Research on Feature Recognition of UAV Acoustic Signal Based on SVM

Hengkang Jin¹ and Yiwen Zhang²

¹College of Information Engineering, Engineering University of PAP, Xi’an, China, 573905527@qq.com
²College of Information Engineering, Engineering University of PAP, Xi’an, China, 137477045@qq.com

Abstract. At present, the analysis of UAV flight acoustic signals is mainly based on traditional speech signal processing methods, and has not been analyzed in depth. According to the flight signal of UAV, combined with the aerodynamic characteristics of UAV, the characteristics of UAV’s acoustic signal are analyzed. The three feature extraction algorithms of pitch period, FFT and Mel Cepstral Coefficient (MFCC) are analyzed and compared. Feature extraction is performed, and a support vector machine (SVM) classification algorithm is applied to perform multi-classification model recognition. The measured and experimental results show that based on SVM classification and recognition, the three feature recognition methods all realize the classification of the model. The comprehensive FFT is the best, and the MFCC is the second. The pitch period is not suitable as the feature extraction method alone.

1. Introduction

With the continuous application of consumer-grade UAVs, the problem of “black flying” of drones in daily work and life has become an urgent problem to be solved. In particular, civil aviation systems, major conference sites, and fixed security units are facing huge threats from low-altitude drones. At present, the UAV detection system invented at home and abroad mainly focuses on radio monitoring and acoustic wave monitoring, which can alleviate the problem of "black fly" to a certain extent. However, in the complex electromagnetic environment and the "quiet" state of the drone, there are certain limitations. UAV acoustic signal detection technology is an effective complement to existing detection methods and systems.

UAV sound signals are used as an acoustic signal. At present, most experts and scholars use transplanted speech signal processing methods to study them. Among them, Wang Wei identifies the UAVs with different motion heights by optimizing the MFCC and GMM algorithms; Qiu Yubin et al. proposed a feature extraction algorithm based on modal decomposition (EMD) and MFCC, combined with vector quantization classifiers (VQ) classifies and identifies different types of UAV targets; Xiao Hanchun uses the improved Mel cepstrum coefficients to extract features of low-altitude aircraft to determine whether there are UAVs in the airspace. Most of these studies are based on the classical method MFCC of speech feature extraction, and the difference between UAV and ambient sound is applied. The signal analysis for UAV is only extended by the traditional speech signal method, and the UAV acoustic signal is not analyzed in depth, but other The research and comparison of feature extraction methods, the search for optimal feature extraction methods, and the different types of UAVs are less differentiated.

The feature extraction algorithms of acoustic signals mainly include Fourier transform (FFT), short-time Fourier transform (STFT), pitch period, Mel cepstral coefficient (MFCC), short-time
zero-crossing detection method, etc.. The various feature extraction algorithms have different degrees of complexity, and the characteristics of the acoustic signals are also different. This paper mainly analyzes the aerodynamic characteristics of UAVs combined with aerodynamic characteristics, and explores the problems of different feature extraction algorithms for the recognition and complexity of UAV types.

2. UAV Flight Sound Characteristics

The flight noise of a certain drone during flight aerial photography is collected by actual measurement, and the audio signal characteristic analysis is performed. Analyze its time domain characteristics, as shown in Figure 1(a). The signal strength of the time domain changes with the relative distance of the UAV's flight over the collection point. The distance increases and the signal strength gradually decreases, which is in line with the human's intuitive experience.

The time-frequency characteristics are analyzed, as shown in Figure 1(b). When the drone is hovering, the frequency component is relatively fixed; when the drone is flying relative to the collection point, the high-frequency component is significantly increased, showing a Doppler effect. The data processing of the short-time Fourier transform results of a large number of samples, according to the principle that the frequency components account for 80% of the total energy, it is found that the flight noise is mainly concentrated in the low-frequency components, mainly at frequencies below 1900 Hz.

According to the time-frequency characteristics, the frequency domain characteristics below 2000 Hz are analyzed, as shown in Figure 1(c). The frequency signal exhibits a relationship between the fundamental frequency and the frequency multiplication. The signal has a part of low frequency noise in the frequency range of less than 150 Hz, mainly due to the airflow noise existing in the environment.

![Figure 1. UAV acoustic signal characteristics](image)

Through extensive analysis of the aerodynamic principle and mechanical structure of different types of UAVs, it is found that the acoustic signals generated by the UAV during flight are mainly composed of two parts, one is the mechanical vibration noise during the operation of the motor, and the other is the pneumatic generated by the rotor. Noise [10]. Among them, the running noise of the motor is higher than the aerodynamic noise frequency, and the amplitude is low. Under certain distance conditions, the capturing is difficult. Therefore, the acoustic noise is mainly collected by the acoustic sensor. Among them, the main component of aerodynamic noise is that the rotor produces air during the rotation process.
Therefore, the characteristics of the unmanned aerial sound signal are closely related to the appearance structure of the drone, the size of the rotor, the number of rotors, and the motor speed. Different UAVs have different hardware characteristics. By analyzing the characteristics of UAV's acoustic signals, different UAV models can be distinguished and classified. Using the above features, the UAV sound signature database is built to provide support for UAV detection.

3. Feature Extraction

3.1. Fourier Transform

According to the above analysis, the time-frequency characteristics of the UAV show a certain quasi-stationary characteristic during the hovering process, and the frequency components do not change significantly with time. Therefore, the Fourier transform can be performed on the acquired acoustic signal. FFT key steps:

\[ X(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn} \quad k = 0,1,\ldots,N-1, W_N = e^{-\frac{j2\pi}{N}} \]

Feature extraction is performed according to the spectrum of different UAV hardware differences. Figure 2 shows the spectral characteristics of three different drones.

![Figure 2](image)

**Figure 2** Comparison of frequency domain comparison of different drone acoustic signals

The frequency domain characteristics of the three UAV signals are significantly different, mainly reflecting the fundamental frequency, frequency doubling and frequency amplitude. The main reason is the difference in motor speed and rotor size for the three signals. Therefore, by calculating the FFT of the drone acoustic signal, frequency domain features, especially peak frequency and amplitude, can be extracted as the acoustic characteristics of the model.

3.2. Pitch Period Estimation

The pitch period estimation, also known as pitch detection, is applied to speech signal processing, and the ultimate goal is to find a frequency variation curve that matches the vibration frequency of the vocal cord. The pitch period is an important parameter that effectively describes the vibration of the excitation source. The methods for estimating the pitch period estimation usually include correlation method, cepstrum method and linear prediction method. The autocorrelation method is usually used to estimate the pitch period, which mainly includes the following steps: preprocessing (framing, filtering), threshold setting, clipping processing, cross-correlation to solve the pitch frequency. Among them, the key step short-time autocorrelation function is defined as:
\[ R_n(\tau) = \sum_{m=0}^{N-1} \left[ s(n+m)w(m) \right] \left[ s(n+m+\tau)w(m+\tau) \right] \]

(2)

The period of the autocorrelation function is equal to the original sequence period, the pitch frequency is close to the first formant frequency, and the first formant is extracted as the pitch period estimation result.

3.3. Mel Cepstral Coefficient (MFCC)

The MFCC proposed according to the nonlinear characteristics of human hearing is one of the important algorithms for speech feature extraction. According to the difference of the subjective perception frequency domain, the conversion to the Mel field and the determination of the Mel perception frequency are as follows:

\[ F_{mel} = 1125 \log \left( 1 + \frac{f}{700} \right) \]

(3)

The expression of the MFCC coefficient obtained by discrete cosine transform (DCT) is:

\[ C(n) = \sum_{m=0}^{N-1} s(m) \cos \left( \frac{\pi n (m-0.5)}{M} \right) n=1,2,\ldots,L \]

\[ s(m) = \ln \left( \sum_{k=0}^{N-1} X_a(k) H_m(k) \right)^2 \]

The three feature extraction methods have their own characteristics. The FFT captures the characteristics of the acoustic signal from the perspective of the full spectrum. The MFCC amplifies the local frequency characteristics through the nonlinear Mel transform domain. The pitch period distinguishes different models from the vibration angle of the UAV sound source.

4. Classification and Experiment

4.1. Classification Algorithm

This paper mainly realizes the classification and identification between different UAV models, and Support Vector Machine (SVM) has better performance in solving low sample number, high dimensionality, nonlinearity and local minimum value. At the same time, the optimized SVM also supports better multi-classification, so this paper uses SVM for classification and identification. The commonly used kernel functions of SVM are linear kernel function, polynomial kernel function, radial kernel function and two-layer neural network kernel function. The kernel function uses radial kernel function (RBF), and the expression is:

\[ K(x_i, x_j) = \exp \left( -\frac{|x_i - y_j|^2}{2\sigma^2} \right) \]

(4)

The SVM needs to optimize the parameters \( c \) and \( g \), where the parameter \( c \) (penalty factor), the larger the value, indicates that we pay more attention. The parameter \( g \) is the setting of the gamma function in the kernel function. The values of \( c \) and \( g \) have a great influence on the classification accuracy. In this paper, the function SVMcgForClass is used to determine \( c \) and \( g \). Determine \( c \) and \( g \) in two steps, first rough selection, \( c \) and \( g \) range is, ..., and then finely select, the range of \( c \) is: , ...., the range of \( g \) is: , ...., and finally train and predict the sample.

4.2. Experimental Results and Analysis

The experiment was conducted according to the measured 50 training samples and 100 predicted samples. Through the SVM classification and recognition algorithm, according to different models,
flying heights 0m, 10m, 20m, 30m, 50m, record the time of each feature extraction method, the recognition rate, the time taken to extract features, and the time it takes for the classification algorithm to train and predict. Table 1 shows the recognition rate of each feature extraction method at different flight heights; Table 2 shows the time consuming of each feature extraction method (on a computer with a CPU frequency of 3.3 GHz, quad-core, eight threads, and 8 GB of memory);

### Table 1 Identification rate of each feature extraction method at different heights (%)

| Method    | Drone model | Flight altitude(m) | 0     | 10    | 20    | 30    | 50    |
|-----------|-------------|---------------------|-------|-------|-------|-------|-------|
| FFT       |              |                     |       |       |       |       |       |
| Mavic     | 85.0        | 97.0                | 95.7  | 95.3  | 96.3  |       |       |
| Spark     | 77.0        | 96.3                | 94.3  | 96    | 94.3  |       |       |
| Phantom   | 67.7        | 96.7                | 96.3  | 94.3  | 95.3  |       |       |
| MFCC      |              |                     |       |       |       |       |       |
| Mavic     | 98.3        | 89                  | 96.3  | 90    | 93.3  |       |       |
| Spark     | 94.3        | 91.3                | 92.3  | 93.6  | 94.6  |       |       |
| Phantom   | 93.3        | 92.6                | 94.3  | 94.6  | 94.7  |       |       |
| Pitch period |          |                     |       |       |       |       |       |
| Mavic     | 81.3        | 88.3                | 91.0  | 90.3  | 87.1  |       |       |
| Spark     | 92.1        | 67.7                | 71.3  | 77.3  | 78.7  |       |       |
| Phantom   | 87.7        | 64.3                | 66.3  | 69.3  | 69.7  |       |       |

### Table 2. Time consuming for each feature extraction method (s)

| Method    | Drone model | Flight altitude(m) | Avg. (s) |
|-----------|-------------|---------------------|----------|
|           | Extraction  |                     |          |
| FFT       | 8.97        | 9.76                | 9.46     | 9.53  | 9.22  | 9.39  |
| Training  | 25.96       | 25.54               | 24.76    | 23.84 | 23.71 | 24.76 |
| Classification | 0.78       | 0.76                | 0.77     | 0.76  | 0.78  | 0.77  |
| MFCC      | 35.67       | 36.64               | 35.53    | 35.24 | 34.32 | 35.48 |
| Training  | 11.92       | 11.30               | 11.31    | 11.49 | 10.73 | 11.35 |
| Classification | 0.78       | 0.77                | 0.76     | 0.76  | 0.76  | 0.76  |
| Pitch period | Extraction | 10.78               | 10.25    | 10.16 | 10.24 | 10.31 |
| Training  | 18.62       | 20.08               | 21.29    | 23.59 | 23.59 | 20.76 |
| Classification | 0.81       | 0.80                | 0.82     | 0.77  | 0.75  | 0.79  |

From Table 1, the recognition rate of FFT and MFCC is equivalent, which is obviously better than the pitch period. The recognition rate of feature extraction method is not affected by height change, height changes, and the recognition rate fluctuates less. Mainly because the height change mainly affects the intensity of the time domain signal and the amplitude of the high frequency signal. As the height increases, the signal strength decreases and the high frequency signal attenuates. However, each feature extraction algorithm mainly focuses on the frequency domain features and the frequency is relatively low; FFT extraction, the method of feature is poor in the case of face-to-fly, mainly because the ground sound characteristics are complex, the spectral components are complex, and the fundamental frequency and frequency-doubled relationship are not obvious. The pitch period estimation is weaker at different heights. The other two methods.

From Table 2, in general, the length of time-consuming reflects the complexity of the operation; the feature extraction process FFT takes the shortest time, and the MFCC takes the longest time, because the FFT only needs to perform the transform domain and extract the fundamental frequency and multiplier. Features, but the MFCC also needs to change the Mel domain on the basis of the frequency domain and perform complex logarithm operations, which increases the time consuming; the training process MFCC takes the shortest time, the pitch period estimation is second, and the FFT
takes the longest time. It is inversely proportional to the extracted feature dimension, because the most time-consuming process in the training process is mainly the optimization process of parameters c and g. The increase of the dimension will increase the difficulty of the optimization process and increase the time; the prediction process takes time, the three quite.

5. Conclusion
By combining the aerodynamic characteristics of the UAV, this paper deeply studies the sound characteristics of the UAV, analyzes and compares the three characteristics comparison methods and conducts experiments. Model classification is performed using the SVM classification method. Experiments show that FFT focuses on the full frequency domain characteristics of the signal, MFCC focuses on the characteristics of the local amplification band, and gene cycle estimation focuses on the fundamental frequency characteristics of the acoustic signal. Among them, the complexity of FFT feature extraction is low, and the feature dimension extracted by MFCC is low. The low feature dimension mainly affects the model training process, but the model training as a pre-preparation work has less influence on the real-time UAV detection and recognition efficiency; the feature extraction process occurs in the real-time detection process of the detection system, which affects the system in real time. The important factor of sex; the overall recognition rate of pitch period estimation is low, especially in the face of drones with close aerodynamic characteristics, which cannot be effectively identified.

6. Acknowledgement
Thanks for support of National Natural Science Foundation of China (Youth Science Fund Project) (Grand No.61101238)

7. References
[1] Yang Donghai. Research and implementation of four-rotor UAV detection technology based on voice recognition[D]. Xiangtan University, 2017.
[2] Xia Peng. Research on interference and countermeasure technology for civilian small drones[D]. University of Electronic Science and Technology of China, 2018.
[3] Huang Qiang. Research on passive acoustic detection and location method for low-altitude drones[D]. University of Electronic Science and Technology of China, 2018.
[4] Wang Xuemeng. Research and implementation of passive detection system for unmanned aerial vehicles with weak sound source[D]. University of Electronic Science and Technology of China, 2018.
[5] Yu Yu-yang. Design of a UAV Monitoring Device Based on Microphone Array[J]. China High-tech Zone, 2018(07):9-10.
[6] Wang Wei. An Teng-fei, Ou Jian-ping. Research on passive audio detection and recognition technology of UAV[J]. Journal of Acoustics, 2018(01): 89-93.
[7] Qiu Yu-bin, Li Jian-liang. Audio Feature Extraction Method for UAV Recognition[J]. Noise and Vibration Control, 2018(02): 188-192.
[8] Xiao Han-chun, Guo Jun-feng, Zhang Li. Application of Improved Mel Cepstrum Coefficient in Feature Extraction of Low Altitude Aircraft[J]. Applied Acoustics, 2018(06):909-915.
[9] Zhao Hong-xu, Zhang Yi-wen, Yang Wen-shuai. Research on Vehicle Type Recognition of Mixed Vehicle Audio Signal Based on Support Vector Machine[J]. Journal of Laser Science, 2018(09):133-137.
[10] Wang Bing. Bionic Noise Reduction of Small UAV Rotor [D]. University of Science and Technology of China, 2018.
[11] DAVID A, ROBERTO G, MANUEL U, et al. An evolutionary algorithm to optimize the microphone array configuration for speech acquisition in vehicles[J]. Engineering Applications of Artificial Intelligence, 2014, 34:37-44.
[12] M. Artuso, C. Betancourt, I. Bezshyiko et al.. Signal coupling to embedded pitch adapters in silicon sensors[J]. Nuclear Inst. and Methods in Physics Research, A, 2018, 877.
[13] Rajeev Rajan, Hema A. Murthy. Two-pitch tracking in co-channel speech using modified group delay functions[J]. Speech Communication, 2017, 88-92.

[14] Anett Antony, R. Gopikakumari. Speaker identification based on combination of MFCC and UMRT based features[J]. Procedia Computer Science, 2018, 250-257.