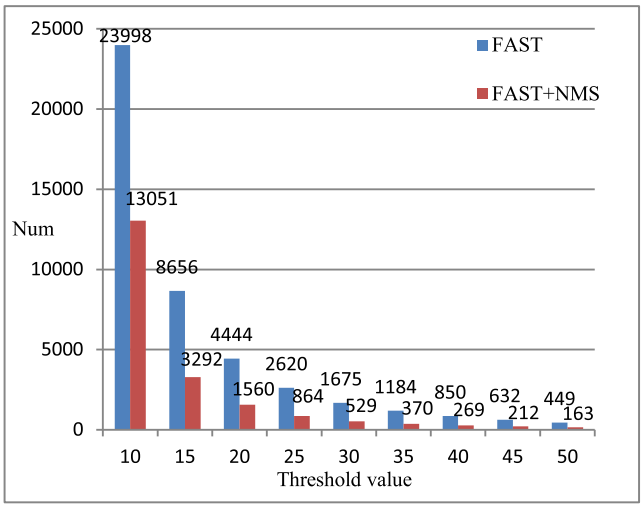


FIGURE 6. Relationship of the common FAST algorithm and threshold values.

TABLE 1. Comparison of feature extraction algorithms.

Performance / Algorithms	Feature points' number	Running time (s)
SIFT	2065	0.5174
SURF	1246	0.3263
Harris	100	0.2924
FAST+NMS(T=30)	529	0.1285
The paper	730	0.1306

where, R is the covariance matrix of observation noise, which is set to the constant.

The advantage of fusing optical flow and feature points matching is to reduce the mean square error. Because optical flow is the main positioning method of this paper, its result is affected by illumination change to produce deviation. Meanwhile, when ground texture is unclear, the number of feature points is less, and produce higher error percentage. Through kalman filter, the error can be effectively reduced to improve the precision.

III. EXPERIMENTAL RESULTS AND ANALYSIS

This paper conducts tests using the intelligent car of Chang'an University, whose experiment devices are shown in Fig.4. German basler aca1600-60gm-gc industrial camera is applied to collect images with resolution of 1600∗1200. Sampling frequency is 60 Hz, and the average speed of vehicle is 25 *km/h*. Image sequence are collected in testing ground.

In one experiment, the 1632*th* frame of pavement image sequence is processed by the novel FAST algorithm with self-adaptive threshold of the study to extract feature points, whose results are shown in Fig.5. Data show that the number of feature points using the global dynamic threshold is 6851, the number is 1545 after adding the local dynamic threshold, and it is 730 by using NMS. This process ensures

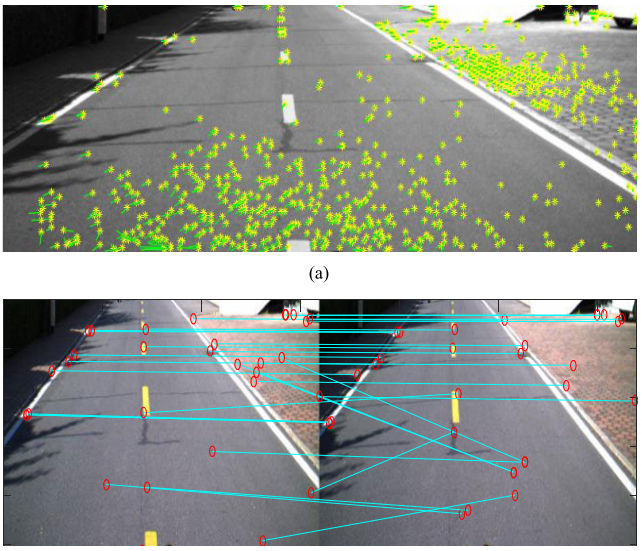


FIGURE 7. Matching effects: (a) optical flow; (b) feature matching method.

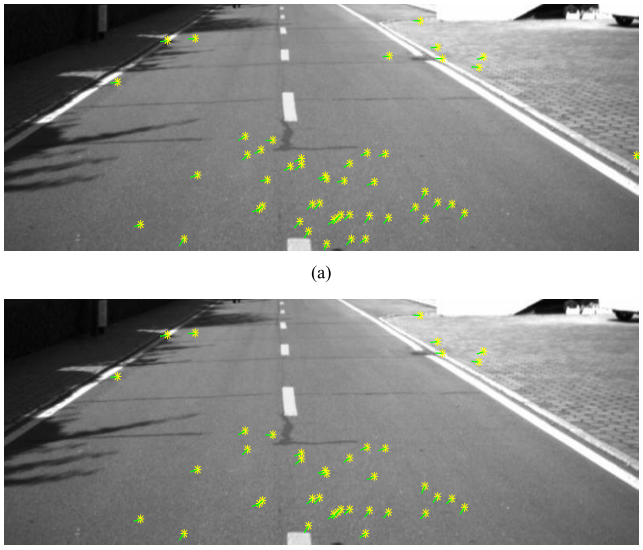


FIGURE 8. Result of LARSAE: (a) $\alpha = 0.05$; (b) $\alpha = 0.1$.

the stability of feature points, and solves the phenomenon of feature point blocks. Fig.6 shows the number of feature points using the common FAST with different thresholds. Through comparison with it, the improved FAST can adjust adaptively threshold according to the actual situation, to obtain more effective feature points, which has more adaptable. Observing Table 1, the speed of FAST algorithm is very fast, the optimized FAST algorithm increases the complexity and time-consuming, but it is still quick and meets the needs of real-time system compared with other algorithms.

When estimating vehicle position based on optical flow, in order to quicken running speed, a part of image is only selected to calculate optical flow. For 1632*th* and 1633*th* frames, its result is shown in Fig. 7(a). Every 10 frames,

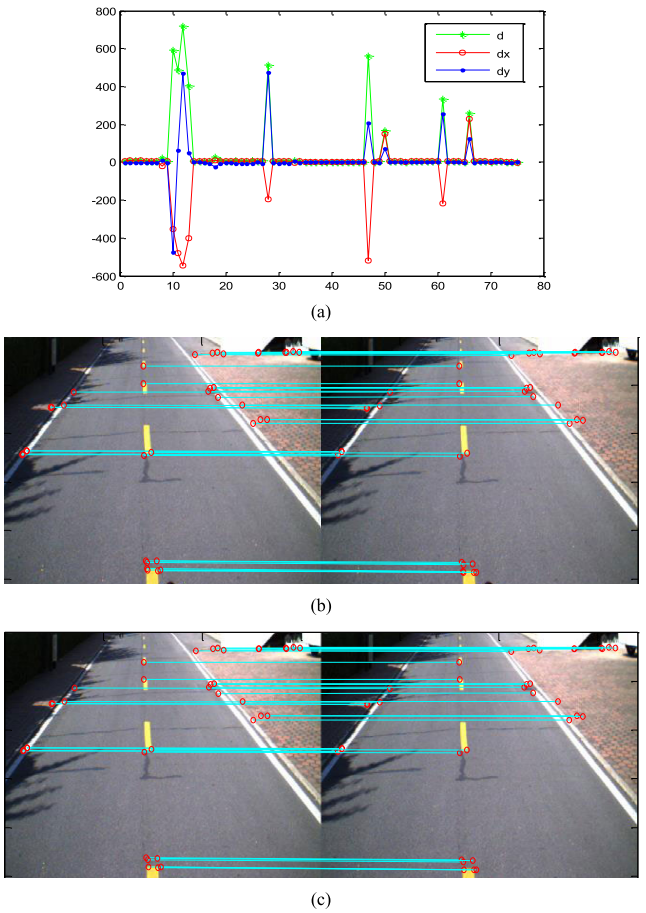


FIGURE 9. Result of literature [7]: (a) different offset values; (b) eliminating effect using *dx*; (c) eliminating effect using *d*.

feature points matching are used to estimate the vehicle position, whose matching result is shown in Fig. 7(b) for 1630*th* and 1639*th* frames. We can find that the number of matching points for two algorithms meets the requirement of vehicle offsets calculation, but they also exist false matching.

In order to remove the false matching points of optical flow method, LARSAE is applied, whose effect is shown in Fig.8. Different confidence probability $(1 - \alpha)$ has certain impact on the number of feature points. The smaller the confidence probability is, the less the number of points is. Correct matching points are obtained by using the approach in literature [7] for feature matching method, as shown in Fig.9, where d is the euclidean distances of dx and dy. When the different offsets (dx, dy, and d) are used to remove, the number of getting feature points is different, so it can be chosen according to the actual situation.

After correct matching points are obtained through the two methods, they are transformed to the new values in three-dimensional space of camera coordinate system according to calibration algorithm [24], where the Z-offset is known as 0. Selecting a lot of images as a test case, for two adjacent frames, optical flow method is applied to calculate offsets of rotation and translation, which are continued to accumulate, and makes a result $[l_dx, l_dy, l_d\theta]$ every 10 frames

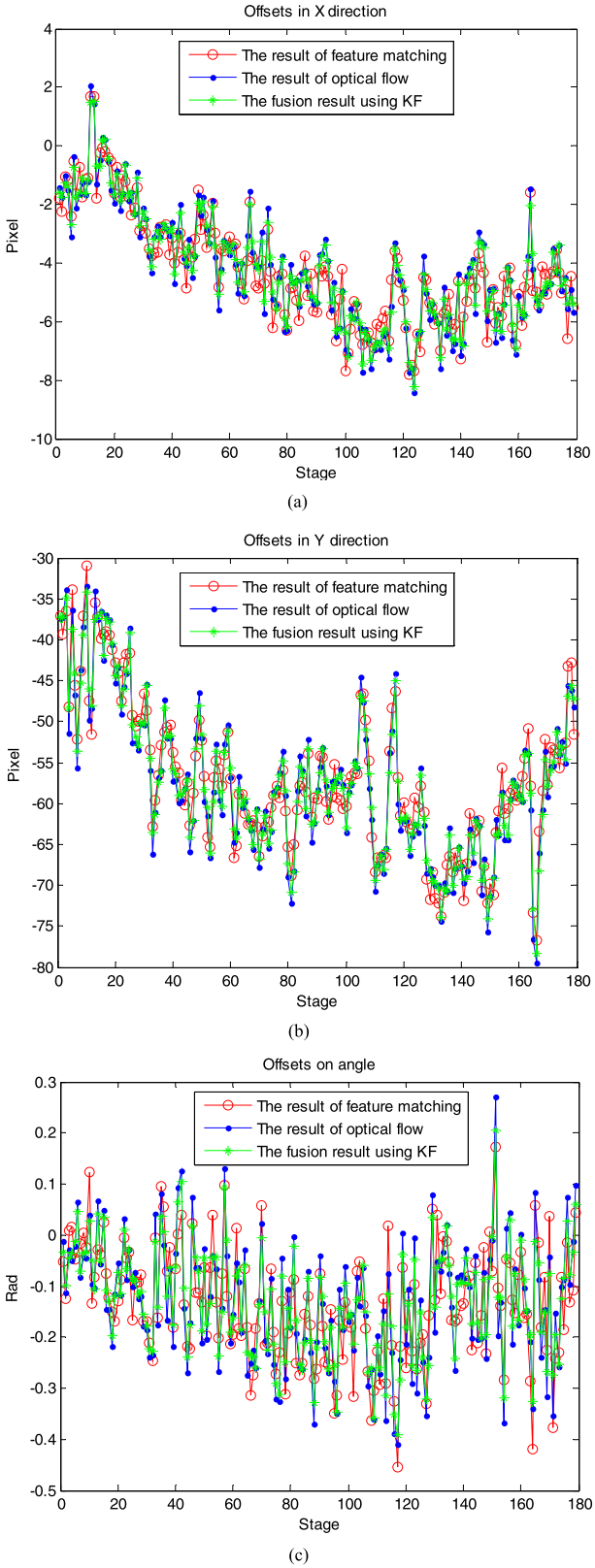


FIGURE 10. Vehicle offsets analysis of three algorithms: (a) *dx*; (b) *dy*; (c) $d\theta$.

(called one stage). Meanwhile, in every stage, the matching result $[t_dx, t_dy, t_d\theta]$ of the first frame and the $10th$ frame using feature matching method can be calculated.

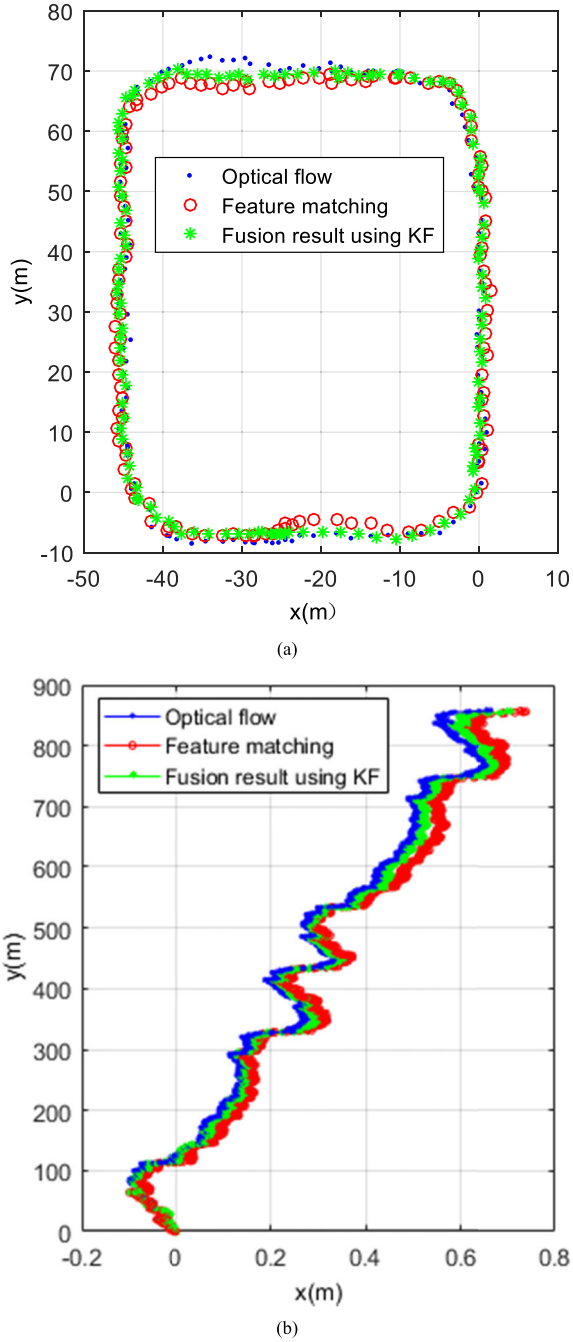


FIGURE 11. Experiment results of three algorithms: (a) circular trajectory; (b) straight trajectory.

According to the previous experiments and literatures, we can easily know that the error of the optical flow method is higher than feature points matching method. Therefore, in the fusion process based on the discrete kalman filter, the parameters of process noise covariance matrix Q and observation noise covariance matrix R are relatively set to 0.1 and 0.05. The optimized result is $[dx, dy, d\theta]$, as shown in Fig. 10, improving the offsets' accuracy. Because feature points matching needs a longer time, after a completion, a few frames have been computed by optical flow. Therefore, the fusion result will be used as the initial value to update the next stage of

computing. Finally, this fusion process is repeated until all images are calculated.

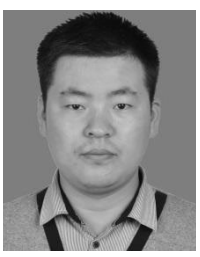
The study is used to test vehicle trajectories, which is shown in Fig. 11. For straight trajectory estimation, the optical flow-based method can obtain more accurate results. However, for the bend, it will make the angle error larger, and the cumulative error will be increased constantly. Large scale scene moving can be conducted by the positioning algorithm based on feature points matching, which has low requirements for illumination change. When the shortage and mismatching of feature points caused by the insufficient of the ground texture appear, the accuracy will be influenced. Therefore, the optical flow method and the feature point matching algorithm are fused by kalman filter to reduce the error and make the curve smooth. When lacking ground texture, vehicle ego-localization can be realized with the help of optical flow. For large illumination change, feature matching-based method is mainly used. Two algorithms complement each other, and the fusion results can provide the more accurate location information and more abundant details, to improve the robustness of vehicle positioning system.

IV. CONCLUSION

In order to ensure better real-time performance and higher accuracy of vehicle positioning, the paper puts forward a vehicle motion estimation method based on the fusion of optical flow and feature points matching. The novel FAST algorithm with self-adaptive threshold enhances the stability and adaptability of the feature points, suppressing the phenomenon of multiple feature point block. The custom LARSAE has a good effect for vehicle position estimation based on optical flow. These two kinds of methods are fused by using kalman filter, to overcome the low accuracy of optical flow and the long processing time of feature matching. On the hardware, using monocular camera reduces the requirement for hardware devices, and is easy to implement. Experimental results show that the study can achieve higher positioning accuracy in real time, and has a better result for ego-localization in the flat pavement. Under the circumstances such as illumination change and low pavement texture, it proves to have more precise and stronger noise resistance, and can be very useful for the development of odometer with quick visual.

REFERENCES

[1] A. de la Escalera, E. Izquierdo, B. Musleh, F. García, J. M. Armingol, and D. Martín, "Stereo visual odometry in urban environments based on detecting ground features," *Robot. Auto. Syst.*, vol. 80, pp. 1–10, Jun. 2016.
[2] S. Lovegrove, "Parametric dense visual SLAM," Ph.D. dissertation, Imperial College London, London, U.K., 2012.
[3] H. Shi, L. Yan, B. Liu, and J. Zhu, "GPS/SINS integrated navigation algorithm aided by scene matching," *J. Tsinghua Univ., Sci. Technol.*, vol. 48, no. 7, pp. 1182–1185, 2008.
[4] X. Song, Z. Song, K. Althoefer, and L. D. Seneviratne, "Optical flow-based slip and velocity estimation technique for unmanned skid-steered vehicles," in *Proc. Intell. Robots Syst.*, Nice, France, 2008, pp. 101–106.
[5] H. Uchiyama, D. Deguchi, I. Ide, H. Murase, and T. Takahashi, "Ego-localization using streetscape image sequences from in-vehicle cameras," in *Proc. Intell. Vehicles Symp.*, Xi'an, China, 2009, pp. 185–190.
[6] T. Wu and A. Ranganathan, "Vehicle localization using road markings," in *Proc. Intell. Vehicles Symp. (IV)*, Gold Coast, QLD, Australia, 2013, pp. 1185–1190.
[7] J. Zhou, X. Zhao, Z. Xu, and Z. Zhou, "Algorithm of vehicle trajectory extraction using fusion of Hu moments and grayscale feature," *Inf. Technol. J.*, vol. 13, no. 11, pp. 1770–1778, 2014.
[8] Y. Gao, "Pose estimation of intelligent vehicles based on monocular visual geometry," M.S. thesis, Dept. Control Sci. Eng., Zhejiang Univ., Hangzhou, China, 2018.
[9] F. Zhang, G. Chen, H. Stähle, C. Buckl, and A. Knoll, "Visual odometry based on random finite set statistics in urban environment," in *Proc. Intell. Vehicles Symp. (IV)*, Alcala de Henares, Spain, 2012, pp. 69–74.
[10] F. Zhang, H. Stähle, A. Gaschler, C. Buckl, and A. Knoll, "Single camera visual odometry based on Random Finite Set Statistics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vilamoura, Portugal, Oct. 2012, pp. 559–566.
[11] F. Zhang, H. Stähle, G. Chen, C. Buckl, A. Knoll, and C. C. C. Simon, "A sensor fusion approach for localization with cumulative error elimination," in *Proc. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Hamburg, Germany, 2012, pp. 1–6.
[12] O. Pink, F. Moosmann, and A. Bachmann, "Visual features for vehicle localization and ego-motion estimation," in *Proc. Intell. Vehicles Symp.*, Xi'an, China, 2009, pp. 254–260.
[13] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, 2010, pp. 2432–2439.
[14] Y. Ban, "Vision-based mobile robot navigation based on optical flow," M.S. thesis, Nanjing Univ. Sci. Technol., Jiangsu, China, 2012.
[15] S. Lovegrove, A. J. Davison, and J. Ibañez-Guzmán, "Accurate visual odometry from a rear parking camera," in *Proc. Intell. Vehicles Symp. (IV)*, Baden-Baden, Germany, 2011, pp. 788–793.
[16] C. Zheng, Z.-Y. Xiang, and J.-L. Liu, "Monocular vision odometry based on the fusion of optical flow and feature points matching," *J. Zhejiang Univ., Eng. Sci.*, vol. 48, no. 2, pp. 279–284, 2014.
[17] C. Zheng, "The research of monocular vision odometry based on optical flow," M.S. thesis, Zhejiang Univ., Hangzhou, China, 2013.
[18] C. Pan, H. Deng, X. F. Yin, and J. G. Liu, "An optical flow-based integrated navigation system inspired by insect vision," *Biological*, vol. 105, nos. 3–4, pp. 239–252, 2011.
[19] C. G. Harris and M. M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, Manchester, U.K., Jan. 1988, pp. 147–151.
[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[21] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[22] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010.
[23] J.-Y. Bouguet, "Pyramidal implementation of the Lucas–Kanade feature tracker: Description of the algorithm, OpenCV documentation," Microprocessor Res. Lab., Santa Clara, CA, USA, Intel Corp., 1999.
[24] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, Jan. 1999, pp. 666–673.
[25] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.



CHENG XIN received the B.S. degree in software engineering and the M.S. and Ph.D. degrees in traffic information engineering and control from Chang'an University, Xi'an, China, in 2012, 2014, and 2017, respectively. In 2017, he joined Chang'an University, where he has been an Assistant Professor, since 2019. His current interests include artificial intelligence and vehicle networking technology, high-precision positioning and environment perception for connected and automated vehicles, traffic big data analysis, and information fusion.

ZHOU JINGMEI received the B.S. degree in software engineering and the Ph.D. degree in traffic information engineering and control from Chang'an University, Xi'an, China, in 2012 and 2017, respectively. In 2017, she joined Chang'an University, where she has been an Assistant Professor, since 2019. Her current interests include high-precision positioning and environment perception for connected and automated vehicles, image recognition, and information fusion.

ZHAO XIANGMO received the B.S. degree from Chongqing University, China, in 1987, and the M.S. and Ph.D. degrees from Chang'an University, China, in 2002 and 2005, respectively. He is currently a Professor and the Vice President with Chang'an University. He has authored or coauthored over 130 publications and received many technical awards for his contribution to the research and development of intelligent transportation systems. His research interests include intelligent transportation systems, distributed computer networks, wireless communications, and signal processing.

WANG HONGFEI received the B.E. degree in software engineering from Chang'an University, Xi'an, China, in 2019, where he is currently pursuing the M.E. degree in software engineering with the School of Information. His current research interests include deep learning and traffic sign detection.

CHANG HUI received the B.E. degree in computer science and technology from Chang'an University, Xi'an, China, in 2019, where she is currently pursuing the M.E. degree in computer science and technology with the School of Information. Her current research interests include traffic flow and tidal lanes setting.