A two-step abnormal data analysis and processing method for millimetre-wave radar in traffic flow detection applications

Haiqing Liu | Kunmin Teng | Laxmisha Rai | Yu Zhang | Shengli Wang

1 College of Transportation, Shandong University of Science and Technology, Qingdao, China
2 College of Electronic and Information Engineering, Shandong University of Science and Technology, Shandong, China
3 Ocean Science and Engineering College, Shandong University of Science and Technology, Qingdao, China

Correspondence
Shengli Wang, Ocean Science and Engineering College, Shandong University of Science and Technology, Qingdao, China.
Email: shlwang@sduet.edu.cn

Funding information
Shandong Provincial Natural Science Foundation, Grant/Award Number: ZR2019QH017; Qingdao Municipal Science and Technology Commission, Grant/Award Number: 19-6-2-3-cg

Abstract
In a new application scenario where the millimetre-wave radar is installed above the road for detecting traffic flow in downward looking direction, the original data of the radar includes all kinds of background noises and false targets. In order to acquire effective vehicle trajectories, a two-step abnormal data processing method for millimetre-wave radar in traffic flow detection application is proposed. In the first step, the rational range of distance, angle and speed are studied, and proper thresholds are presented for reducing the samples of which the single parameter is with the obvious abnormality. Moreover, the nearest neighbour analysis method is used to further extract vehicle trajectories based on the similarity and slope characteristics of each sample to its neighbours. Taking actual detected data as samples, the weighting coefficients, similarity threshold, average slope threshold and standard deviation threshold are calibrated for the proposed nearest neighbour analysis method. The two-step processing method presents a higher performance in extracting effective trajectory samples, and the ratio of noise points is reduced to 4.1%, compared with 239.9% in the original data sample. The proposed method can provide an effective reference for further applications, such as driving behaviour analysis and traffic flow parameter identification.

1 INTRODUCTION

Real-time road traffic parameter collection and accurate road congestion evaluation are the prerequisites for applying better traffic congestion avoidance strategies to improve the traffic flow [1–3]. Among all kinds of traffic detection methods, the radar-based method has certain advantages. For example, there is no need to destroy roads when deploying the devices, compared with the embedded detectors, such as loop and geomagnetic detectors [4, 5]. Besides, the radar is highly accurate when detecting the vehicle speed and minimally affected by the environmental fluctuations, compared to video-based devices [6, 7]. All these advantages make the radar widely used in urban traffic and expressway management, such as the traffic volume collection and over speed detection [8–10].

However, the mainstream radar devices used in traffic flow detection are generally achieved in centimetre, decimetre or metre waves, which work at K and X frequency bands [11–13]. These kinds of radars are insensitive to low-speed objects, and they are unable to detect stationary vehicles. The millimetre-wave radar generally works at a 30–300 GHz frequency band and the wavelength is from 1 to 10 mm. Since the wavelength is between the centimetre and light waves, it has the advantages of integrating both the microwave and photoelectric characteristics. Especially, the frequency-modulated continuous wave (FMCW) millimetre-wave radar has been used in the road traffic environment, as it has significant advantages in high-speed measurement precision including the low-speed targets, high multi-objective resolution capacity and imaging quality, wide detection range and strong anti-interference capability [14]. Besides, millimetre-wave radar also has the capability of detecting stationary objects by identifying radar cross section (RCS) energy information. All these characteristics suitably compensate for the drawbacks of traditional microwave radar but also bring noises and false targets from the surroundings.

Currently, the millimetre-wave radar is mainly used in target identification and trajectory tracking for perceiving driving environment in advanced driver assistance systems and unmanned
aerial vehicles. In [15, 16], the authors proposed a blind spot detection and warning system using millimetre-wave radar. The FMCW millimetre-wave radar system is used to monitor the moving targets, which are in the blind spot warning area behind a certain vehicle, maintaining higher detection and lower false detection rate. In [17], the authors presented work on information processing of millimetre-wave radar obstacle in self-driving car environment sensing module. Considering tracking, a method has been added to provide edge information by surrounding obstacles to determine the direction of travel, improving the accuracy and stability of the algorithm. In [18], a fast-square root cubature kalman filter (CKF) method for automotive millimetre-wave radar target tracking is proposed. In [19], to solve the problems such as false alarm and missing detection in the FMCW radar, an improved multi-target detection algorithm is proposed combining the phase and power differences for frequency matching. The corresponding research mainly focused on target tracking and simultaneously reducing noises and false targets by echo signal optimisation or filtering method.

In order to overcome the defects of a single sensor and improve the target recognition rate, the millimetre-wave radar is also used for data fusion with other sensors, mainly vision-based detectors. In [20], the authors use the millimetre-wave radar to detect targets on road and transmit the location and size of the region of interest (ROI) to image sequences captured by the monocular camera. Using the active contour method, false targets are recognised when there is no vehicle in the ROI. The proposed fusion method can achieve a 0% false alarm rate for on-board under a real-world dataset. In [21], a multi-sensor fusion algorithm based on a centralised fusion strategy where the fusion centre takes a unified track management is proposed. The multi-target tracking with the prediction of current tracks is performed according to the range and the azimuth angles of targets obtained by vision sensor and radar, respectively. This work can also be used to help to identify the effective samples with the noises. Similar studies are also presented in [22]. In [23], in order to discriminate a vehicle from a non-vehicle target, radar detection is projected to the camera image for target width estimation. This strategy helps to recognise false targets in the road environment and significantly improves the accuracy of the radar detection. Since the vision sensor has high accuracy on the detection of azimuth angle, the multi-sensor fusion is an effective method for avoiding the background noise. However, this strategy certainly induces computational complexity.

There are substantial differences between the data collected from the on-board installation case and the on-road installation case. For the former, there are speed values of relative movement between all detecting targets and the on-board unit. The Doppler effect can fully play its role, and the effective vehicle information can be extracted referring to the speed characteristics. For the latter, the situation is more complicated. That is because the high level of background noises and false targets are embedded in the effective vehicle trajectories, and they cannot be identified by a single parameter. The abnormal data mainly comes from two resources: (1) The background noise of the radar device, such as the side-lobe interferes [24, 25], and (2) the false targets caused by the non-vehicle objects in the road conditions, such as road surface, guardrails, lamp posts and others. In [26], the authors use the FMCW millimetre-wave radar for the on-road detection scenario and analyse the data features collected from the radar. However, the research only gives a preliminary presentation of the data’s statistical characteristics. Further analysis and applications, such as abnormal data processing and traffic event identification are not elaborated in detail.

As elaborated in the aforementioned literature, limited research has been carried out regarding the installation of the millimetre-wave radar above the road for detecting road traffic flow towards the downward direction. Aiming at this new application scenario, a data analysis and processing method is proposed in this paper. Based on the data features acquired in [26], a two-step data processing method is proposed, namely, threshold analysis for the first step and nearest neighbour analysis [27, 28] for the second step. In the first step, the rational range of distance, angle and speed are analysed, respectively. By setting proper thresholds, obvious abnormal data from the original data are rejected so that the data volume can be reduced. In the second step, data samples that have similar moving and slope characteristics with their neighbours are clustered together to extract the vehicle trajectories. Experimental study shows that the two-step processing method presents a better performance in extracting effective trajectory samples. This accomplishment can provide an effective reference for further applications, such as driving behaviour analysis and traffic flow parameter identification.

The rest of the paper is organised as follows. Section 2 introduces the experimental scenario and shows the example of the original data. In Sections 3 and 4, the two-step data preprocessing method is proposed, namely, the threshold analysis for the first step and nearest neighbour analysis for the second step, respectively. In Section 5, the proposed method is verified based on the actual road traffic data. Finally, in Section 6, we conclude the paper and provide discussions for future work.

2 | INTRODUCTION TO EXPERIMENTAL SCENARIO AND ORIGINAL DATA PRESENTATION

To collect traffic data for further analysis, we have conducted the experiment in an actual traffic scenario. The device is installed 7.5 m above the experimental road, and the system hardware architecture is shown in Figure 1. The detailed parameters of the installation and the road section are introduced in Table 1. The radar used in this study is an improved version in terms of signal quality and radiation pattern, compared to the device used in [26]. Compared with the radar in the previous work, the planar microstrip array antenna contains two groups for transmission and four groups for receiving; it can detect a much farther target, and the angle value is more stable. The radar device detects the information of the target vehicle, including the distance, angle, speed and the RCS energy per
FIGURE 1  Experimental hardware architecture

TABLE 1  Parameters of the installation environment and the road section

| Parameter                   | Value |
|-----------------------------|-------|
| Height of the radar         | 7.5 m |
| Pitch angle of the radar surface | 12°  |
| Number of lanes             | 3     |
| Lane width                  | 3.5 m |

50 ms outputting the data. The parameters are introduced as follows:

1. Distance (m): The distance between the vehicle and radar.
2. Speed (m/s): The relative velocity of the vehicle target.
3. Angle: The horizontal angle between the target and the vertical detecting direction of the radar.
4. RCS energy (dB): The signal strength reflected by the electromagnetic reflection capability of the target and the signal level attenuation along the electromagnetic wave transmission path.

The detailed technical parameters of the radar device are introduced in Table 2.

TABLE 2  Technical parameters of the millimetre-wave radar

| Parameter                   | Value                  |
|-----------------------------|------------------------|
| Frequency range             | 76–77 GHz              |
| Modulation mode             | Frequency-modulated    |
| Distance measuring range    | 1–150 m                |
| Distance error precision    | ±0.15 m                |
| Velocity measurement range  | −70~70 m/s             |
| Velocity error precision    | ±0.5 m/s               |
| Angle measurement range     | ±15°                   |
| Angle error precision       | <1°                    |
| Maximum measurement target numbers | 32                 |
| Detection period            | 50 ms                  |

Under the experimental scenario described above, the original data of the road traffic flow are collected. Part of the sampling data are selected and presented in Figure 2 following the two principles: (1) The samples should not last too long for the purpose of clear presentation of the results. (2) There are enough points for the verification of the proposed method. In the figure, the continuous sloping lines are the effective vehicle trajectories including seven vehicles, while others are all background noises. It is hard to identify the original data with trajectory points that embed the noises within. To solve this problem, we present a two-step strategy to reject them.

3  ABNORMAL DATA PREPROCESSING METHOD BASED ON THRESHOLD ANALYSIS

In the research scenario, the detection parameters, mainly the distance, angle and speed, are within a certain range according to the radar device itself, similar to the installation condition and road environment. In this section, the abnormal data identification and rejection method based on the distance, angle and speed thresholds is introduced.

3.1  Distance threshold analysis

Referring to [26], when the radar device is installed above the road, the distance values of vehicles are distributed between $L_{\text{min}}$ and $L_{\text{max}}$ as described by Equations (1) and (2) in detail:

$$L_{\text{min}} = \frac{H - b_i}{\cos(90 - \theta - \delta)}$$

$$L_{\text{max}} = \frac{H}{\cos(90 - \theta + \delta)}$$

where $H$ is the height of the radar device from the ground, $\theta$ is the pitch angle of the radar device, $\delta$ is the pitch angle of the radar signal, and $b_i$ is the height of the vehicle.

Hence, the effective range of distance is expressed by $[L_{\text{min}}, L_{\text{max}}]$. Data samples that are not within a certain scope is considered abnormal and should be rejected.

3.2  Angle threshold analysis

In the traffic flow detection scenario, the angle measurement can be shown in Figure 3.

In Figure 3, $\alpha$ is the angle of the target vehicle. $\alpha_{\text{max}}$ is the maximum and $-\alpha_{\text{max}}$ is the minimum measurement angles. From Figure 3, it is evident that the detecting angle gets a maximum value at a certain distance when the vehicle travels along the outside line of the road. The top and bottom values of the detecting angle in the case scenario is calculated by Equation (3).

Besides, the detecting angle values are also confined by the angle measurement range of the radar device presented in Table 2 as
shown in Equation (4):

\[
\begin{align*}
\alpha_{\text{top}} &= \arcsin\left(\frac{d}{L}\right) + \tau \\
\alpha_{\text{bottom}} &= -\arcsin\left(\frac{d}{L}\right) - \tau \\
\end{align*}
\]

\begin{align*}
\alpha_{\text{top}} &\leq \alpha_{\text{max}} \\
\alpha_{\text{bottom}} &\geq -\alpha_{\text{max}}
\end{align*}

In Equation (3), \(d\) is the distance from the edge of the lane in the radar coverage area to the central vertical line of the radar antenna surface. \(L\) is the detecting distance of the data sample. \(\tau\) is the systematic error of the angle of the radar device. We assume that the parameter \(\tau\) follows some mathematical distribution in the detecting range values in \([-\tau_{\text{max}}, \tau_{\text{max}}]\), where \(\tau_{\text{max}}\) denotes the maximum angle error of the radar device. In [26], the authors use the normal distribution function to describe the probability distribution characteristics. Since this systematic error is caused by the radar device itself that cannot be eliminated, we set \(\tau\) as \(\tau_{\text{max}}\) to indistinguishably cover all the possible detecting values, avoiding excessive elimination of the effective data.

Referring to Equations (3) and (4), for the experimental scenario in Section 2, the variation of detecting angle values under different distances is shown in Figure 4.

In Figure 4, the effective angle top and bottom boundaries of the target show convergence tendency when it moves farther away from the radar. Due to the systematic error caused by the radar device, the two boundaries present a certain confidence level for describing the effective angle of the target. In this paper, we define the effective confidence level as the proportion of \(\tau_{\text{max}}\) with the boundary value by Equation (5) and is also shown in Figure 5:
It is evident from Figure 4 that the angular values decline to the minimum value as the target moves farther from the radar device. Considering this, the systematic error makes an ever-increasing effect on the accuracy of the angle value of the target as shown in Figure 5. At a far distance, the angle value is not accurate enough to locate the horizontal position of the target. On the contrary, we can also make a reasonable assumption that the angle value is feasible if the target is close enough to the radar. Under a certain confidence level $p^\dagger$, the angle information can be used to recognize abnormal data.

Based on the above analysis, the effective range of the angle is expressed by $[\max(\alpha_{\text{bottom}}(L), -\alpha_{\text{max}}), \min(\alpha_{n_{\text{np}}}(L), \alpha_{\text{max}})]$ a certain confidence level $p_1$. Data samples that are not in a certain scope are considered abnormal and should be rejected.

### 3.3 Speed threshold analysis

In this paper, the experiment mainly focuses on moving vehicles. The background noises caused by fixed metal objects on road, such as guardrails, lamp posts and other sensitive objects, are with distinct characteristics. Moreover, the distance changes in a limited scope while the speed remains zero. In the experimental scenario introduced in Section 2, there is no vehicle parking on the road, and the speed limit is 60 km/h. Considering the 10 km/h allowance for the abnormal driving behaviour where some drivers may ignore the speed limit and take some over-speed operations, 70 km/h is taken as the speed threshold. The number of data samples under different speed values are shown in Figure 6.

From Figure 6, it is evident that the data samples with zero speed present a huge set of values. Since the millimetre-wave radar is sensitive enough to capture a low-speed moving target, the zero speed targets are all caused by the background noises. Besides, when the speed is higher than 19.4 m/s (70 km/h), the samples are also false targets caused by the background noises. As a conclusion, the data samples with zero speed or higher than the maximum speed threshold are considered abnormal and should be rejected.

### 4 Abnormal data preprocessing method based on nearest neighbor analysis

From data preprocessing based on threshold analysis presented in Section 3, it is obvious that abnormal data cannot be completely eliminated from the original data. There are still some data samples affected by the background noises of radar, and simultaneously the distance, speed and angle values are in the rational threshold ranges. As shown in Figure 2, to some extent, the trajectory of the vehicle is regular and uninterrupted, compared with surrounding noises. In this section, we use the nearest neighbour analysis method to further recognize the effective trajectory and eliminate the abnormal data.

In the original radar data shown in Figure 2, the effective vehicle trajectory points are consecutively ordered in the time-distance graph. This characteristic can be described by the Euclidean distance for neighbour sampling points. We set the point $p$ as the datum point. The distance of point $p$ and $q$ in the time-distance graph is calculated by Equation (6):

$$
dist(p, q) = \sqrt{(t_p - t_q)^2 + (L_p - L_q)^2}
$$

Based on the relative time-distance position relationship, we present the definition of $k$-nearest neighbours of point $p$ for the data preprocessing of millimetre radar detecting data as follows.

### 4.1 $k$-nearest neighbours

The $k$-nearest neighbours of point $p$ denote the set of neighbour points that are in a certain spatial distance in the time-distance graph as shown in Equation (7):

$$
N_k(p) = \{p_1, p_1, ..., p_k\} \ni R, \forall q \in R - \{p_1, p_1, ..., p_k\},
$$

$$
dist(p, q) \geq dist(p, p_1), 1 \leq i \leq k
$$
In the \(k\)-nearest neighbours, two kinds of points are included: Noise and vehicle trajectory points. Constrained by the vehicle moving properties, the trajectory points are generally similar in the detecting parameters. Further, we present Equation (8) to calculate the similarity of \(p\) and \(q\).

\[
sim(p, q) = \max \delta f - \delta f(p, q) \quad (8)
\]

where \(\delta f(p, q)\) is the dissimilarity degree, which denotes the weighted difference of the detecting parameters that are the speed, angle and RCS energy as shown in Equation (9):

\[
\delta f(p, q) = \mu_v |v_q' - v_p| + \mu_a |a_q' - a_p| + \mu_r |r_q' - r_p| \quad (9)
\]

where \(v_p, a_p\) and \(r_p\) denote the dimensionless speed, angle and RCS energy values for point \(p\), respectively. \(v_q', a_q'\) and \(r_q'\) denote the dimensionless forecast values based on the moving parameters at the point \(p\), respectively. \(\mu_v, \mu_a\) and \(\mu_r\) are the weighting coefficients.

In Equation (9), we assume that the vehicle moves at uniform variable motion in a short time. The forecast values can be calculated by Equation (10):

\[
\begin{align*}
    v_q' &= v_p + \frac{2(L_q - L_p - r_p(t_q - t_p))}{t_q - t_p} \\
    a_q' &= \arcsin \left( \frac{L_p \sin a_p}{L_q} \right) \\
    r_q' &= r_p
\end{align*} \quad (10)
\]

In Equation (9),

\[
\mu_v + \mu_a + \mu_r = 1 \quad (11)
\]

The valuing of \(\mu_v, \mu_a\) and \(\mu_r\) are associated with variation features of the speed, angle and RCS energy. They are calibrated by the ratio of mean value and standard deviation of definite sampling points in a short time as shown in Equations (12) and (13):

\[
\begin{align*}
    \mu_v'(i) &= \frac{|\bar{v}_i - \bar{v}_w|}{\sigma_v'} \\
    \mu_a'(i) &= \frac{|\bar{a}_i - \bar{a}_w|}{\sigma_a'} \\
    \mu_r'(i) &= \frac{|\bar{r}_i - \bar{r}_w|}{\sigma_r'} \quad (12)
\end{align*}
\]

\[
\mu_{v,a,r} = \frac{\mu_v' + \mu_a' + \mu_r'}{\mu_v + \mu_a + \mu_r} \quad (13)
\]

In Equation (12), \(\bar{v}_i, \bar{a}_i, \bar{r}_i\) and \(\bar{v}_w\) are the average speed, angle and RCS energy of the definite \(i\)-th vehicle, respectively, in a certain short time. Similarly, \(\sigma_v', \sigma_a'\) and \(\sigma_r'\) are the standard deviations. \(\bar{v}_w, \bar{a}_w, \bar{r}_w\) are the average values of the sampling points except for the definite \(i\)-th vehicle in the same sampling time period. The numerator of each sub-equation in Equation (12) represents the difference of parameters between the vehicle points and others. A high numerator value implies that the feature is obvious, and the parameter is significant for identifying the vehicle. The denominator represents the variability of parameter values of the same vehicle in different sampling frames. A low denominator value implies that the parameter changes steadily, and it is worthy for trajectory tracking.

In Equation (8), \(\max \delta f\) is the maximum dissimilarity degree of any two points in the original samples as shown in Equation (14):

\[
\max \delta f = \max \{\delta f(p, q) | \forall p, q \in R^3\} \quad (14)
\]

By setting a certain minimised threshold of the similarity degree for any two points in the original samples, the vehicle trajectory can be effectively recognised. However, some special noise points caused by the radar device also have similar
parameters. The similarity degree is not sufficient to recognise this type of noise. In order to get a more accurate vehicle trajectory, we further analyse the slope features and find some potential new characteristics.

In the time-distance graph built by the sampling time and the target distance, the average slope of the total vehicle trajectory is positively correlated with the vehicle speed. Figure 7(a) presents sections of the actual original data samples. As shown in the figure, the slope of the vehicle trajectory is fairly constant in a short time. For the effective trajectory points, the slope of the datum point to its neighbours present certain low values (which is positively correlated with the vehicle speed) and do not vary greatly, compared with the average slope of the total vehicle trajectory as shown in Figure 7(b). While for the background noise points, the slope values change greatly because the noise points generally deviate from the vehicle trajectory as shown in Figure 7(c).

Based on the analysis provided in this paper, the average value and the average variance of the slope of the datum point to its neighbours are referred to distinguish the effective vehicle trajectory points and the background noise. The average slope value and the standard deviation are calculated by Equations (15) and (16), respectively:

\[
K(p) = \frac{1}{k} \sum_{i=1}^{k} \frac{I_{p_i} - I_p}{t_{p_i} - t_p}
\]  
(15)

\[
S(p) = \sqrt{\frac{1}{k} \sum_{i=1}^{k} \left( \frac{I_{p_i} - I_p}{t_{p_i} - t_p} - K(p) \right)^2}
\]  
(16)

4.2 Core point

The concept of the core point of the proposed method is provided as follows.

For each \( p \in N_k(p) \), if \( p \) is the core point, it should satisfy the equation shown in Equation (17):

\[
\begin{cases}
K(p) \leq K_T \\
S(p) \leq S_T
\end{cases}
\]  
(17)
In the definition, \( N'_k(p) \) is the set of neighbours that satisfies the following constraint:

\[
N'_k(p) = \{p_1, p_2, \ldots, p_k | p \in N_k(p), sim(p, p_i) > sim, 1 \leq i \leq k \}
\]  

(18)

In Equation (17), \( K_T \) is the average slope threshold, and \( \sigma \) is the standard deviation threshold.

### 4.3 Direct density-reachable

Based on the concept of the core point, the direct density-reachable relationship is given as follows.

If \( p \) and \( q \) are core points and satisfy the equation shown in Equation (19), they are defined as direct density-reachable:

\[
N_k(p) \cap N_k(q) \neq \emptyset
\]  

(19)

Equation (19) implies that two core points are direct density-reachable when they share the same points in their \( k \)-nearest neighbours.

In this paper, the core points are the effective vehicle trajectory sampling points that should be recognised, while others are background noise and should be rejected. The vehicle trajectory can be formed by gathering all the core points together based on the direct density-reachable relationship. The details of the data preprocessing method are presented as follows:

1. **Step 1.** Find the \( k \)-nearest neighbours for each sampling data point based on a certain Euclidean spatial distance.
2. **Step 2.** Extract the possible effective vehicle samples according to the similarity degree for any two points in the \( k \)-nearest neighbours.
3. **Step 3.** Identify the core points based on the slope features between the datum point to its neighbours in the acquired possible effective vehicle samples obtained in step 2.
4. **Step 4.** Carry out the clustering operation for the core points to identify the vehicle trajectory using the breadth-first search algorithm based on the direct density-reachable relationship.
5. **Step 5.** Reject the background noise points that do not belong to the identified vehicle trajectory, and complete the data preprocessing operation.

### 5 EXPERIMENTAL RESULTS AND ANALYSIS

Using the original data shown in Figure 2, the two-step abnormal data processing method is verified, and the results are analysed as follows:

#### 5.1 Experimental results based on threshold analysis

Based on the installation parameters, road conditions and configurations of the radar device, the rational detecting range is calculated, and samples beyond the effective range are rejected. The results after the distance threshold analysis are shown in Figure 8(b). Further, results after speed threshold analysis are shown in Figure 8(c), and the results after angle threshold analysis are shown in Figure 8(d).

In addition, we use the ratio of noise points to evaluate the performance of results after different steps of the threshold analysis. The index is calculated by Equation (20):

\[
\text{Ratio} = \frac{N_n}{N_e} \times 100\%
\]

(20)

In Equation (20), \( N_n \) is the number of noise points, and \( N_e \) is the number of effective vehicle trajectory points. In this paper, \( N_e \) is acquired by manual analysis. The ratio of the noise points after the steps of threshold analysis is given in Table 3.

It is evident from Table 3 that after each step of threshold analysis, certain abnormal sampling points are rejected from the original data, and the ratio of noise points decreases. Especially, speed threshold analysis presents the most significant effect because major background noises are caused by stationary targets in the road environment. The threshold analysis method rejects most of the background noises at the expense of bringing a low complexity to the system. This definitely improves the efficiency of further analysis. However, there are still 38.7% noise points after three steps of threshold analysis. These noise points are mainly caused by the radar device, and the data features are not obvious enough for rejection of the threshold analysis method.

#### 5.2 Experimental results based on nearest neighbour analysis

##### 5.2.1 Calibration of the weighting coefficients \( \mu_s \), \( \mu_a \) and \( \mu_r \)

In order to acquire the definition samples, the effective vehicle trajectory data of seven vehicles are extracted from the original data by manual analysis method. The ratio of the mean value and the standard deviation for each vehicle and the final weight coefficient values are presented in Table 4.
In Table 4, the results show that the weight coefficient value of speed is the biggest (0.64), following is the angle (0.28) and RCS energy is the lowest (0.08). We make qualitative analysis on the original data, and the results are in accordance with the actual situation. Moreover, the speed measurement of the target is so precise that the samples can form continuous features for adjacent frames. Besides, for a definite vehicle, the variation of the speed values are not significant in different frames in a short time. Hence, the speed value is the most important parameter for tracking the vehicle and identifying the vehicle from the background noise. For angle, data jumping occurs around the actual moving trajectory because of the detecting error. This reduces the reliability of the parameter to describe the vehicle moving characteristics. For the RCS energy, it varies in the range of (80, 110), even for different samples of the same vehicle. There is no substantial difference between the vehicle point and others. This parameter is not obvious enough to distinguish the target vehicle and other noise points. Hence, it should be assigned a low weight value.

5.2.2 Calibration of the similarity threshold

Based on the definite trajectory of vehicles, the similarity value is calculated referring to Equation (8). In order to calibrate the similarity threshold, two kinds of similarity values are statistically analysed, namely, (1) similarity values between any two vehicle points and (2) others, including values between noise point and vehicle point, and values between any two noise points. When the number of nearest neighbours is set to 20, the distribution of the two kinds of values is shown in Figure 9.

In Figure 9, the left part of the figure presents the variation similarity values from the largest to smallest for different samples. Combined with the box-plot shown in the right part of the figure, it is evident that the similarity between two vehicle points presents a high average value and changes smoothly (with a standard deviation of 0.2). However, for other cases, the distribution is discrete (with a standard deviation of 0.98), and most of the samples are located below 3.20. This verifies that it is effective and reasonable for the proposed method to describe the moving properties of the vehicle. This characteristic is obvious enough to select a proper similarity threshold.
to distinguish the effective vehicle points and the background noise points. In this paper, the bottom boundary of the box-plot is taken as the similarity threshold to filter effective neighbours from the original neighbour set established by the Euclidean distance.

For a certain sample point, neighbours with larger similarity than the obtained threshold are considered effective neighbours, while others are removed from the neighbour set. We conducted a statistical analysis of the number of neighbours for the effective vehicle trajectory points, and the results are shown in Figure 10. For the case samples, all the vehicle points include at least six neighbours. This is because the vehicle trajectory points are consecutive and densely distributed, compared with the noise. The number of neighbours under the similarity threshold constraint can also be used as an index to select core points for further nearest neighbour analysis.

5.2.3 Calibration of average slope threshold $K_T$ and the standard deviation threshold $S_T$

By selecting effective neighbours based on the similarity threshold as the research object, the average slope and standard deviation for each sample to its neighbours are calculated. The distribution characteristics for the vehicle trajectory and the noise points are presented in Figures 11 and 12, respectively.

In Figure 11, the distribution of slope values for effective vehicle trajectory points is relatively centralised, and the average value is much lower, compared with the noise points. In the box-plot, unusual values above the upper boundary are also significant for vehicle trajectory. In order to reject the noise points as much as possible, and simultaneously, to reduce the mis-rejection of the effective trajectory data, the 98th percentile is taken as the average slope threshold $K_T$. In Figure 12, the average standard deviation distribution for effective vehicle trajectory samples also presents centralised characteristics.
example, as shown in the figure, the 75th percentile for the effective vehicle trajectory points is only 0.95, which is slightly larger than the 25th percentile for the noise points. However, under the upper boundary of noise points, there are still many effective trajectory points that are denoted as unusual values in the box-plot. In order to reduce the mis-rejection of the effective trajectory points, the 98th percentile of the standard deviation is taken as the threshold $ST$.

5.2.4 Final processing results

Final samples after the nearest neighbour analysis are presented in Figure 13.

Referring to the original samples shown in Figure 2, it is evident that vehicle trajectory points are effectively extracted. Compared with the results after threshold analysis shown in Table 4, the ratio of the noise points has been reduced to 4.1%. Since the referenced effective vehicle trajectory points are selected by manual experience, some samples may fail to be recognised as an effective trajectory point unavoidably because of subjective judging error. Hence, the proposed method reduces the actual number of noise points to much less than 4.1%. This conclusion can be verified by an evident phenomenon that most of the noise points (isolated points in the figure) are close to the trajectory or distributed along the extending line. Besides, these samples also have similar characteristics to the vehicle trajectory points in the original data. Hence, the proposed method shows high performance for reducing noise points.

In the final results, there are also 2.2% of effective trajectory points that are mis-rejected. The reason is that the 98th percentile of the average slope and standard deviation for effective vehicle trajectory points are selected as the thresholds to identify core points. Under this condition, small samples of effective vehicle trajectory points are treated as noises and finally rejected. This mis-rejection probability is much lower than the radar inherent frame loss that makes little influence on the identification of vehicle trajectory and further higher application analysis.

6 CONCLUSION AND FUTURE WORK

Limited research has been carried out regarding the installation of the millimetre-wave radar above the road for detecting road traffic flow towards the downward direction. Aiming at this new application scenario, an innovative, two-step process for analysing abnormal data is proposed to reduce background noise. The method combines the advantage of speed and efficiency in reducing the obvious abnormal data for the first threshold analysis step and the advantage of the possibility of extracting the indiscernible abnormal samples by intrinsic similar driving characteristics for the second nearest neighbour analysis step.
The experimental results present obvious features in which there are significant differences between effective vehicle trajectory points and noise points in parameter similarity, average slope and standard deviation of the slope. All these differences provide stable references for identifying and rejecting abnormal data in the application scenario. In conclusion, the proposed method displays higher performance in rejecting the noise points, and the ratio of the noise points has been reduced to 4.1% after two-step processing.

The future work mainly focuses on two aspects: (1) The corresponding thresholds will be calibrated and optimised using a larger sample set so that the final ratios of the noise points and the mis-rejected samples can be further improved. (2) In our work, the radar is tested under good weather conditions. Data collection and verification will be carried out under different weather conditions, such as the wind, rain and snow.

ACKNOWLEDGEMENTS
This research is supported by Shandong Provincial Natural Science Foundation, grant number ZR2019QF017 and Basic Research Plan on Application of Qingdao Science and Technology, grant number 19-6-2-3-cg.

REFERENCES
1. Sanguesa, J.A., et al.: Sensing traffic density combining V2V and V2I wireless communications. Sensors 15, (12), 31794–31810 (2015)
2. Liu, H.Q., et al.: A new approach for real-time traffic delay estimation based on cooperative vehicle-infrastructure system at the signal intersection. Arabian J. Sci. Eng. 44(3), 2613–2625 (2019)
3. Ma, D.F., et al.: Estimating maximum queue length for traffic lane groups using travel times from video-imaging data. IEEE Intell. Transp. Syst. Mag. 10(3), 123–134 (2018)
4. Al-Jameel, H.A., Al-Jumaili, M.A.: Analysis of traffic stream characteristics using loop detector data. Jordan J. Civ. Eng. 10(4), 403–416 (2016)
5. Kwon, Y.J., Kim, D.H., Choi, K.H.: A single node vehicle detection system using an adaptive signal adjustment technique. In: 20th International Symposium on Wireless Personal Multimedia Communications, Yogyakarta, Indonesia (2018)
6. Gao, Z., et al.: Synergizing appearance and motion with low rank representation for vehicle counting and traffic flow analysis. IEEE Trans. Intell. Transp. Syst. 19(8), 2675–2685 (2018)
7. Du, Y.C., et al.: An open data platform for traffic parameters measurement via multirotor unmanned aerial vehicles video. J. Adv. Transp. 2017, 1–12 (2017)
8. Zhao, H., et al.: Method of recognizing background power spectrum used in millimeter wave traffic flow detection radar. J. Infrared Millimeter Waves 27, 437–441 (2008)
9. Miyamoto, K., Nakagawa, Y.: 79 GHz-band millimeter-wave radar high precision and wide field of view technologies. In: Proceedings of Asia-Pacific Microwave Conference, Sendai, Japan, pp. 546–548 (2015)
10. Ho, T.J., Chung, M.J.: An approach to traffic flow detection improvements of non-contact microwave radar detectors. In: 2016 International Conference on Applied System Innovation, Okinawa, Japan, pp. 26–30 (2016)
11. Wang, S., Sharma, A.: Assessing the impact of speed-limit reduction near signalized high-speed intersections equipped with advance-warning flashers. J. Transp. Eng., Part A: Systems 143(6), 1–3 (2017)
12. Jia, Y.Z., Chen, X.W., Zou, Y.L.: A high-precision K-band LFMCW radar for range measurement. In: Selected Papers of the Chinese Society for Optical Engineering Conferences, Changshun, China (2016)
13. Kusmad, Munir, A.: Performance analysis of FMCW-based X-band weather radar. In: 2nd International Conference on Wireless and Telematics, Yogyakarta, Indonesia, pp. 40–43 (2016)
14. Li, C.L., et al.: A noncontact FMCW radar sensor for displacement measurement in structural health monitoring. Sensors 15(4), 7412–7433 (2015)
15. Liu, G.R., Wang, L.L., Zou, S.: A radar-based blind spot detection and warning system for driver assistance. In: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, pp. 2204–2208 (2017)
16. Liu, G.R., et al.: A blind spot detection and warning system based on millimeter wave radar for driver assistance. Optik 135, 335–365 (2017)
17. Song, X.S., Du, Y.: Obstacle tracking based on millimeter wave radar. In: 3rd International Conference on System and Informatics, Shangai, China, pp. 531–535 (2016)
18. Liu, H.J., Lai, S.F.: Fast square root CKF for automotive millimeter-wave radar target tracking. J. Nanjing Univ. Sci. Technol. 40(1), 56–66 (2016)
19. Du, J., Song, C.L.: A modified millimeter-wave radar multi-target detection algorithm. Commun. Technol. 46, 808–813 (2015)
20. Wang, X., et al.: On-road vehicle detection and tracking using MMW radar and monovision fusion. IEEE Trans. Intell. Transp. Syst. 17(7), 2075–2084 (2016)
21. Wu, X., et al.: Study on target tracking based on vision and radar sensor fusion. In: WCX World Congress Experience, Detroit, MI, vol. 1, pp. 6–13 (2018)
22. Oishi, Y., Matsunami, I.: Radar and camera data association algorithm for sensor fusion. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. E100.A(2), 510–514 (2017)
23. Wang, J.G., et al.: Vehicle detection and width estimation in rain by fusing radar and vision. In: 2018 15th International Conference on Control, Automation, Robotics and Vision, Singapore, pp. 1063–1068 (2018)
24. Sabanci, K., et al.: A hue-domain filtering technique for enhancing spatial sampled compressed sensing-based SAR images. IET Radar Sonar Navig. 13(3), 357–367 (2019)
25. Chae, S.C., et al.: Array antenna with suppressed side lobe level for millimeter-wave applications. In: 2018 International Symposium on Antennas and Propagation, Busan, Korea (2019)
26. Liu, H.Q., et al.: Data feature analysis of non-scanning multi target millimeter-wave radar in traffic flow detection applications. Sensors 18(9), 2756 (2018)
27. Jarvis, R.A., Patrick, E.A.: Clustering using a similarity measure based on shared near neighbors. IEEE Trans. Comput. C-22(11), 1025–1034 (1973)
28. Qiu, B.Z., Xin, H.: Shared nearest neighbor affinity based clustering algorithm. Comput. Eng. Appl. 54(18), 184–187 (2018)

How to cite this article: Liu H, Teng K, Rai L, Zhang Y, Wang S. A two-step abnormal data analysis and processing method for millimetre-wave radar in traffic flow detection applications. IET Intell Transp Syst. 2021;15:671–682. https://doi.org/10.1049/itr2.12052