IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2020, accepted June 28, 2020, date of publication July 1, 2020, date of current version July 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006332

Traffic State Evaluation Using Traffic Noise

QINGLU MA[1,2] AND ZHENG ZOU[1]

[1]School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China
[2]Key Laboratory of Traffic System and Safety in Mountain Cities, Chongqing 400074, China

Corresponding author: Zheng Zou (zouzhn@mails.cqjtu.edu.cn)

This work was supported by the National Key Research and Development Plan: Theory and Method of Intelligent Perception of Road Infrastructure under Grant 2018YFB1600200.

ABSTRACT Traffic state information is widely applied into all aspects of Intelligent Transportation System (ITS), such as the macro-control of government departments, the implementation of traffic managers' plans, the decision-making of residents travel, and so on. At present, Mel Frequency Cepstrum Coefficient (MFCC) is generally used as characteristic of traffic noise to characterize different traffic states, and performs well in simple noise environment, but performs poorly in complex noise environment. Based on the analysis of traffic noise acquired from a roadside-installed acoustic acquisition equipment, the evaluation problem of traffic state in complex noise environment is considered in this paper. Traffic state is divided into three categories according to traffic speed in our work: free flow (40 km/h and above), saturated flow (10-40 km/h), and jammed flow (0-10 km/h). Teager Energy Operator (TEO) is introduced to improve the MFCC characteristic, thus a novel characteristic called T-MFCC is proposed. Principal Component Analysis (PCA) is introduced to reduce dimension of T-MFCC characteristic, thus a novel characteristic called PT-MFCC is proposed. Support Vector Machine (SVM) optimized by Particle Swarm Optimization (PSO) algorithm is applied as classifier to identify traffic state. Characterization capabilities of two modified characteristics and traditional MFCC characteristic for traffic state are compared in this paper. Experimental results demonstrate that the evaluation accuracy of traffic state based on T-MFCC characteristic is 3.685% higher than that based on MFCC characteristic, and the evaluation accuracy of traffic state based on PT-MFCC characteristic is 26.466% lower than that based on MFCC characteristic. Therefore, T-MFCC characteristic is superior to MFCC characteristic, while MFCC characteristic is superior to PT-MFCC characteristic, namely, T-MFCC characteristic can better characterize traffic state than MFCC characteristic, meanwhile, there are no redundancy attributes in T-MFCC characteristic, thus PCA is not needed to reduce the dimension of T-MFCC characteristic.

INDEX TERMS Traffic state, traffic noise, intelligent transportation systems, mel frequency cepstrum coefficient, teager energy operator, principal component analysis.

I. INTRODUCTION

Statistics from China Statistical Yearbook and the Ministry of Public Security show that from 2005 to 2019, the ownership of civil vehicles (small passenger vehicles) increases from 16.1835 million to 220 million, and the ownership of private vehicles (small passenger vehicles and miniature passenger vehicles) increases from 13.544 million to 207 million. The rapid increase of private car ownership and the continuous rise of residents travel demand lead to the ceaseless prolongation of peak time and the gradual expansion of traffic congestion. The growing traffic congestion has caused a series of life problems, such as frustration and ''road rage'', meanwhile, it has also caused huge economic losses (such as fuel consumption and time waste) and security risks for residents travel. How to effectively solve the problem of traffic congestion has attracted extensive attention of the transportation industry personnel, government departments, and the public. Nowadays, the application of Intelligent Transportation System (ITS) to automatically collect traffic information and identify traffic state has become a new breakthrough to solve the problem of traffic congestion [1].

In ITS, various infrastructure-based technologies (such as magnetic induction coils, and video cameras) are applied into vehicles and roads to detect various traffic parameters. The existing evaluation research of traffic state are mostly based on traditional detectors such as magnetic induction coil, video camera, and cellular networks [2]–[9]. Although

The associate editor coordinating the review of this manuscript and approving it for publication was Maurice J. Khabbaz.

these traffic detectors are widely used and have achieved good detection results, they have their own advantages and disadvantages: The advantages of magnetic inductive coil detector are that it has high detection accuracy and only needs a small investment in the preliminary stage, and the disadvantages are that once the coil detector broke down, it needs to close the lane and dig the road surface for maintenance, thus seriously reducing the road capacity and greatly increasing the maintenance difficulty of the coil detector; The advantages of video camera detector are that it can acquire visual images and obtain various traffic information such as traffic volume, density, vehicle type and license plate, and the disadvantages are that the detection results are easily affected by light conditions, such as day and night alternation, and ponding reflection, at the same time, the occlusion between vehicles will also affect the detection results, and the image processing is relatively time-consuming [10]. Compared with the magnetic inductive coil detector, the acoustic detector has lower maintenance cost and will not damage the road surface. Compared with the video camera detector, the acoustic detector is not affected by occlusion, day and night alternation, and light intensity, meanwhile, the calculation load required by the acoustic signal processing is reduced. Therefore, acoustic detector owns great development potential in the future, and it is meaningful to explore and study the application of traffic noise in traffic information detection. In recent years, the research of traffic information detection technology based on acoustic sensors is gradually increasing.

In order to understand and analyze traffic noise better, a few scholars have studied the internal properties of traffic noise. Traffic noise is mainly composed of tire and road friction noise, engine noise, exhaust noise, and aerodynamic noise [11]. In-depth research is conducted on the relevant properties of traffic noise: Tire/road noise increases with the increase of vehicle speed, it also changes with the change of road materials, and the change will become more evident when the vehicle speed is above 50 km/h; The engine noise always exists at all vehicle speeds, and its power depends on the engine load and rotation speed, in the meanwhile, the low-frequency noise component in the engine noise is in the majority, but a small number of high-frequency harmonics also exist; The exhaust noise belongs to low-frequency noise, and it has less impact on traffic noise under standard conditions; The aerodynamic noise increases with the increase of vehicle speed, and the increase will become more evident when the vehicle speed is above 90 km/h [10]. Traffic noise has non-stationary property, and it is difficult to extract robust acoustic characteristics due to the interference of environmental noise (such as vehicles in the adjacent lane on the highway) [12].

Currently some scholars utilize traffic noise to identify traffic state. In literature [13], Mel Frequency Cepstrum Coefficient (MFCC) characteristic of traffic noise is extracted, and Support Vector Machine (SVM) is used as classifier to identify free flow (40 km/h and above), saturated flow (20-40 km/h), and jammed flow (0-20 km/h). In literature [14] and [15], a traffic state evaluation method based on vehicle whistle noise is proposed, and the corresponding software and hardware systems are developed for near real-time traffic congestion monitoring on chaotic roads. In literature [16], a traffic state evaluation method based on MFCC characteristic of traffic noise is proposed: First, the traffic state is divided into free flow (40 km/h and above), saturated flow (10-40 km/h), and jammed flow (0-10 km/h); Second, the traffic noise is collected from three different places (in city), namely the traffic noise collection locations of each traffic state are different; Third, the MFCC characteristic of traffic noise is extracted, and the Vector Quantization (VQ), Artificial Neural Network (ANN), and K Nearest Neighbor (KNN) classification algorithms are compared in the experiment; Finally the experimental results show that VQ classification algorithm has better classification effect. In literature [17], a traffic state evaluation method based on Gray Level Co-Occurrence Matrix (GLCM) characteristic of traffic noise is proposed: First, the traffic state is also divided into jammed flow (0-10 km/h), saturated flow (10-40 km/h), and free flow (40 km/h and above); Second, the traffic noise data is also collected from three different places (including city and suburb); Third, the GLCM characteristic of traffic noise is extracted, and the SVM classifier is applied as classifier to identify the traffic state; Finally the experimental results show that the proposed method owns high evaluation accuracy. In literature [18], a traffic state evaluation method based on MFCC characteristic of traffic noise is proposed: First, the traffic state is still divided into free flow (40 km/h and above), saturated flow (10-40 km/h), and jammed flow (0-10 km/h); Second, the traffic noise data is still collected from three different places (in city); Third, the MFCC characteristic of traffic noise is extracted, the Gaussian Mixture Model (GMM) and SVM are utilized as classifier to identify the traffic state, and the SVM classifier is implemented by LIBSVM toolbox; Finally, the experimental results show that SVM owns higher evaluation accuracy than GMM. In literature [19], the traffic state is divided into ''Busy Street'' and ''Quiet Street'', and a traffic state evaluation method based on smart phone is proposed (smart phone is used to collect traffic noise data): First, the collected traffic noise data is divided into fixed size frames; Second, the characteristics based on time and frequency domain from each frame is extracted (including Zero Crossing Rate (ZCR), Short Time Energy (STE), Root Mean Square (RMS) and MFCC); Third, the SVM and Neural Network (NN) are used as classifier to identify the traffic state; Finally, the experimental results shows that the evaluation accuracy of SVM is slightly higher than that of NN.

There are also some scholars using traffic noise for vehicle type classification. Literature [20] mainly focuses on how to select the description characteristics of traffic noise that can make the vehicle classifier work normally. Literature [21] utilizes traffic noise and Probabilistic Neural Network (PNN) to identify different vehicle positions and types. Literature [22] studies the automatic vehicle type classification based on the spectrum characteristic of traffic noise. Literature

[23] proposes a new algorithm for accurately estimating the fundamental frequency of automobile engine sound, and then uses the proposed algorithm to identify different civil vehicles. Literature [24] introduces a sensor equipped with magnetometer and microphone to collect traffic noise and detect emergency vehicles. Literature [25] uses the Linear Prediction Coefficient (LPC) characteristic and Time Delay Neural Network (TDNN) to identify a single moving vehicle.

Traffic noise analysis is also applied to vehicle speed measurement. Literature [26] makes use of a small microphone array to collect traffic noise data and estimate vehicle speed. Literature [27] extracts the peak characteristic of traffic noise, and detects the number of vehicles, then estimates the passing time of vehicles using Generalized Cross Correlation (GCC) function, finally obtains the passing distance of vehicles through video in order to estimate the vehicle speed. Literature [28] finds through experiments that if the target vehicle moves rapidly or the traffic noise signal presents high frequency component, then the serious deviation may happen to the vehicle speed estimation, aiming at this problem, an improved vehicle speed estimation method is proposed.

Traffic noise analysis is also applied to traffic volume estimation. Literature [29] proposes a Triangular Wave Analysis (TWA) characteristic extraction algorithm for traffic volume estimation in order to solve the problem of intersectant Vehicle-Pass-Signals (VPSs) identification. Literature [30] uses traffic noise to study vehicle positioning so as to detect the existence of vehicle, and the number of vehicles finally can be converted into hourly traffic volume. Literature [31] proposes a real-time automatic vehicle detection algorithm based on the acoustic characteristic of wavelet packet transform. Literature [32] extracts the peak characteristic of the power envelope spectrum of traffic noise, then the number of vehicles is counted according to the peak number, finally the number of vehicles is converted into hourly traffic volume. Literature [10] conducts an 11.5-day experimental study on the roadside of Paris Ring Road, then extracts the MFCC characteristic of traffic noise and estimates the traffic volume with the Support Vector Regression (SVR) method, finally the experimental results demonstrate a good application prospect. Literature [33] constructs a Multilayer Perceptron Neural Network (MPNN), then estimates the short-term traffic volume of urban roads by taking the acoustic spectrum profile characteristic of traffic noise as the input data.

Research achievements above demonstrate that utilizing traffic noise to detect traffic information is feasible and effective. In the meanwhile, these research achievements also have important guiding significance for further research on the application of traffic noise in traffic information detection. Although some scholars have made use of traffic noise to evaluate traffic state, generally speaking, there is still a lack of comprehensive consideration on the following two aspects:

(1) Research on the evaluation of traffic state of the same road section in the city. At present, most scholars study the traffic state evaluation under the conditions of different road sections or suburban road section, while few scholars study traffic state evaluation under the conditions of the same urban road sections. Compared with the conditions of different road sections, the transition among different traffic states in condition of the same road section is more coherent and smooth, and the similarity among different traffic states is stronger, so it is more difficult to evaluate traffic state in condition of the same road section. The surrounding environment of suburban road section is usually quiet, so it is easier to evaluate traffic state compared with urban road section. However, the population density and the residents travel demand in urban are both bigger than those in suburban, thus urban road section is more prone to traffic congestion than suburban road section. Therefore, it is still of great significance to study the traffic state evaluation in condition of the same urban road section.

(2) Research on the acoustic characteristic extraction of traffic noise in condition of complex noise environment. So far, most scholars tend to detect traffic information by extracting MFCC characteristic of traffic noise, this also shows that it is reliable to use MFCC characteristic to characterize traffic state. Some scholars use peak frequency characteristic or GLCM characteristic of traffic noise to characterize traffic state, however, MFCC characteristic is more consistent with human ears auditory properties than peak frequency and GLCM characteristics. Although it is feasible to use MFCC characteristic to characterize traffic state, most scholars study traffic state evaluation problem in the condition that the surrounding environment interference is weak (such as no traffic noise interference from other road sections), while few scholars study traffic state evaluation problem in the condition that the surrounding environment interference is strong (such as traffic noise interference from other road sections). The stronger the environmental noise interference is, the more complex the traffic noise acquired is. Therefore, it remains to be further studied whether MFCC characteristic is still appropriate for characterizing traffic state in the condition that the surrounding environment interference is strong.

In Section I, the present situation of research of traffic information detection using traffic noise is introduced, and the problems remain to be solved in current traffic state evaluation research using traffic noise are analyzed. Section II introduces the pretreatment of traffic noise. Section III introduces the SVM classifier and LIBSVM toolbox. In Section IV, traditional traffic state evaluation method using traffic noise is introduced. In Section V, the modified traffic state evaluation methods using traffic noise are proposed. Traffic state evaluation results based on different methods are compared and discussed in Section VI. The research contents of this paper are concluded in Section VII.

II. PRETREATMENT OF TRAFFIC NOISE

Traffic noise should be pretreated before extracting its acoustic characteristics. The pretreatment operation of traffic noise mainly includes framing and windowing. The two steps are detailed below.

A. FRAMING

Traffic noise owns non-stationary property, thus it is not appropriate to analyze frequency domain directly [12]. Once the traffic noise is framed, then each frame can be regarded as a stationary acoustic signal. Generally, two adjacent frames will overlap in part each other. The framing diagram of traffic noise is shown as Figure 1.

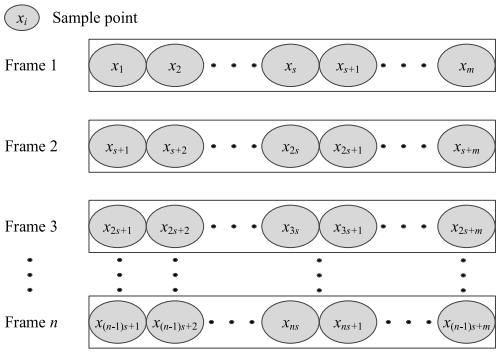


FIGURE 1. Framing diagram of traffic noise.

The traffic noise signal changes over time. Δt represents the time length of a short signal from the traffic noise signal. Constantly changing Δt, it can be observed that the larger the Δt is, the more obvious the change of the short signal is. Therefore, when Δt is small enough, the short signal then can be treated as a steady-state signal, commonly, the short signal is called frame, and the frame length is equal to Δt. In order to make the characteristic parameter of adjacent frames of traffic noise change continuously and smoothly, some sample points between adjacent frames is added, thus forming the overlap between adjacent frames.

In Figure 1, m represents the frame length (i.e. the number of sample points contained in a frame), and s represents the frame shift (i.e. m minus the number of overlapping sample points between adjacent frames). Sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ represents the traffic noise signal, N represents the number of sample points in the sequence, and f (unit: Hz) represents the sampling frequency of traffic noise, then the frame number n can be calculated via formula (1), and the time t_k (unit: s) of the kth frame can be calculated as formula (2),

$$n = [(N - m)/s] + 1 \tag{1}$$

$$t_k = (1/f) * ((k - 1)s + [m/2]) \tag{2}$$

where [] represents the rounding operation (namely directly discarding the decimal part), $k = 1, 2, \ldots, n$. It can be also seen from formula (1) that if the length of last frame is less than m, then the last frame will be directly discarded. Formula (2) shows that the time of each frame can be expressed through the time of center sample point of each frame.

B. WINDOWING

After the traffic noise signal is divided into frames, each frame should be multiplied by a window function, and the frame multiplied by the window function can be regarded as a periodic signal approximately. Most of the window functions have low-pass properties, in addition, the bandwidth and spectrum leakage caused by different window functions are also different. The frequently used window functions are rectangular window [34], hanning window [35], and hamming window [36]. Hanning window and hamming window are respectively defined as formula (3) and formula (4), where m represents the length of window function (generally equal to the frame length), and $j = 1, 2, \ldots, m$.

$$a_j = 0.5\left(1 - \cos\left(2\pi\ (j-1)\big/m - 1\right)\right) \tag{3}$$

$$b_j = 0.54 - 0.46\cos\left(2\pi\ (j-1)\big/m - 1\right) \tag{4}$$

Hamming window owns small spectrum leakage, what's more, literature [19] shows that hamming window can better eliminate the influence of sinusoidal signal sidelobe than hanning window, so hamming window is usually selected as the windowing function of traffic noise signal. The result after windowing the frame is shown as formula (5).

$$y_k\ (j) = b_j \boldsymbol{x}\ ((k-1)\ s + j) \tag{5}$$

III. SVM CLASSIFIER OPTIMIZED BY PSO ALGORITHM

In the case of less training data and linear indivisibility, Support Vector Machine (SVM) is still able to achieve high classification accuracy. SVM has become one of the most popular machine learning theories in the fields of data analysis and pattern recognition. In addition, most scholars also use SVM classifier to identify traffic state. In the aspect of SVM parameters optimization, Particle Swarm Optimization (PSO) algorithm is frequently used.

SVM is originally proposed to solve the problem of binary classification, but binary classifier is helpless in dealing with the problem of multi classification. The emergence of multi classification SVM is to make up for this defect of binary classifier. In order to solve practical problems efficiently, some SVM toolkits have been developed and designed, such as LIBSVM, LSSVM, SVMlight, and WEKA, where LIBSVM is most frequently used. LIBSVM toolbox is a convenient and efficient software package developed at the beginning of the 21st century, and its main function is to realize SVM pattern recognition and regression. The software package can be used across platforms, and the codes are open-source, namely, it can be second developed by users. In addition, LIBSVM involves relatively less parameter adjustment for SVM, and provides many default parameters, so it can still solve many problems. In practical application, Radial Basis Function (RBF) is usually used as kernel function. When SVM is used for classification and prediction, generally, penalty parameter c and RBF parameter g are adjusted to obtain ideal accuracy. LIBSVM provides Cross Validation (CV) function, dependent on the CV function, the best parameters can be obtained, and to some extent, the occurrence of ''over learning'' and ''under learning'' can be prevented, thus the better prediction effect can be achieved.

PSO algorithm belongs to the global stochastic search heuristic algorithm based on swarm intelligence. PSO

algorithm is usually utilized to optimize the penalty parameter c and the RBF parameter g for SVM. The idea of PSO algorithm originates from the foraging behavior of birds.

In the field of present pattern recognition, the problem that how to select appropriate penalty parameter c and RBF parameter g for SVM is still unsolved. In other words, c and g are generally determined according to experience, experimental test or some toolbox including CV function. While CV function is usually combined with PSO algorithm to improve the ability of optimization. Steps for optimizing SVM parameters c and g with PSO algorithm (namely PSO-SVM algorithm) are as follows:

Step 1: Initialize basic parameters. Let $F(\boldsymbol{x})$ represent the fitness function and set $F(\boldsymbol{x}) = -H(\boldsymbol{x})$, where $H(\boldsymbol{x})$ represents the objective function, $\boldsymbol{x} = [c\ g]^{\mathrm{T}}$, generally, $H(\boldsymbol{x})$ is equal to 100 multiplied by the accuracy of CV. The number of population particles is set as N_{pop}. Let j represent the particle counter and set $j = 1$. Let N_{gen} represent the maximum evolution generation. Let i represent the generation counter and set $i = 1$. Let $\boldsymbol{x}_j^{(0)}$ represent the initial location of particle, and set $\boldsymbol{x}_j^{(0)} = \left[c_j^{(0)}\ \ g_j^{(0)}\right]^{\mathrm{T}}$. Let $\boldsymbol{v}_j^{(0)}$ represent the initial moving speed of particle, and set $\boldsymbol{v}_j^{(0)} = \left[v_{j,1}^{(0)}\ \ v_{j,2}^{(0)}\right]^{\mathrm{T}}$. Let $\boldsymbol{y}_j$ represent the historical optimal location of the jth particle in the population, and set $\boldsymbol{y}_j = \boldsymbol{x}_j^{(0)}$. Let $\boldsymbol{z}$ represent the global optimal location of the population, and $\boldsymbol{z}$ satisfies $F(\boldsymbol{z}) = \min\limits_{1\le j\le N_{\mathrm{pop}}} F\left(\boldsymbol{x}_j^{(0)}\right)$. Let γ_1 represent the local search capability, and set $\gamma_1 = 1.5$. Let γ_2 represent the global search capability, and set $\gamma_2 = 1.7$. Let ω_0 represent the inertia weight, and set $\omega_0 = 1.0$. Let $c_{\max}$ represent the high threshold of penalty parameter c, and set $c_{\max} = 100$. Let $c_{\min}$ represent the low threshold of penalty parameter c, and set $c_{\min} = 0.1$. Let $g_{\max}$ represent the high threshold of RBF parameter g, and set $g_{\max} = 1000$. Let $g_{\min}$ represent the low threshold of RBF parameter g, and set $g_{\min} = 0.01$. Let κ represent the relationship coefficient between speed threshold and location threshold, and set $\kappa = 0.6$. Let $v_{c\max} = \kappa c_{\max}$, $v_{c\min} = -v_{c\max}$, $v_{g\max} = \kappa g_{\max}$, and $v_{g\min} = -v_{g\max}$, where $v_{c\min}$ and $v_{c\max}$ respectively represent the minimum updating speed and the maximum updating speed of penalty parameter c, while $v_{g\min}$ and $v_{g\max}$ respectively represent the minimum updating speed and the maximum updating speed of RBF parameter g.

Step 2: Update speed and location. Update the moving speed of particle according to formula (6), and update the particle location according to formula (7), where $\boldsymbol{v}_j^{(i)}$ represent the moving speed of particle, $\boldsymbol{x}_j^{(i)}$ represent particle location, ϑ_1 and ϑ_2 both represent a random number between 0 and 1. Judge whether the location and moving speed exceed their respective thresholds. If exceeding the threshold, then the threshold is taken directly.

$$\boldsymbol{v}_j^{(i)} = \omega_0 \boldsymbol{v}_j^{(i-1)} + \gamma_1 \vartheta_1 \left(\boldsymbol{y}_j - \boldsymbol{x}_j^{(i-1)}\right) + \gamma_2 \vartheta_2 \left(\boldsymbol{z} - \boldsymbol{x}_j^{(i-1)}\right) \quad (6)$$

$$\boldsymbol{x}_j^{(i)} = \boldsymbol{x}_j^{(i-1)} + \boldsymbol{v}_j^{(i)} \quad (7)$$

Step 3: Update the historical optimal location of particle. If $F\left(\boldsymbol{x}_j^{(i)}\right) < F\left(\boldsymbol{y}_j\right)$, then set $\boldsymbol{y}_j = \boldsymbol{x}_j^{(i)}$.

Step 4: Update the global optimal location of the population. If $F\left(\boldsymbol{x}_j^{(i)}\right) < F(\boldsymbol{z})$, then set $\boldsymbol{z} = \boldsymbol{x}_j^{(i)}$.

Step 5: Judge whether all particles of the ith generation are completely updated. If $j < N_{\mathrm{pop}}$, then set $j = j + 1$, and return to Step 2. Otherwise, set $j = 1$, and return to Step 6.

Step 6: Determine whether to finish the PSO-SVM algorithm. If $i < N_{\mathrm{gen}}$, then set $i = i + 1$, and return to Step 2. Otherwise, finish the PSO-SVM algorithm.

IV. THE TRADITIONAL TRAFFIC STATE EVALUATION METHOD BASED ON MFCC CHARACTERISTIC

Mel Frequency Cepstrum Coefficient (MFCC) is one of the most commonly used acoustic characteristics in the fields of speech endpoint detection and speech recognition. At present, most scholars also utilize MFCC characteristic of traffic noise to characterize different traffic states. The traditional traffic state evaluation method based on MFCC characteristic of traffic noise is reviewed in this section.

A. REVIEW OF MFCC CHARACTERISTIC

The MFCC characteristic extraction flowchart is shown as Figure 2. MFCC extraction process mainly includes Fast Fourier Transform (FFT), Mel filter bank design, and Discrete Cosine Transform (DCT).

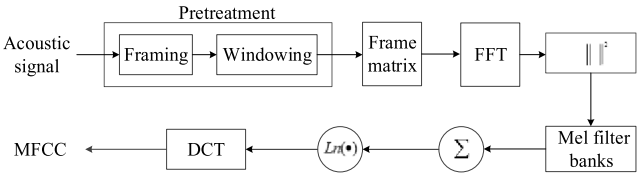


FIGURE 2. Flowchart of MFCC characteristic extraction.

1) FAST FOURIER TRANSFORM (FFT)

If $x(n)$ is a complex sequence and its length is N, then its N-point Discrete Fourier Transform (DFT) sequence $X(k)$ usually also represents a complex sequence and its length is still N. The discrete sequence $x(n)$ and periodic complex index sequence can be written as the complex form of real part plus virtual part, i.e. $x(n) = \mathrm{Re}[x(n)] + \mathrm{jIm}[x(n)]$, $W_N^{nk} = \mathrm{Re}\left[W_N^{nk}\right] + \mathrm{jIm}\left[W_N^{nk}\right] = \cos\left(\frac{2\pi}{N}nk\right) - \mathrm{j}\sin\left(\frac{2\pi}{N}nk\right)$, so the DFT of $x(n)$ can be defined as formula (8), where $k = 0, 1, \ldots, N-1$.

$$\begin{aligned}
&\mathrm{DFT}[x(n)] \\
&= X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk} \\
&= \sum_{n=0}^{N-1} \left\{ \mathrm{Re}[x(n)] \cos\left(\frac{2\pi}{N}nk\right) + \mathrm{Im}[x(n)] \sin\left(\frac{2\pi}{N}nk\right) \right\} \\
&\quad + \mathrm{j}\left\{ \mathrm{Im}[x(n)] \cos\left(\frac{2\pi}{N}nk\right) - \mathrm{Re}[x(n)] \sin\left(\frac{2\pi}{N}nk\right) \right\}
\end{aligned} \quad (8)$$

Since one-time complex number multiplication is equivalent to four times of real number multiplication and two times of real number addition, while one-time complex number addition is equivalent to two times of real number addition, therefore, every $X(k)$ calculated is equivalent to performing $4N$ times of real number multiplication and $2N + 2(N-1) = 4N - 2$ times of real number addition. From formula (8), it can be concluded that DFT$[x(n)]$ needs to execute N times of calculation of $x(k)$ in total. Therefore, in DFT calculation process, $4N^2$ times of real multiplication operations and $4N^2 - 2N$ times of real addition operations are required. In conclusion, the longer the sequence is, the larger the calculation of DFT is, what's more, when N is very large, the DFT calculation will increase dramatically.

FFT is essentially a fast algorithm of DFT, and FFT can greatly reduce the DFT operation load. Compared with DFT algorithm, the longer the sequence length is, the larger the operation load saved by FFT algorithm is, and the more obvious the advantages are. FFT algorithm can be divided into two categories: time extraction algorithm and frequency extraction algorithm.

a: TIME EXTRACTION ALGORITHM

Assuming that the sequence length N satisfies $N = 2^m$, where m represents a positive integer, the sequence $x(n)$ is decomposed into the sum of even and odd terms, as shown in formula (9),

$$\begin{cases} x(2r) = x_1(r) \\ x(2r+1) = x_2(r) \end{cases} \quad r = 0, 1, \cdots, N/2-1 \tag{9}$$

correspondingly, the DFT is also decomposed into the combination of odd and even terms, as shown in formula (10),

$$x(k) = \sum_{n=0}^{N-1} x(n) w_N^{nk} = \sum_{r=0}^{N/2-1} x(2r) w_N^{2rk} + w_N^k \sum_{r=0}^{N/2-1} x(2r+1) w_N^{2rk} \tag{10}$$

according to the symmetry, it can be derived that $w_N^{2n} = e^{-j\frac{2\pi}{N}2n} = w_{N/2}^n$. $X(k)$ can be also expressed as formula (11).

$$X(k) = \sum_{r=0}^{N/2-1} x(2r) W_{\frac{N}{2}}^{rk} + W_N^k \sum_{r=0}^{N/2-1} x(2r+1) W_{\frac{N}{2}}^{rk} \tag{11}$$

Therefore, an N-point DFT can be decomposed into two $N/2$-point DFTs, and these two $N/2$-point DFTs can be recombined into one N-point DFT, and so on. If N satisfies $N = 2^m$ (m is a positive integer), then the N-point DFT can finally become a linear combination of 2-point DFT after m-times decomposition. The above decomposition process is also known as m-level operations from $x(n)$ to $X(k)$. Each level of operations experiences $N/2$-times complex multiplication and N-times complex addition, so the m-level operations finally experience $N/2\log_2 N$-times complex multiplication and $N\log_2 N$-times complex addition.

b: FREQUENCY EXTRACTION ALGORITHM

Compared with time extraction algorithm, in frequency extraction algorithm, the sequence $X(k)$ is not divided into two parts according to the way of odd and even division, but according to the way of front and back half division. The specific implementation of decomposition of N-point DFT in frequency extraction algorithm is shown as formula (12).

$$\begin{aligned} X(k) &= \sum_{n=0}^{N/2-1} x(n) W_N^{nk} + \sum_{n=N/2}^{N-1} x(n) W_N^{nk} \\ &= \sum_{n=0}^{N/2-1} x(n) W_N^{nk} + \sum_{n=0}^{N/2-1} x\left(n+N/2\right) W_N^{\left(n+\frac{N}{2}\right)k} \\ &= \sum_{n=0}^{N/2-1} \left[x(n) + W_N^{\frac{N}{2}k} x\left(n+N/2\right)\right] W_N^{nk} \end{aligned} \tag{12}$$

It can be further divided into even and odd groups, as shown in formulas (13), (14), and (15). If N satisfies $N = 2^m$, then $N/2$ is still an even number, that is to say, after m-times decomposition of N-point DFT, the final DFT is composed of 2-point DFT, and only includes addition and subtraction operations.

$$X(k) = \sum_{n=0}^{N/2-1} \left[x(n) + (-1)^k x\left(n+N/2\right)\right] W_N^{nk} \tag{13}$$

$$X(2r) = \sum_{n=0}^{N/2-1} \left[x(n) + x\left(n+N/2\right)\right] W_{N/2}^{2nr} \tag{14}$$

$$X(2r+1) = \sum_{n=0}^{N/2-1} \left[x(n) + x\left(n+N/2\right)\right] W_N^{n(2r+1)} \tag{15}$$

2) MEL FILTER BANK

When designing Mel filter bank, some band-pass filters $H_n(k)$ are usually inserted into different frequency ranges of signals, and the triangle filter is generally selected as the band-pass filter. The central frequency of the filter is represented by $f(n)$. In Mel frequency range, the filter bandwidth is equal. The transfer function of band-pass filter in different frequency range can be expressed by formula (16),

$$H_n(k) = \begin{cases} 0 & k < f(n-1) \\ \dfrac{k - f(n-1)}{f(n) - f(n-1)} & f(n-1) \le k \le f(n) \\ \dfrac{f(n+1) - k}{f(n+1) - f(n)} & f(n) \le k \le f(n+1) \\ 0 & k > f(n+1) \end{cases} \tag{16}$$

where $f(n)$ can be expressed in the form of formula (17),

$$f(n) = \left(\frac{N}{f_s}\right) F_{\text{mel}}^{-1}\left(F_{\text{mel}}(f_l) + n\frac{F_{\text{mel}}(f_h) - F_{\text{mel}}(f_l)}{m+1}\right) \tag{17}$$

where f_l represents the lowest frequency, f_h represents the highest frequency, N represents the sequence length, f_s represents the sampling frequency, F_{mel} represents the Mel function ($F_{\text{mel}} = 1125\ln(1 + f/700)$),

and F_{mel}^{-1} represents the inverse operation of Mel function ($F_{\text{mel}}^{-1}(x) = 700(\text{e}^{x/1125} - 1)$).

3) DISCRETE COSINE TRANSFORM (DCT)

DCT owns many advantages, such as rich spectrum components, energy concentration, and no phase estimation. DCT can achieve good signal enhancement effect under low computational complexity. Let $x(n)$ denote a signal sequence and its length is N, where $n = 0, 1, \ldots, N-1$, then the complete orthogonal normalization function of DCT can be expressed as formula (18),

$$\begin{cases} X(k) = \delta(k) \sum_{n=0}^{N-1} x(n) \cos\left(\dfrac{\pi(2n+1)k}{2N}\right) \\ x(n) = \sum_{k=0}^{N-1} \delta(k) X(k) \cos\left(\dfrac{\pi(2n+1)k}{2N}\right) \end{cases} \tag{18}$$

where $\delta(k)$ can be defined according to formula (19),

$$\delta(k) = \begin{cases} \sqrt{1/N} & k = 0 \\ \sqrt{2/N} & 1 \le k \le N-1 \end{cases} \tag{19}$$

therefore, $X(k)$ can be also expressed as formula (20), as shown at the bottom of the next page.

Another representation of DCT can be obtained from formula (20), as shown in formula (21), where $k = 0, 1, \ldots, N-1$, and $C(k)$ represents the orthogonal factor, which ensures the normalization of the transformation basis.

$$\begin{cases} X(k) = \sqrt{\dfrac{2}{N}} \sum_{n=0}^{N-1} C(k)\, x(n) \cos\left[\dfrac{\pi(2n+1)k}{2N}\right] \\ C(k) = \begin{cases} \sqrt{2}/2 & k = 0 \\ 1 & k = 1, 2, \cdots, N-1 \end{cases} \end{cases} \tag{21}$$

Therefore, the DCT matrix can be written as $\boldsymbol{X} = \boldsymbol{C}_N\, \boldsymbol{x}$, where $\boldsymbol{X}$ and $\boldsymbol{x}$ are both column vector, $\boldsymbol{X}$ represents the output sequence of DCT, and $\boldsymbol{x}$ represents the input sequence of DCT, $\boldsymbol{C}_N$ represents the transformation matrix, and the element of $\boldsymbol{C}_N$ can be solved according to formula (21).

B. THE TRADITIONAL TRAFFIC STATE EVALUATION METHOD BASED ON MFCC CHARACTERISTIC

The flowchart of traffic state evaluation based on MFCC and PSO-SVM is shown as Figure 3. First, the traffic noise is collected under different traffic states, and each traffic state is marked. Second, the traffic noise is pretreated through framing and windowing, thus the frame matrix of traffic noise is acquired. Third, the MFCC characteristic of traffic noise is extracted according to the frame matrix. Fourth, the MFCC characteristic sets are divided into two parts: training sets and testing sets. The optimal parameters c and g are acquired after executing the PSO-SVM algorithm. Finally, the SVM prediction model constructed by the optimal parameters c and g is used to predict the testing sets, thus different traffic states are evaluated.

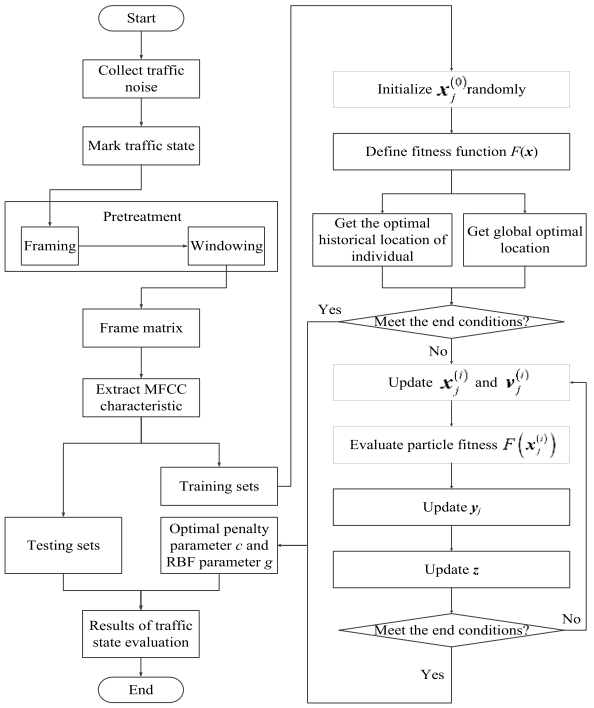


FIGURE 3. Flowchart of traffic state evaluation based on MFCC.

V. THE PROPOSED TRAFFIC STATE EVALUATION METHOD BASED ON THE MODIFIED MFCC CHARACTERISTIC

A. TEAGER ENERGY OPERATOR

Teager Energy Operator (TEO) is a non-linear operator proposed by Kaiser [37]. TEO can effectively reflect the signal energy, and has been successfully applied to acoustic signal processing. TEO can enhance the stable or semi-stable signal, and also attenuate the unstable signal, meanwhile, it also owns the non-linear energy tracking property. In addition, TEO is also extremely sensitive to the amplitude envelope of amplitude modulated signal and the instantaneous frequency change of frequency modulation signal.

The calculation of TEO can be realized by formula (22) and (23), where $i = 2, 3, \ldots, r-1$, and $j = 1, 2, \ldots, n$. Acoustic signal after framing and windowing, can be represented by matrix $\boldsymbol{Y}$ (also called frame matrix), where each column represents a frame, r represents the frame length (number of sample points of a frame), n represents the number of frames, $\boldsymbol{T}$ represents the TEO characteristic matrix of acoustic signal, and each element t_{ij} in $\boldsymbol{T}$ can be calculated by formula (23).

$$\boldsymbol{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{r,1} & y_{r,2} & \cdots & y_{r,n} \end{pmatrix},$$
$$\boldsymbol{T} = \begin{pmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{r,1} & t_{r,2} & \cdots & t_{r,n} \end{pmatrix} \tag{22}$$

$$\begin{cases} t_{i,j} = y_{i,j}^2 - y_{i-1,j}y_{i+1,j} \\ t_{1,j} = 2t_{2,j} - t_{3,j} \\ t_{r,j} = 2t_{r-1,j} - t_{r-2,j} \end{cases} \tag{23}$$

B. PRINCIPAL COMPONENT ANALYSIS

In order to solve the practical problem, two or more variables will be introduced more often than not, but some variables may correlate strongly, and it will not only increase the amount of calculation but also may produce many mistakes if using strongly correlated variables directly to analyze problems. To increase the utilization rate of data, more old variables are usually replaced with fewer new variables, meanwhile, the new variables are required to reflect the original information of data as much as possible. Therefore, Principal Component Analysis (PCA) is designed under this background. The essence of PCA is to replace multiple variables with a few comprehensive variables [38].

1) PRINCIPAL COMPONENT ANALYSIS MODEL

Let $\boldsymbol{X}_i$ represent the ith random variables, and $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p)^{\mathrm{T}}$. Expectation of $\boldsymbol{X}$ is recorded as $E(\boldsymbol{X}) = (E(\boldsymbol{X}_1), E(\boldsymbol{X}_2), \ldots, E(\boldsymbol{X}_p))^{\mathrm{T}}$, and the covariance matrix $\boldsymbol{\Sigma}$ is shown as formula (24), where Cov represents the covariance function.

$$\boldsymbol{\Sigma} = \left(\sigma_{i,j}\right)_p = Cov\left(\boldsymbol{X}, \boldsymbol{X}\right) = \boldsymbol{E}\left(\left(\boldsymbol{X} - \boldsymbol{E}\left(\boldsymbol{X}\right)\right)\left(\boldsymbol{X} - \boldsymbol{E}\left(\boldsymbol{X}\right)\right)^{\mathrm{T}}\right) \tag{24}$$

The new variable $\boldsymbol{Y}_i$ can be created by old variables $\boldsymbol{X}$. It is calculated by formula (25),

$$\begin{cases} \boldsymbol{Y}_1 = l_1^{\mathrm{T}}\boldsymbol{X} = l_{11}\boldsymbol{X}_1 + l_{12}\boldsymbol{X}_2 + \cdots + l_{1p}\boldsymbol{X}_p \\ \boldsymbol{Y}_2 = l_2^{\mathrm{T}}\boldsymbol{X} = l_{21}\boldsymbol{X}_1 + l_{22}\boldsymbol{X}_2 + \cdots + l_{2p}\boldsymbol{X}_p \\ \vdots \\ \boldsymbol{Y}_p = l_p^{\mathrm{T}}\boldsymbol{X} = l_{p1}\boldsymbol{X}_1 + l_{p2}\boldsymbol{X}_2 + \cdots + l_{pp}\boldsymbol{X}_p \end{cases} \tag{25}$$

where $\boldsymbol{l}_i = (l_{i1}, l_{i2}, \cdots, l_{ip})^{\mathrm{T}}$ denotes the composite variable coefficient, and $i = 1, 2, \cdots, p$. Meanwhile, formula (25) is required to meet the following three restrictions:

(1) The composite variable coefficient $\boldsymbol{l}_i$ is a unit vector, as shown in formula (26), where $i = 1, 2, \cdots, p$.

$$\boldsymbol{l}_i^{\mathrm{T}}\boldsymbol{l}_i = l_{i1}^2 + l_{i2}^2 + \cdots + l_{ip}^2 = 1 \tag{26}$$

(2) $\boldsymbol{Y}_i$ is not related to $\boldsymbol{Y}_j$, namely, the covariance of $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ is 0, as shown in formula (27).

$$Cov\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j\right) = Cov\left(\boldsymbol{l}_i^{\mathrm{T}}\boldsymbol{X}, \boldsymbol{l}_j^{\mathrm{T}}\boldsymbol{X}\right) = \boldsymbol{l}_i^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{l}_j = 0 \tag{27}$$

(3) The variance of $\boldsymbol{Y}_i$ decreases monotonically with the increase of i, as shown in formula (28).

$$Var(Y_1) \geq Var(Y_2) \geq \cdots \geq Var(Y_p) \geq 0 \tag{28}$$

Therefore, $\boldsymbol{Y}_i$ is called the ith principal component, coefficient $\boldsymbol{l}_i$ is called the ith principal component coefficient, where $i = 1, 2, \cdots, p$.

2) MODEL SOLUTION

According to the limit conditions from formula (26) to formula (28), the principal component and its related parameters (such as variance and contribution rate) can be calculated.

a: PRINCIPAL COMPONENT SOLUTION

Let λ_i represent eigenvalue of the covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \cdots, e_{ip},)$ denote the unit orthogonal eigenvector. If $\boldsymbol{\Sigma}$ has been determined, then $\boldsymbol{Y}_i$ can be calculated via formula (29), in addition, the variance of principal component and the covariance between principal components can be calculated by formula (30), where $i \neq j$, meanwhile $i, j = 1, 2, \cdots, p$.

$$\boldsymbol{Y}_i = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{X} = e_{i1}\boldsymbol{X}_1 + e_{i2}\boldsymbol{X}_2 + \cdots + e_{ip}\boldsymbol{X}_p \tag{29}$$

$$\begin{cases} Var\left(\boldsymbol{Y}_i\right) = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{e}_i = \lambda_i \\ Cov\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j\right) = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{e}_j = 0 \end{cases} \tag{30}$$

b: TOTAL VARIANCE OF PRINCIPAL COMPONENTS AND ITS CONTRIBUTION RATE

Let $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_p)^{\mathrm{T}}$, and $\boldsymbol{P} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_p)$, then $\boldsymbol{Y} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{X}$, and the covariance matrix of $\boldsymbol{Y}$ is shown in formula (31). The total variance of $\boldsymbol{Y}_i$ is equal to the total variance of $\boldsymbol{X}_i$, as shown in formula (32).

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{P} = \boldsymbol{\Lambda} = Diag\left(\lambda_1, \lambda_2, \cdots, \lambda_p\right) \tag{31}$$

$$\sum_{i=1}^{p} Var\left(\boldsymbol{Y}_i\right) = \sum_{i=1}^{p} Var\left(\boldsymbol{X}_i\right) \tag{32}$$

Let α_m represent the sum of the contribution rate of the first m ($m \leq p$) principal components, α_m is generally also called the cumulative contribution rate. In practical problems, m principal components are usually selected to make the cumulative contribution rate reach more than 80%.

$$\alpha_m = \sum_{i=1}^{m} \lambda_i \Big/ \sum_{j=1}^{p} \lambda_j \tag{33}$$

$$\begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} = \sqrt{\frac{2}{N}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \\ \cos\frac{\pi}{2N} & \cos\frac{3\pi}{2N} & \cdots & \cos\frac{(2N-1)\pi}{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \cos\frac{(N-1)\pi}{2N} & \cos\frac{3(N-1)\pi}{2N} & \cdots & \cos\frac{(2N-1)(N-1)\pi}{2N} \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \tag{20}$$

C. THE PROPOSED TRAFFIC STATE EVALUATION METHOD BASED ON T-MFCC CHARACTERISTIC

1) THE PROPOSED T-MFCC CHARACTERISTIC

We utilize TEO to modify the MFCC characteristic, thus a new acoustic characteristic is acquired, and this new characteristic is named T-MFCC by us in this paper. First, the MFCC characteristic is extracted; second, the characteristic value of each dimension of MFCC is fused with the mean value of TEO characteristic; finally, the T-MFCC characteristic is acquired, and its extraction flowchart is as shown in Figure 4, where the content in the red dotted box represents the modified part relative to MFCC characteristic.

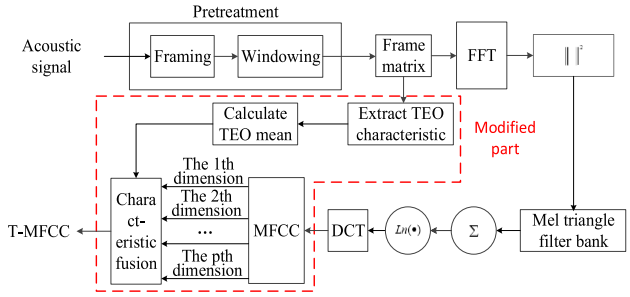


FIGURE 4. Extraction flowchart of T-MFCC characteristic.

The acoustic signal data is pretreated by framing and windowing, and the frame matrix is obtained. At this time, the frame matrix is operated in two aspects: on the one hand, the frame matrix is processed by FFT, Mel filtering and DCT to obtain MFCC characteristic matrix; on the other hand, the TEO characteristic is extracted directly based on the frame matrix. The mean value of each column of TEO characteristic is calculated, and then the TEO mean is multiplied by the characteristic value of each dimension of MFCC to realize the fusion between TEO characteristic and MFCC characteristic.

Let $\boldsymbol{M}$ and $\boldsymbol{\chi}$ represent MFCC characteristic matrix and T-MFCC characteristic matrix respectively, as shown in formula (34), where p represents the characteristic dimension and n represents the total number of frames.

$$\boldsymbol{M} = \begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p,1} & m_{p,2} & \cdots & m_{p,n} \end{pmatrix},$$
$$\boldsymbol{\chi} = \begin{pmatrix} \chi_{1,1} & \chi_{1,2} & \cdots & \chi_{1,n} \\ \chi_{2,1} & \chi_{2,2} & \cdots & \chi_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{p,1} & \chi_{p,2} & \cdots & \chi_{p,n} \end{pmatrix} \tag{34}$$

In formula (34), each column of $\boldsymbol{M}$ represents a frame of MFCC characteristic, and each frame of MFCC characteristic owns p attributes M1, M2, $\cdots$, Mp, where Mi $=$ $(m_{i1}, m_{i2}, \cdots, m_{in})^{\mathrm{T}}$, $i = 1, 2, \cdots, p$. Similarly, each column of $\boldsymbol{\chi}$ represents a frame of T-MFCC characteristic, and each frame of T-MFCC characteristic owns p attributes TM1, TM2, $\cdots$, TMp, where TMj $= \begin{bmatrix} \chi_{j,1} & \chi_{j,2} & \cdots & \chi_{j,n} \end{bmatrix}^{\mathrm{T}}$, $j = 1, 2, \cdots, p$.

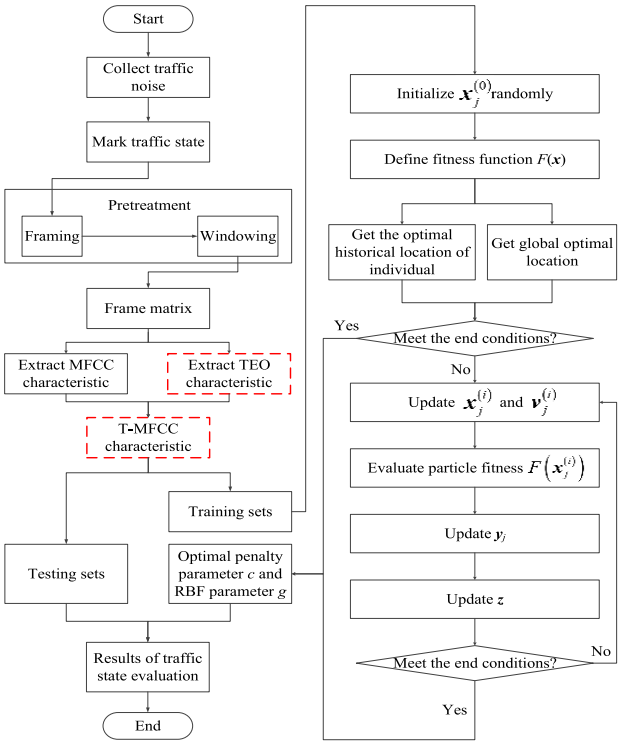


FIGURE 5. Flowchart of traffic state evaluation based on T-MFCC.

Therefore, according to Figure 5, the T-MFCC characteristic extraction algorithm can be concluded as follows:

Step 1: Extract MFCC characteristic matrix according to Figure 3, and initialize $j = 1$.

Step 2: Calculate the TEO mean of frame j, as shown in formula (35), where t_{ij} can be calculated by formula (23).

$$\bar{t}_j = \frac{1}{r} \sum_{i=1}^{r} t_{i,j} \tag{35}$$

Step 3: Define $\chi_{i,j}$ as shown in formula (36).

$$\chi_{i,j} = \bar{t}_j m_{i,j} \tag{36}$$

Step 4: Let $j = j + 1$. If $j \leq n$, then return to Step 2; otherwise, generate T-MFCC characteristic matrix and end the T-MFCC characteristic extraction algorithm.

2) THE PROPOSED TRAFFIC STATE EVALUATION METHOD BASED ON T-MFCC CHARACTERISTIC

The flowchart of traffic state evaluation based on T-MFCC and PSO-SVM is shown as Figure 5. First, the traffic noise is collected under different traffic states, and each traffic state is marked. Second, the traffic noise is pretreated through framing and windowing, thus the frame matrix of traffic noise is acquired. Third, the MFCC and TEO characteristics of traffic noise are extracted according to the frame matrix. Fourth, the mean value of each column of TEO characteristic is calculated, and the TEO mean is multiplied by the characteristic value of each dimension of MFCC to acquire the T-MFCC characteristic. Fifth, the T-MFCC characteristic

sets are divided into two parts: training sets and testing sets. The optimal parameters c and g are acquired after executing the PSO-SVM algorithm. Finally, the SVM prediction model constructed by the optimal parameters c and g is used to predict the testing sets, thus different traffic states are evaluated.

D. THE PROPOSED TRAFFIC STATE EVALUATION METHOD BASED ON PT-MFCC CHARACTERISTIC

1) THE PROPOSED PT-MFCC CHARACTERISTIC

Using PCA to reduce the dimension of T-MFCC characteristic, thus we can generate another new acoustic characteristic, and this new characteristic is named PT-MFCC by us in this paper. The extraction flowchart of PT-MFCC characteristic is shown as Figure 6, where the content in the red dotted box represents the modified part relative to MFCC characteristic, and the content in the blue dotted box represents the modified part relative to T-MFCC characteristic.

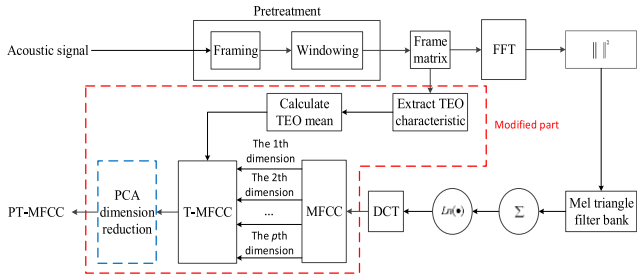


FIGURE 6. Extraction flowchart of PT-MFCC characteristic.

T-MFCC is modified by PCA to reduce its characteristic dimension. The steps of PT-MFCC characteristic extraction algorithm are as follows:

Step 1: Extract the T-MFCC characteristic matrix and initialize the threshold of cumulative contribution rate α_0 $(0 < \alpha_0 < 1)$.

Step 2: Standardize the T-MFCC characteristic matrix. In order to reduce the data level differences, the T-MFCC characteristic matrix is standardized according to formula (37), and the standardization process is shown as formula (38), where $1 \le i \le p$, $1 \le j \le n$, $\bar{\chi}_i$ represents the sample mean, and s_{ij} represents sample variance.

$$\boldsymbol{\chi}^* = \begin{pmatrix} \chi^*_{1,1} & \chi^*_{1,2} & \cdots & \chi^*_{1,n} \\ \chi^*_{2,1} & \chi^*_{2,2} & \cdots & \chi^*_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \chi^*_{p,1} & \chi^*_{p,2} & \cdots & \chi^*_{p,n} \end{pmatrix} \tag{37}$$

$$\begin{cases} \chi^*_{i,j} = \dfrac{\chi_{i,j} - \bar{\chi}_i}{\sqrt{s_{i,i}}} \\ \bar{\chi}_i = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} \chi_{i,k} \\ s_{i,i} = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n} \left(\chi_{i,k} - \bar{\chi}_i\right)^2 \end{cases} \tag{38}$$

Step 3: Calculate the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\chi}^*}$ of $\boldsymbol{\chi}^*$, as shown in formula (39), where $1 \le j \le n$.

$$\begin{cases} \boldsymbol{\Sigma}_{\boldsymbol{\chi}^*} = \dfrac{1}{n-1}\displaystyle\sum_{j=1}^{n} \left(\boldsymbol{\chi}^*_j - \bar{\boldsymbol{\chi}}^*\right)\left(\boldsymbol{\chi}^*_j - \bar{\boldsymbol{\chi}}^*\right)^{\mathrm{T}} \\ \bar{\boldsymbol{\chi}}^* = \frac{1}{n}\displaystyle\sum_{j=1}^{n} \boldsymbol{\chi}^*_j \\ \boldsymbol{\chi}^*_j = \begin{pmatrix} \chi^*_{1,j} & \chi^*_{2,j} & \cdots & \chi^*_{p,j} \end{pmatrix}^{\mathrm{T}} \end{cases} \tag{39}$$

Step 4: Calculate the eigenvalues $\boldsymbol{\Omega}_i$ of the principal component and the eigenvalues λ_i of $\boldsymbol{\Sigma}_{\boldsymbol{\chi}^*}$, as shown in formula (40) and (41) respectively, where $i = 1, 2, \cdots, p$.

$$\boldsymbol{\Omega}_i = \boldsymbol{\chi}^{*\mathrm{T}} \boldsymbol{e}_i = \left(e_{i,1}\boldsymbol{\chi}^*_1 + e_{i,2}\boldsymbol{\chi}^*_2 + \cdots + e_{i,p}\boldsymbol{\chi}^*_p\right)^{\mathrm{T}} \tag{40}$$

$$\lambda_i = \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\chi}^*} \boldsymbol{e}_i \tag{41}$$

Step 5: Calculate the cumulative contribution rate α_m of the first m principal components according to formula (33). Obviously, α_m strictly monotonically increases with the increase of m. If $\alpha_m \le \alpha_0$, then let $m_{\max}$ represent the maximum m value, so the former $m_{\max}$ principal components can be regarded as the attributes of the PT-MFCC characteristic, that is, $m_{\max}$ is the dimension of PT-MFCC. The characteristic matrix $\boldsymbol{\Omega}$ of PT-MFCC, and can be expressed in the form of $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_2 & \cdots & \boldsymbol{\Omega}_{m_{\max}} \end{pmatrix}^{\mathrm{T}}$, each column of $\boldsymbol{\Omega}$ represents a frame of PT-MFCC characteristic. End the PT-MFCC characteristic extraction algorithm.

2) THE PROPOSED TRAFFIC STATE EVALUATION METHOD BASED ON PT-MFCC CHARACTERISTIC

The flowchart of traffic state evaluation based on PT-MFCC and PSO-SVM is shown as Figure 7. First, the traffic noise is collected under different traffic states, and each traffic state is marked. Second, the traffic noise is pretreated through framing and windowing, thus the frame matrix of traffic noise is acquired. Third, the MFCC and TEO characteristics of traffic noise are extracted according to the frame matrix. Fourth, the mean value of each column of TEO characteristic is calculated, and the TEO mean is multiplied by the characteristic value of each dimension of MFCC to acquire the T-MFCC characteristic. Fifth, PCA is introduced to reduce the T-MFCC characteristic dimension, thus PT-MFCC characteristic is acquired. Sixth, the PT-MFCC characteristic sets are divided into two parts: training sets and testing sets. The optimal parameters c and g are acquired after executing the PSO-SVM algorithm. Finally, the SVM prediction model constructed by the optimal parameters c and g is used to predict the testing sets, thus different traffic states are evaluated.

VI. CASE STUDY

A. TRAFFIC NOISE DATA ACQUISITION

Up to now, there is not a consistent definition regarding traffic state all over the world. In different road conditions, drivers have different feelings about traffic state, that is, the characteristic parameters of traffic flow change with the

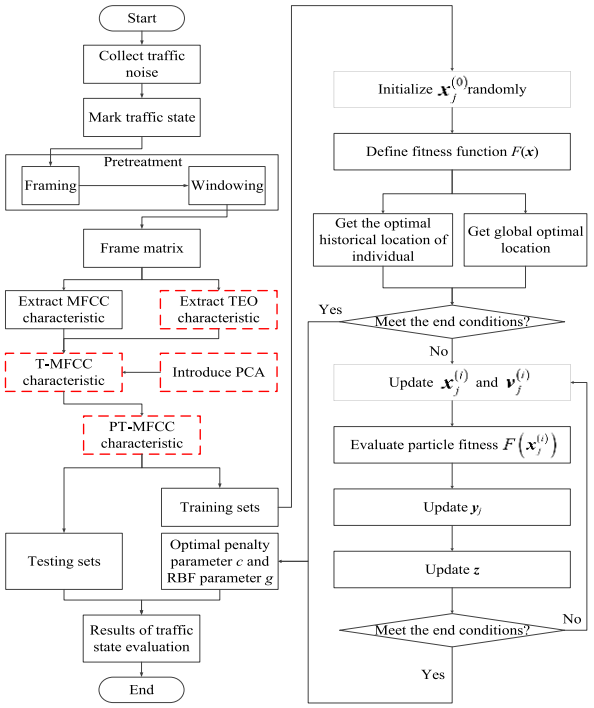


FIGURE 7. Flowchart of traffic state evaluation based on PT-MFCC.

change of traffic state. In order to collect the traffic noise under different traffic states, it is necessary to acquire the traffic state by other ways in advance. The average speed of vehicles is generally used to judge the traffic state. Most scholars divide 0-10 km/h into jammed flow, 10-40 km/h into saturated flow, and 40 km/h above into free flow. A few scholars divide 0-20 km/h into jammed flow, 20-40 km/h into saturated flow, 40 km/h above into free flow. There is not a clear standard regarding which division way is more suitable, therefore, we adopt the popular division way in this paper, namely, 40 km/h and above are divided into free flow, 10-40 km/h is divided into saturated flow, and 10 km/h is divided into jammed flow. The division way of traffic state corresponding to average speed are shown in Table 1.

TABLE 1. The division way of traffic state corresponding to average speed.

Average speed	40 km/h above	10-40 km/h	0-10 km/h
Traffic state	free flow	saturated flow	jammed flow

Chongqing Yangtze River Bridge in China is selected as the road section for traffic noise data acquisition. This road section includes four lanes, one of them is a bus lane. The traffic noise data acquisition scene is shown in Figure 8. The traffic noise data corresponding to three different traffic states are all acquired from this same road section. The number of vehicles per lane are not the same. We obtained the acoustic data in a four-lane road segment, namely we take into account the traffic noise in a set of lanes. Therefore, the number of vehicles per lane does not affect the acoustic analysis.

When acquiring the traffic noise from the expected road section, the acquisition data will be greatly disturbed by the

(a) Acquisition location (b) Acquisition diagram

FIGURE 8. Traffic noise data acquisition scene.

traffic noise of the opposite lanes, the wind noise blowed from the surface of the river, and the traffic noise of other road sections under the bridge, so the traffic noise acquired from the expected road section is more complex. In some areas or segments of city, the surrounding environment is quiet, while others are noisy. However, no matter how the surrounding environment changes, acoustic characteristics is the key factor affecting the accuracy of traffic state evaluation. Therefore, the traffic state evaluation based on traffic noise in different surrounding environments, the problem can still be regarded as the selection of acoustic characteristics, which is also the research content of this paper. If the surrounding environment of segment is complex enough, then we just need to extract a new more robust acoustic characteristics.

When using the acoustic acquisition equipment to collect the traffic noise, the equipment is fixed to the top of the tripod. At present, there is not a specific standard for the placement of tripod, generally, it can be placed within one meter from the edge of the road section. Considering the pedestrians on the bridge, the tripod is placed at the edge of the road section. In order to better acquire the traffic noise, the microphone should face the approaching vehicles. The format of the collected audio data is ''.wav'', and the sampling frequency is 48 kHz. The traffic noise acquired for five seconds under each traffic state is shown in Figure 9, where the class label represents the traffic state corresponding to the traffic noise, namely, ''1'' represents free flow, ''2'' represents saturated flow, and ''3'' represents jammed flow. The traffic noise signal in each traffic state is taken 5 seconds, so we collect 15 seconds of traffic noise for three traffic states.

Nowadays, radar velocimeter is widely used in vehicle speed detection, so we use hand-held radar velocimeter to measure the speed of the passing vehicle. It should be noted that speed measurement and traffic noise collection are carried out at the same time, because we do not know which traffic state the collected traffic noise belongs to before finishing data acquisition.

B. TRAFFIC STATE EVALUATION RESULTS BASED ON MFCC CHARACTERISTIC

1) MFCC CHARACTERISTIC EXTRACTION

In speech recognition, 25 ms is generally selected as the frame length, and the frame shift is generally set as one half of the frame length. In view of this, we also set the frame length of traffic noise equal to 25 ms (namely 1200 sample points)

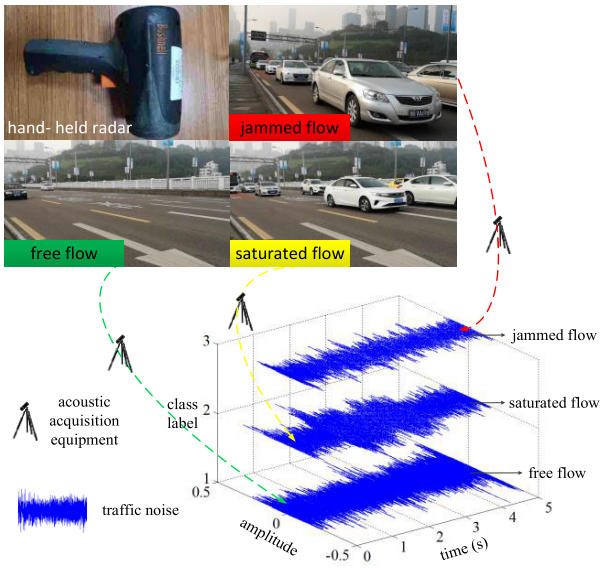


FIGURE 9. Traffic noise data acquisition under different traffic states.

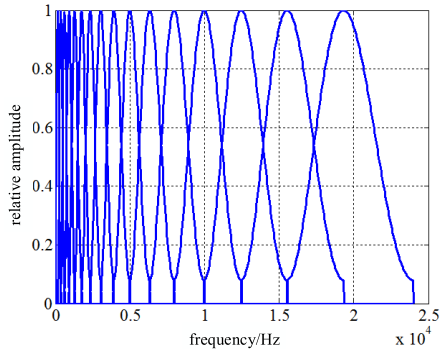


FIGURE 10. Frequency response curve of Mel filter bank.

and frame shift equal to 12.5 ms (namely 600 sample points). Generally, the MFCC characteristic dimension is set equal to 8, correspondingly, the number of Mel filters should be set equal to twice of MFCC characteristic dimension. The frequency response curve of Mel filter bank using hamming window function is shown as Figure 10.

Figure 11 shows the fractal dimension visualization graph of MFCC characteristic under different traffic states, where Mi ($i = 1, 2, \cdots, p$) represents the ith dimension attribute of MFCC characteristic, here, the parameter p is 8. It can be seen from Figure 11 that in the same dimension, the attribute values of saturated flow and jammed flow are close, but different from those of free flow.

The class label of traffic state corresponding to each frame of MFCC is shown as Figure 12. There are 399 frames of MFCC in each traffic state, so the total number of frames of MFCC is 1197.

2) TRAFFIC STATE EVALUATION RESULTS

In the field of digital signal processing, the normalization method is usually utilized to process sample points into values in [0, 1] interval, so as to reduce the differences between different data levels. The MFCC characteristic is normalized according to formula (42), where m_{ij} denotes the ith dimension attribute value of the acoustic characteristic of the jth frame, n represents the number of frames, p represents the acoustic characteristic dimension.

$$m_{i,j} = \frac{m_{i,j} - \min\limits_{1\le j\le n} m_{i,j}}{\max\limits_{1\le j\le n} m_{i,j} - \min\limits_{1\le j\le n} m_{i,j}} \tag{42}$$

The definition of comprehensive evaluation accuracy r_{all} of traffic state is shown in formula (43), where $n^{(i)}_{\text{correct}}$ represents the number of correctly identified test sample frames in the traffic state corresponding to label i, $n^{(i)}$ represents the total number of test sample frames in the traffic state corresponding to label i ($i = 1, 2, 3$), and N denotes the number of traffic state classification. Therefore, r_{all} can be used to compare the advantages and disadvantages of different traffic state evaluation methods.

$$r_{\text{all}} = \left(\sum_{i=1}^{N} n^{(i)}_{\text{correct}} \Big/ \sum_{i=1}^{N} n^{(i)}\right) \times 100\% \tag{43}$$

The first 200 frames of MFCC of each traffic state are selected as the training sets, and the last 199 frames of MFCC of each traffic state are selected as the testing sets. Therefore, the total number of training frames is 600, and the total number of testing frames is 597. The first 200 frames of MFCC of each traffic state is used to train the SVM classifier. PSO algorithm is introduced to optimize the SVM parameters, and the optimal parameter c_{opt} and g_{opt} are shown in Figure 13 (a), where CVAccuracy represents the accuracy of interactive verification under the current optimal parameters, at this time, CVAccuracy is 90.167%. CVAccuracy is also called training accuracy. The trained SVM classifier is used to predict the last 199 frames of MFCC of each traffic state, and the comprehensive evaluation accuracy of traffic state based on MFCC characteristic is 81.742%, as shown in Figure 13 (b).

Let n_{correct} represent the number of correctly identified testing samples, and n_{error} represent the number of incorrectly identified testing samples. Evaluation results of three different traffic states based on MFCC characteristic are shown in Figure 14.

The evaluation results of free flow based on MFCC characteristic are shown in Figure 14 (a) - (d). n_{correct} is 50 and n_{error} is 0 in frame 1-50; n_{correct} is 50 and n_{error} is 0 in frame 51-100; n_{correct} is 47 and n_{error} is 3 in frame 101-150; n_{correct} is 45 and n_{error} is 4 in frame 151-199. So, n_{correct} of free flow is $50 + 50 + 47 + 45 = 192$, and n_{error} of free flow is $0 + 0 + 3 + 4 = 7$.

The evaluation results of saturated flow based on MFCC characteristic are shown in Figure 14 (e) – (h). n_{correct} is 43 and n_{error} is 7 in frame 200-249; n_{correct} is 38 and n_{error} is 12 in frame 250-299; n_{correct} is 27 and n_{error} is 23 in frame 300-349; n_{correct} is 7 and n_{error} is 42 in frame 350-398. So, n_{correct} of saturated flow is $43+38+27+7 = 115$, and n_{error} of saturated flow is $7 + 12 + 23 + 42 = 84$.

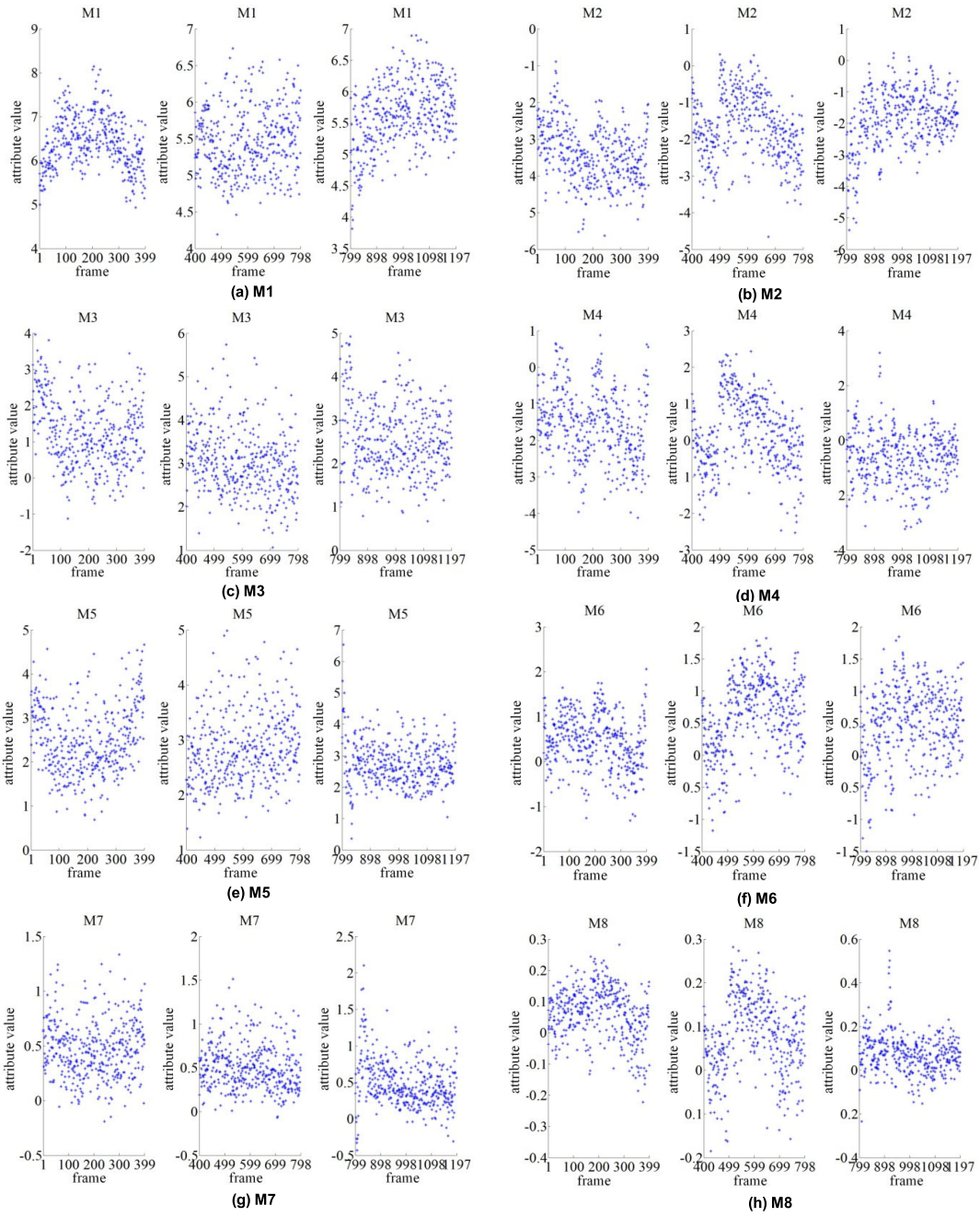


FIGURE 11. Fractal dimension visualization graph of MFCC characteristic.

The evaluation results of jammed flow based on MFCC characteristic are shown in Figure 14 (i) – (l). n_{correct} is 46 and n_{error} is 4 in frame 399-448; n_{correct} is 48 and n_{error} is 2 in frame 449-498; n_{correct} is 43 and n_{error} is 7 in frame 499-548; n_{correct} is 44 and n_{error} is 5 in frame 549-597. So, n_{correct} of jammed flow is $46 + 48 + 43 + 44 = 181$, and n_{error} of jammed flow is $4 + 2 + 7 + 5 = 18$.

In order to better observe the evaluation results of three different traffic states based on MFCC, according to the statistical results of Figure 14, the evaluation accuracies of each traffic state are calculated as shown in Table 2.

3) LIMITATIONS

It can be seen from Table 2 that the evaluation accuracy of free flow is 96.482%, and that of jammed flow is 90.955%. However, the evaluation accuracy of saturated flow is only 57.789%. Because of the low evaluation accuracy of saturated flow, the comprehensive evaluation accuracy of traffic state is also low. Therefore, the traditional MFCC characteristic is relatively strong in characterizing free flow and jammed flow, but very weak in characterizing saturated flow. In addition, saturated flow itself is a transition state from free flow to jammed flow, thus its characteristics are similar to the other

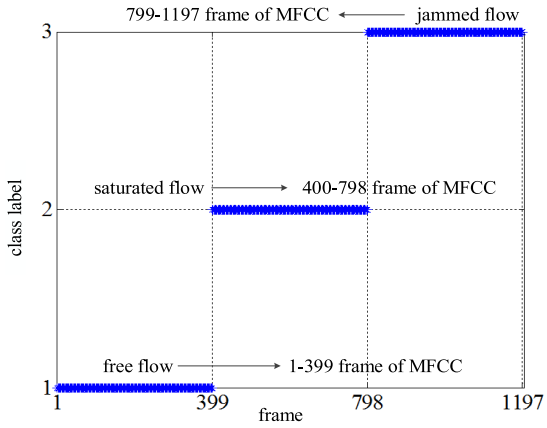


FIGURE 12. Traffic state corresponding to each frame of MFCC.

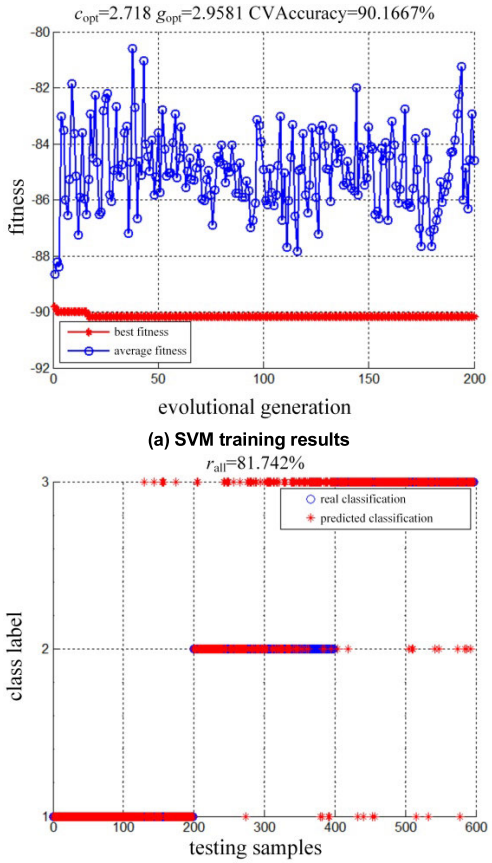


FIGURE 13. Traffic state evaluation results based on MFCC characteristic.

two states. Because the traffic noise is acquired on the same road section, the similarity are stronger. In summary, the recognition of saturated flow is generally more difficult than the other two states, and the MFCC characteristic performs poorly when characterizing saturated flow.

C. TRAFFIC STATE EVALUATION RESULTS BASED ON T-MFCC CHARACTERISTIC

1) T-MFCC CHARACTERISTIC EXTRACTION

Figure 15 shows the fractal dimension visualization graph of T-MFCC characteristic under different traffic states, where

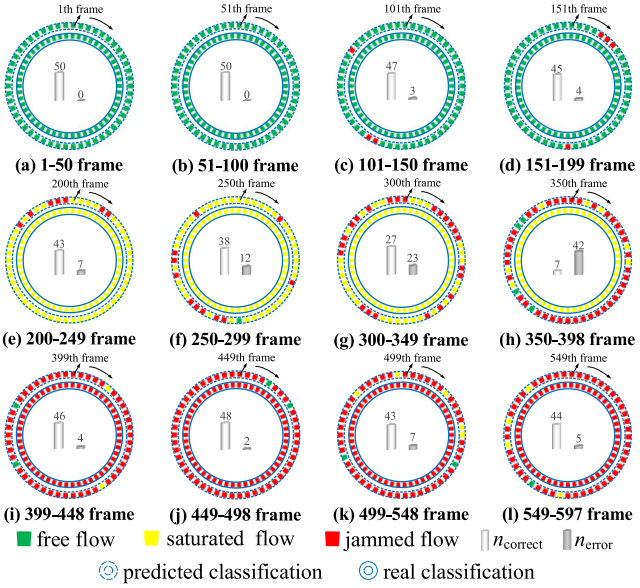


FIGURE 14. Evaluation results of three different traffic states based on MFCC characteristic.

TABLE 2. Evaluation accuracies of three different traffic states based on MFCC characteristic.

Traffic state	n_{correct}	n_{error}	accuracy
free flow	192	7	96.482%
saturated flow	115	84	57.789%
jammed flow	181	18	90.955%

TMi ($i = 1, 2, \cdots, p$) represents the ith dimension attribute of T-MFCC characteristic, here, the parameter p is 8. It can be seen from Figure 15 that in the same dimension, the attribute values of different traffic states are quite different.

The class label of traffic state corresponding to each frame of T-MFCC is shown as Figure 16. There are 399 frames of T-MFCC in each traffic state, so the total number of frames of T-MFCC is 1197.

2) TRAFFIC STATE EVALUATION RESULTS

The first 200 frames of T-MFCC of each traffic state are selected as the training sets, and the last 199 frames of T-MFCC of each traffic state are selected as the testing sets. Therefore, the total number of training frames is 600, and the total number of testing frames is 597. The first 200 frames of T-MFCC of each traffic state is used to train the SVM classifier. PSO algorithm is introduced to optimize the SVM parameters, and the optimal parameter c_{opt} and g_{opt} are shown in Figure 17 (a). The trained SVM classifier is used to predict the last 199 frames of T-MFCC of each traffic state, and the comprehensive evaluation accuracy of traffic state based on T-MFCC characteristic is 85.427%, as shown in Figure 17 (b).

Let n_{correct} represent the number of correctly identified testing samples, and n_{error} represent the number of incorrectly identified testing samples. Evaluation results of three different traffic states based on T-MFCC characteristic are shown in Figure 18.

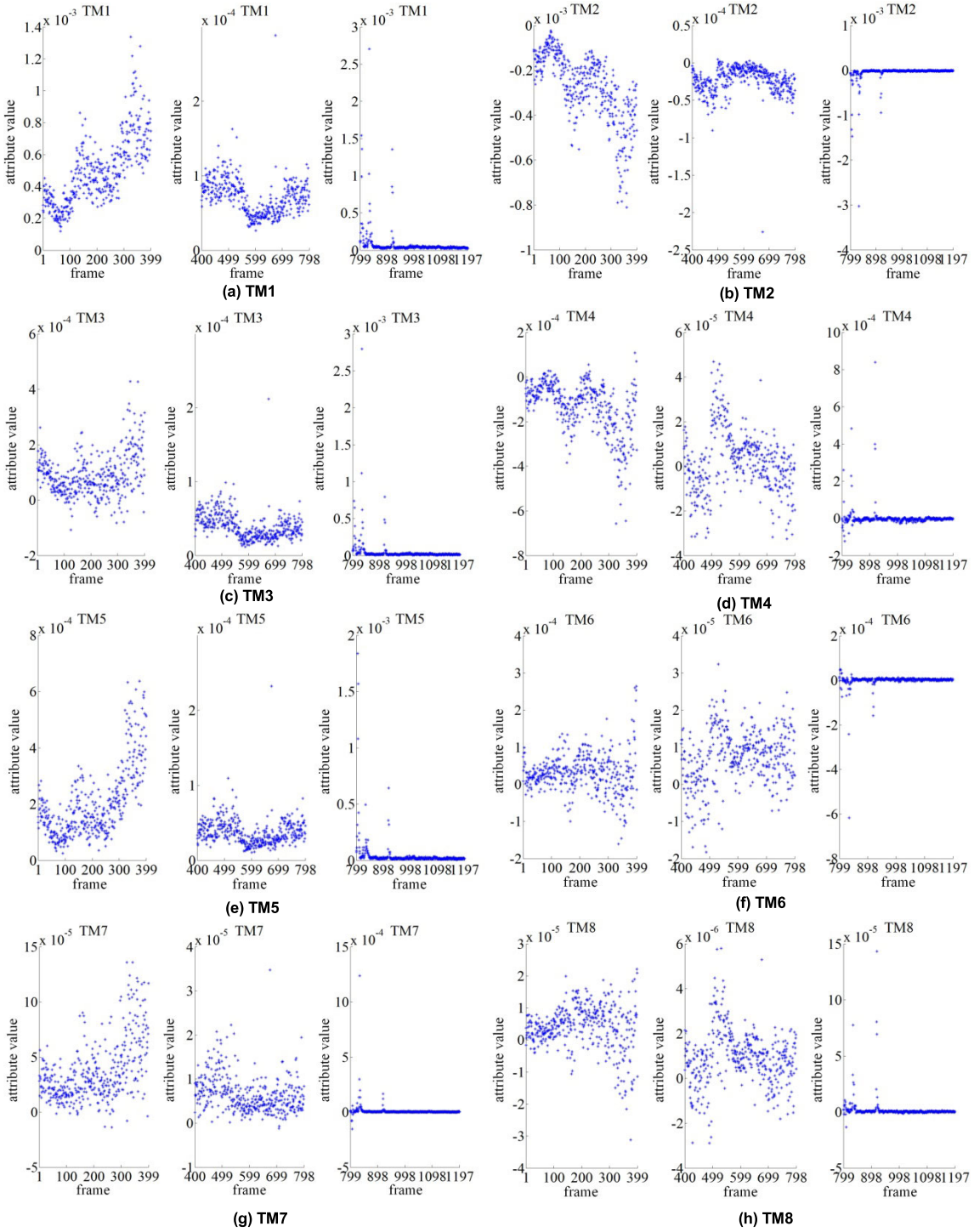


FIGURE 15. Fractal dimension visualization graph of T-MFCC characteristic.

The evaluation results of free flow based on T-MFCC characteristic are shown in Figure 18 (a) - (d). n_{correct} is 50 and n_{error} is 0 in frame 1-50; n_{correct} is 50 and n_{error} is 0 in frame 51-100; n_{correct} is 34 and n_{error} is 16 in frame 101-150; n_{correct} is 32 and n_{error} is 17 in frame 151-199. So, n_{correct} of free flow is $50 + 50 + 34 + 32 = 166$, and n_{error} of free flow is $0 + 0 + 16 + 17 = 33$.

The evaluation results of saturated flow based on T-MFCC characteristic are shown in Figure 18 (e) - (h). n_{correct} is 43 and n_{error} is 7 in frame 200-249; n_{correct} is 38 and n_{error} is 12 in frame 250-299; n_{correct} is 40 and n_{error} is 10 in frame 300-349; n_{correct} is 26 and n_{error} is 23 in frame 350-398. So, n_{correct} of saturated flow is $43 + 38 + 40 + 26 = 147$, and n_{error} of saturated flow is $7 + 12 + 10 + 23 = 52$.

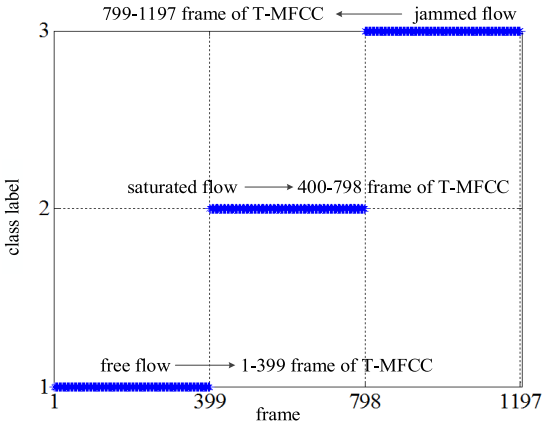
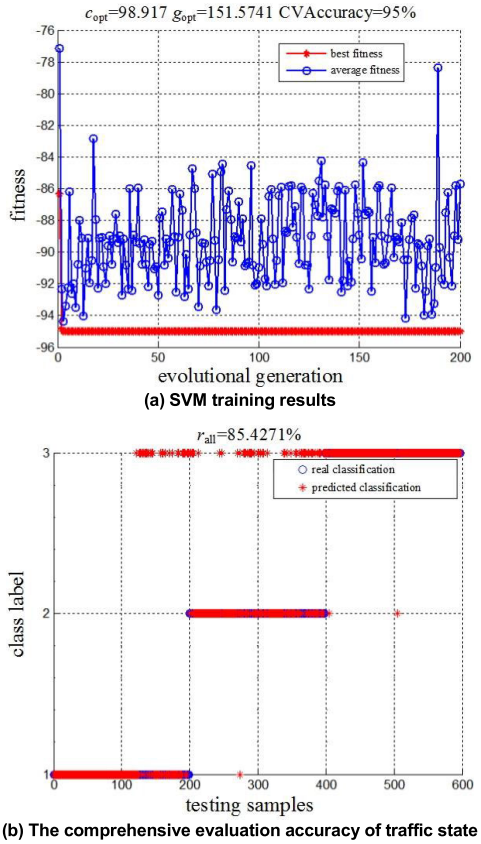


FIGURE 16. Traffic state corresponding to each frame of T-MFCC.



(a) SVM training results

(b) The comprehensive evaluation accuracy of traffic state

FIGURE 17. Traffic state evaluation results based on T-MFCC characteristic.

The evaluation results of jammed flow based on T-MFCC characteristic are shown in Figure 18 (i) - (l). n_{correct} is 49 and n_{error} is 1 in frame 399-448; n_{correct} is 50 and n_{error} is 0 in frame 449-498; n_{correct} is 49 and n_{error} is 1 in frame 499-548; n_{correct} is 49 and n_{error} is 0 in frame 549-597. So, n_{correct} of jammed flow is $49+50+49+49=197$, and n_{error} of jammed flow is $1+0+1+0=2$.

In order to better observe the evaluation results of three different traffic states based on T-MFCC, according to the statistical results of Figure 18, the evaluation accuracies of each traffic state are calculated as shown in Table 3.

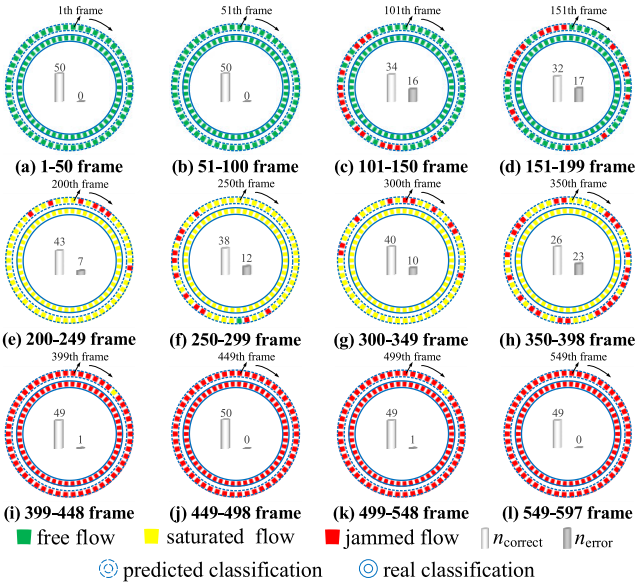


FIGURE 18. Evaluation results of three different traffic states based on T-MFCC characteristic.

TABLE 3. Evaluation accuracies of three different traffic states based on T-MFCC characteristic.

Traffic state	n_{correct}	n_{error}	accuracy
free flow	166	33	83.417%
saturated flow	147	52	73.869%
jammed flow	197	2	98.995%

D. TRAFFIC STATE EVALUATION RESULTS BASED ON PT-MFCC CHARACTERISTIC

1) PT-MFCC CHARACTERISTIC EXTRACTION

On the basis of T-MFCC, the PT-MFCC characteristic is extracted. In the process of the PT-MFCC characteristic extraction, it is found that the first five principal components are the most appropriate attributes for PT-MFCC. Figure 19 shows the fractal dimension visualization graph of PT-MFCC characteristic under different traffic states, where PTMi ($i = 1, 2, \ldots, p$) represents the ith dimension attribute of PT-MFCC characteristic, here, the parameter p is 5. It can be seen from Figure 19 that in the same dimension, the attribute values of saturated flow and free flow are close, but different from those of jammed flow.

The class label of traffic state corresponding to each frame of PT-MFCC is shown as Figure 20. There are 399 frames of PT-MFCC in each traffic state, so the total number of frames of PT-MFCC is 1197.

2) TRAFFIC STATE EVALUATION RESULTS

The first 200 frames of PT-MFCC of each traffic state are selected as the training sets, and the last 199 frames of PT-MFCC of each traffic state are selected as the testing sets. Therefore, the total number of training frames is 600, and the total number of testing frames is 597. The first 200 frames of PT-MFCC of each traffic state is used to train the SVM classifier. PSO algorithm is introduced to optimize the

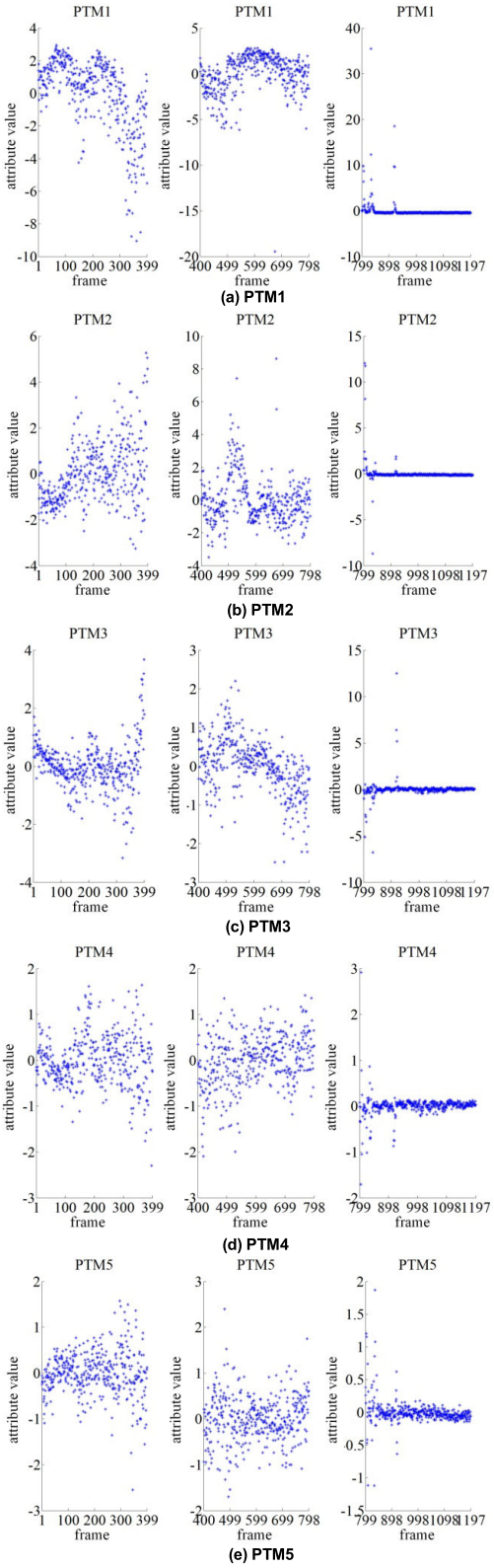


FIGURE 19. Fractal dimension visualization graph of PT-MFCC characteristic.

SVM parameters, and the optimal parameter c_{opt} and g_{opt} are shown in Figure 21 (a). The trained SVM classifier is used to predict the last 199 frames of PT-MFCC of each traffic state,

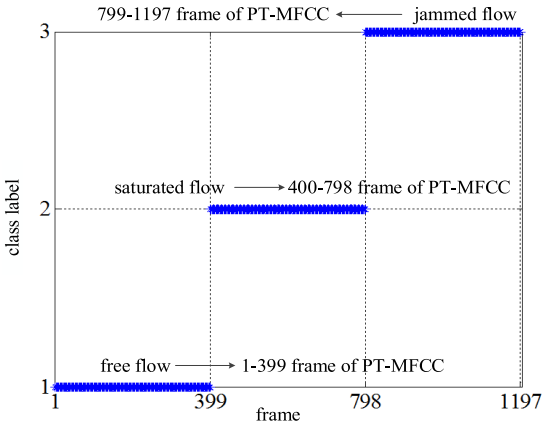


FIGURE 20. Traffic state corresponding to each frame of PT-MFCC.

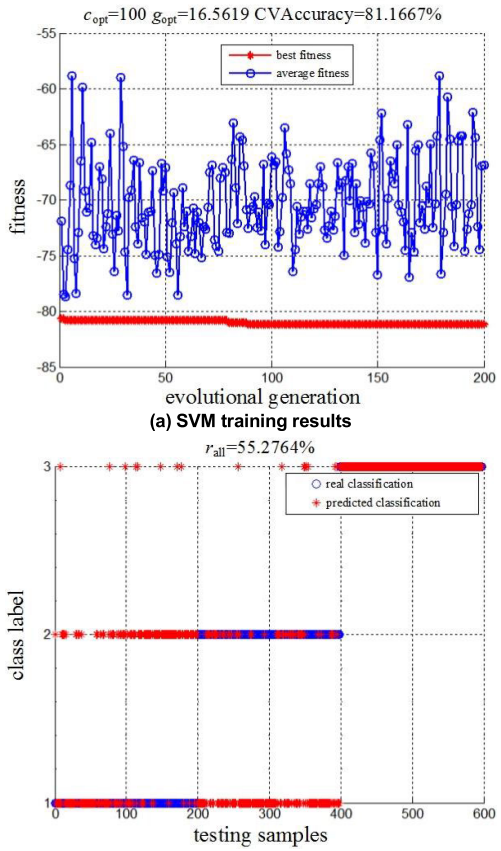


FIGURE 21. Traffic state evaluation results based on PT-MFCC characteristic.

and the comprehensive evaluation accuracy of traffic state based on PT-MFCC characteristic is 55.276%, as shown in Figure 21 (b).

Let n_{correct} represent the number of correctly identified testing samples, and n_{error} represent the number of incorrectly identified testing samples. Evaluation results of three different traffic states based on PT-MFCC characteristic are shown in Figure 22.

The evaluation results of free flow based on PT-MFCC characteristic are shown in Figure 22 (a) - (d). n_{correct} is 38

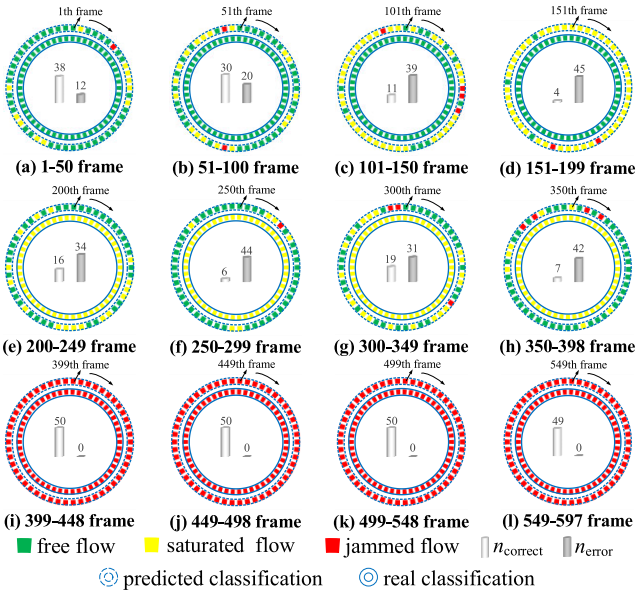


FIGURE 22. Evaluation results of three different traffic states based on PT-MFCC characteristic.

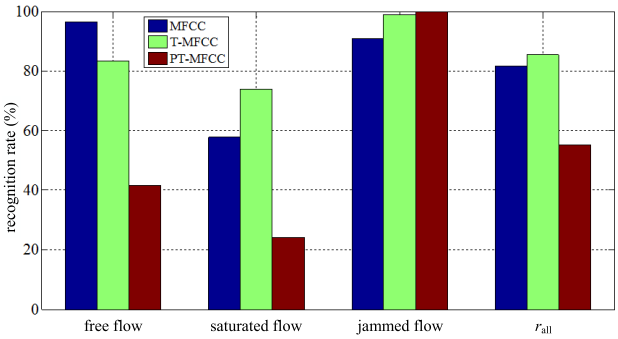


FIGURE 23. Histogram of traffic state evaluation accuracy based on different traffic noise characteristics.

and n_{error} is 12 in frame 1-50; $n_{correct}$ is 30 and n_{error} is 20 in frame 51-100; $n_{correct}$ is 11 and n_{error} is 39 in frame 101-150; $n_{correct}$ is 4 and n_{error} is 45 in frame 151-199. So, $n_{correct}$ of free flow is $38 + 30 + 11 + 4 = 83$, and n_{error} of free flow is $12 + 20 + 39 + 45 = 116$.

The evaluation results of saturated flow based on PT-MFCC characteristic are shown in Figure 22 (e) - (h). $n_{correct}$ is 16 and n_{error} is 34 in frame 200-249; $n_{correct}$ is 6 and n_{error} is 44 in frame 250-299; $n_{correct}$ is 19 and n_{error} is 31 in frame 300-349; $n_{correct}$ is 7 and n_{error} is 42 in frame 350-398. So, $n_{correct}$ of saturated flow is $16 + 6 + 19 + 7 = 48$, and n_{error} of saturated flow is $34 + 44 + 31 + 42 = 151$.

The evaluation results of jammed flow based on PT-MFCC characteristic are shown in Figure 22 (i) - (l). $n_{correct}$ is 50 and n_{error} is 0 in frame 399-448; $n_{correct}$ is 50 and n_{error} is 0 in frame 449-498; $n_{correct}$ is 50 and n_{error} is 0 in frame 499-548; $n_{correct}$ is 49 and n_{error} is 0 in frame 549-597. So, $n_{correct}$ of jammed flow is $50 + 50 + 50 + 49 = 199$, and n_{error} of jammed flow is $0 + 0 + 0 + 0 = 0$.

In order to better observe the evaluation results of three different traffic states based on PT-MFCC, according to the statistical results of Figure 22, the evaluation accuracies of each traffic state are calculated as shown in Table 4. It can be seen from Figure 19 that in the same dimension, the attribute values corresponding to the jammed flow are quite different from those corresponding to the other two states, and the attribute values corresponding to the jammed flow are almost the same, meanwhile, the attribute values of saturated flow and free flow are close. Therefore, the congested flow is with high accuracy and the other flows with very low accuracy.

TABLE 4. Evaluation accuracies of three different traffic states based on PT-MFCC characteristic.

Traffic state	$n_{correct}$	n_{error}	accuracy
free flow	83	116	41.709%
saturated flow	48	151	24.121%
jammed flow	199	0	100%

TABLE 5. Evaluation accuracies of traffic state based on three different traffic noise characteristics (%).

Characteristic	Free flow	Saturated flow	Jammed flow	r_{all}
MFCC	96.482	57.789	90.955	81.742
T-MFCC	83.417	73.869	98.995	85.427
PT-MFCC	41.709	24.121	100	55.276

E. RESULTS ANALYSIS

The traffic state evaluation accuracies acquired under three different traffic noise characteristics are summarized as shown in Table 5.

Compared with the MFCC characteristic, the recognition rate of free flow corresponding to the T-MFCC characteristic is reduced by 13.065%, the recognition rate of saturated flow is increased by 16.080%, the recognition rate of jammed flow is increased by 8.040%, and the comprehensive evaluation accuracy is improved by 3.685%. Compared with the MFCC characteristic, the recognition rate of free flow corresponding to PT-MFCC characteristic is reduced by 54.773%, the recognition rate of saturated flow is reduced by 33.668%, the recognition rate of jammed flow is increased by 9.045%, and the comprehensive evaluation accuracy is reduced by 26.466%. Compared with PT-MFCC characteristic, the recognition rate of free flow corresponding to T-MFCC characteristic is increased by 41.708%, the recognition rate of saturated flow is increased by 49.748%, the recognition rate of jammed flow is reduced by 1.005%, and the comprehensive evaluation accuracy is improved by 30.151%. In order to observe the evaluation results of traffic state more intuitively, data in Table 5 is drawn in the form of histogram, as shown in Figure 23.

Compared with the traditional MFCC characteristic, although the recognition rate of free flow based on T-MFCC characteristic reduces, it still reached more than 80%, in addition, the recognition rates of saturated flow and jammed flow both increase greatly, and the comprehensive evaluation accuracy also increases. Therefore, the T-MFCC characteristic is

superior to the MFCC characteristic, and the MFCC characteristic is superior to the PT-MFCC characteristic.

VII. CONCLUSION

Aiming at the problem that MFCC characteristic of traffic noise in conditions of the same road section and complex noise environment has weak characterization abilities for traffic state (especially saturated flow), two new characteristic extraction algorithms of traffic noise are proposed, namely T-MFCC characteristic extraction algorithm and PT-MFCC characteristic extraction algorithm. T-MFCC characteristic is composed of TEO characteristic and MFCC characteristic. The experimental results show that the T-MFCC characteristic is superior to the MFCC characteristic, especially, the T-MFCC characteristic has a great contribution to the improvement of recognition accuracy of saturated flow. Then PCA is used to reduce the dimension of T-MFCC, so the PT-MFCC characteristic is acquired. However, according to the experimental results, the PT-MFCC characteristic performs much worse than the MFCC characteristic and the T-MFCC characteristic, that is, there are no redundant attributes in T-MFCC characteristic, so it is not needed to utilize the PCA algorithm to reduce the dimensionality of T-MFCC characteristic. Therefore, T-MFCC characteristic owns stronger abilities than other two characteristics to characterize traffic state, especially saturated flow.

In the next work, we will study the traffic state evaluation in conditions of some more complex circumstances, such as thunder, rain, hail and other noises caused by traffic accidents.

REFERENCES

[1] D. Sirohi, N. Kumar, and P. S. Rana, "Convolutional neural networks for 5G-enabled intelligent transportation system: A systematic review," *Comput. Commun.*, vol. 153, pp. 459–498, Mar. 2020.

[2] S. Zhang, H. Wang, W. Quan, and X. Liu, "Measuring variability in freeway traffic states using real-time loop data in jilin," *Procedia-Social Behav. Sci.*, vol. 96, pp. 2676–2683, Nov. 2013.

[3] T. Tsubota, A. Bhaskar, A. Nantes, E. Chung, and V. V. Gayah, "Comparative analysis of traffic state estimation: Cumulative counts-based and trajectory-based methods," *Transp. Res. Rec.*, vol. 2491, no. 1, pp. 43–52, Jan. 2015.

[4] M. Gramaglia, M. Calderon, and C. J. Bernardos, "Abeona monitored traffic: VANET-assisted cooperative traffic congestion forecasting," *IEEE Veh. Technol. Mag.*, vol. 9, no. 2, pp. 50–57, Jun. 2014.

[5] E. S. Canepa and C. G. Claudel, "A dual model/artificial neural network framework for privacy analysis in traffic monitoring systems," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 126–144, Aug. 2019.

[6] X. Shi, Z. Shan, and N. Zhao, "Learning for an aesthetic model for estimating the traffic state in the traffic video," *Neurocomputing*, vol. 181, pp. 29–37, Mar. 2016.

[7] X. Li, Y. She, D. Luo, and Z. Yu, "A traffic state detection tool for freeway video surveillance system," *Procedia-Social Behav. Sci.*, vol. 96, pp. 2453–2461, Nov. 2013.

[8] R. Ke, Z. Li, S. Kim, J. Ash, Z. Cui, and Y. Wang, "Real-time bidirectional traffic flow parameter estimation from aerial videos," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 890–901, Apr. 2017.

[9] A. B. Habtie, A. Abraham, and D. Midekso, "Artificial neural network based real-time urban road traffic state estimation framework," in *Computational Intelligence in Wireless Sensor Networks*. Cham, Switzerland: Springer, 2017, pp. 73–97.

[10] N. Lefebvre, X. Chen, P. Beauseroy, and M. Zhu, "Traffic flow estimation using acoustic signal," *Eng. Appl. Artif. Intell.*, vol. 64, pp. 164–171, Sep. 2017.

[11] A. Wolfindale, G. Dunne, and S. J. Walsh, "Vehicle noise primary attribute balance," *Appl. Acoust.*, vol. 73, no. 4, pp. 386–394, Apr. 2012.

[12] A. Starzacher and B. Rinner, "Single sensor acoustic feature extraction for embedded realtime vehicle classification," in *Proc. Int. Conf. Parallel Distrib. Comput., Appl. Technol.*, Higashi Hiroshima, Japan, Dec. 2009, pp. 378–383.

[13] P. Borkar and L. G. Malik, "Acoustic signal based traffic density state estimation using SVM," *Int. J. Image, Graph. Signal Process.*, vol. 5, no. 8, pp. 37–44, Jun. 2013.

[14] R. Sen, B. Raman, and P. Sharma, "Horn-ok-please," in *Proc. 8th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, San Francisco, CA, USA, Jun. 2010, pp. 137–150.

[15] R. Sen, P. Siriah, and B. Raman, "RoadSoundSense: Acoustic sensing based road congestion monitoring in developing regions," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Salt Lake City, UT, USA, Jun. 2011, pp. 125–133.

[16] V. P. Warghade and M. S. Deshpande, "Road traffic condition estimation based on road acoustics," in *Proc. Int. Conf. Comput., Commun., Control Autom. (ICCUBEA)*, Aug. 2017, pp. 1–5.

[17] P. Jadhav, S. D. Sawarkar, and D. J. Pete, "Roadside acoustic signals based road traffic density estimation," in *Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2018, pp. 1–5.

[18] V. Tyagi, S. Kalyanaraman, and R. Krishnapuram, "Vehicular traffic density state estimation based on cumulative road acoustics," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1156–1166, Sep. 2012.

[19] A. Kaur, N. Sood, N. Aggarwal, D. Vij, and B. Sachdeva, "Traffic state detection using smartphone based acoustic sensing," *J. Intell. Fuzzy Syst.*, vol. 32, no. 4, pp. 3159–3166, Mar. 2017.

[20] E. Alexandre, L. Cuadra, S. Salcedo-Sanz, A. Pastor-Sánchez, and C. Casanova-Mateo, "Hybridizing extreme learning machines and genetic algorithms to select acoustic features in vehicle classification applications," *Neurocomputing*, vol. 152, pp. 58–68, Mar. 2015.

[21] M. P. Paulraj, A. H. Adom, S. Sundararaj, and N. B. A. Rahim, "Moving vehicle recognition and classification based on time domain approach," *Procedia Eng.*, vol. 53, pp. 405–410, Jan. 2013.

[22] A. Wieczorkowska, E. Kubera, T. Słowik, and K. Skrzypiec, "Spectral features for audio based vehicle and engine classification," *J. Intell. Inf. Syst.*, vol. 50, no. 2, pp. 265–290, Apr. 2018.

[23] S. Kozhisseri and M. Bikdash, "Spectral features for the classification of civilian vehicles using acoustic sensors," in *Proc. IEEE Workshop Comput. Intell. Vehicles Veh. Syst.*, Mar./Apr. 2009, pp. 93–100.

[24] O. Karpis, "System for vehicles classification and emergency vehicles detection," *IFAC Proc. Vols.*, vol. 45, no. 7, pp. 186–190, 2012.

[25] A. Y. Nooralahiyan, H. R. Kirby, and D. McKeown, "Vehicle classification by acoustic signature," *Math. Comput. Model.*, vol. 27, nos. 9–11, pp. 205–214, May 1998.

[26] P. Marmaroli, J.-M. Odobez, X. Falourd, and H. Lissek, "A bimodal sound source model for vehicle tracking in traffic monitoring," in *Proc. Eur. Signal Process. Conf.*, Aug./Sep. 2011, pp. 1327–1331.

[27] S. Ishida, J. Kajimura, M. Uchino, S. Tagashira, and A. Fukuda, "SAVeD: Acoustic vehicle detector with speed estimation capable of sequential vehicle detection," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 906–912.

[28] R. Lopez-Valcarce, "Broadband analysis of a microphone array based road traffic speed estimator," in *Proc. Process. Workshop Sensor Array Multichannel Signal*, 2004, pp. 609–612.

[29] Z. Zou, Q. Ma, F. Liu, and S. Ullah, "A novel acoustic characteristic extraction algorithm for traffic volume estimation," *IEEE Access*, vol. 7, pp. 168899–168913, Nov. 2019.

[30] K. Kim, "Lightweight filter architecture for energy efficient mobile vehicle localization based on a distributed acoustic sensor network," *Sensors*, vol. 13, no. 9, pp. 11314–11335, Aug. 2013.

[31] A. Schclar, A. Averbuch, N. Rabin, V. Zheludev, and K. Hochman, "A diffusion framework for detection of moving vehicles," *Digit. Signal Process.*, vol. 20, no. 1, pp. 111–122, Jan. 2010.

[32] T. Toyoda, N. Ono, S. Miyabe, T. Yamada, and S. Makino, "Traffic monitoring with ad-hoc microphone array," in *Proc. 14th Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 318–322.

[33] A. J. Torija and D. P. Ruiz, "Using recorded sound spectra profile as input data for real-time short-term urban road-traffic-flow estimation," *Sci. Total Environ.*, vols. 435–436, pp. 270–279, Oct. 2012.

[34] D. Mirri, G. Iuculano, P. A. Traverso, and G. Pasini, "Performance evaluation of cascaded rectangular windows in spectral analysis," *Measurement*, vol. 36, no. 1, pp. 37–52, Jul. 2004.

[35] J. Barros and R. I. Diego, "On the use of the Hanning window for harmonic analysis in the standard framework," *IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 538–539, Jan. 2006.
[36] A. Basit, I. M. Qureshi, W. Khan, S. U. Rehman, and M. M. Khan, "Beam pattern synthesis for an FDA radar with Hamming window-based nonuniform frequency offset," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 2283–2286, Jun. 2017.
[37] H. Zhao and L. Li, "Fault diagnosis of wind turbine bearing based on variational mode decomposition and teager energy operator," *IET Renew. Power Gener.*, vol. 11, no. 4, pp. 453–460, Mar. 2017.
[38] S. Yi, Z. Lai, Z. He, Y.-M. Cheung, and Y. Liu, "Joint sparse principal component analysis," *Pattern Recognit.*, vol. 61, pp. 524–536, Jan. 2017.

QINGLU MA received the Ph.D. degree in computer technology and science from Chongqing University, Chongqing, China, in 2012. He is currently an Associate Professor with Chongqing Jiaotong University. He is the co-inventor of ten issued Chinese patents and coauthored 30 articles. His current research interests include intelligent transportation and safety, big data processing technology, cloud computing, and the Internet of Things.

ZHENG ZOU received the bachelor's degree in traffic control and information engineering from Chongqing Jiaotong University, Chongqing, China, in 2016, where he is currently pursuing the master's degree. His current research interests include traffic control, acoustic signal processing, and traffic parameters extraction.