REAL-TIME CLASSIFICATION OF FLYING QUADCOPTER AUDIO SIGNALS BY MEL SPECTROGRAMS

Andrey Kadomskij

St. Petersburg State University
st099001@student.spbu.ru

Abstract

In the article we suggest a method for detection of quadcopters by sound by means of Mel Spectrograms. For the experiment, audio recordings of both flying quadcopters signals and noises were collected. The experiments were performed by means of spectrograms with 2 classes, Mel Spectrograms with 2 classes and Mel Spectrograms with 4 classes. The method has proved to be effective at testing and can be used for quadcopter real-time detection software.

1. INTRODUCTION

To classify quadcopters various techniques are applied, such as

- Visual analysis. Analysis of camera image for the search of a quadcopter. This method is significantly dependable on the camera scope. If the quadcopter is at certain distance, it will be difficult to differentiate it from other small objects (for example birds). The efficiency is also reduced in heavy weather conditions (fog, rain, snow). [1]
- Radar. Due to the narrow scope radars cannot insure efficient quadcopter detection. [2]
- Sound analysis. The interest to the sound detection of quadcopters has appeared recently, and only few researches concern this theme. [3]

In this work the sound analysis is used for a classification.

The paper [3] analyses the frequencies of quadcopter monitor and rotors. This method does not allow detecting other quadcopters, because the frequencies of rotors and motors may differ. But the above mentioned method allows detecting quadcopters which produce sounds 3 decibels lower than the surrounding noises.

It the competition dedicated to the classification of sounds of birds [4] the Mel Spectrograms showed the best results. [5]

In this work for classification we applied the method used by the winners of the competition -- Mel Spectrograms and Resnet-50 neural network. It allowed detecting different types of quadcopters in the selected set of data.

2. MAIN NOTIONS

Mel Spectrogram is a spectrogram where frequency is expressed not in hertz, but in mel's.

There exist different ways of transform hertz to mels, and the most common is the following:

$$m = 1127 \ln(1 + \frac{f}{700}) \tag{1}$$

Mel Spectrogram helps to separate out the bass frequencies better, while they propagate in nature further than high frequencies.

The difference between a spectrogram and a Mel Spectrogram is represented in Figure 1.

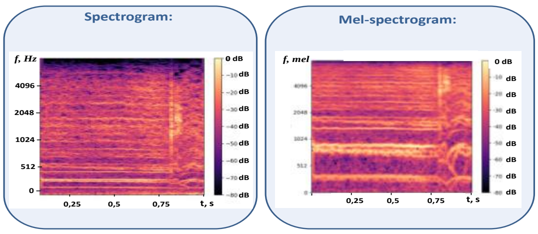

Figure 1. The difference between spectrogram and Mel Spectrogram

2.1. Description of Test dataset

In the test set of data the sounds of flying DJI P3 Pro, FPV250, DJI – Agras T30 quadcopters were used. Many different sounds occurred in the noise record, such as technical, city noise, speech, wind etc.

Each sound was saved in wav format; the sound-track from the video from Youtube.com was also cut out and saved in wav format.

2.2. Procedure of a nalysis

A Mel Spectrogram of flying quadcopter is shown in Figure 2.

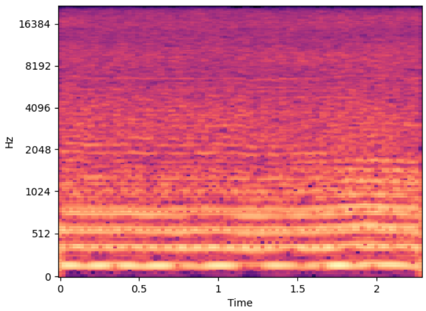

Figure 2. Mel-Spectrogram of flying quadcopter

The audio signals were read at 20000 Hz, because the equipment used to test the program for detecting quadcopter in real-time reads sounds in diapason between 20 and 20000 Hz.

Length of the windowed signal for fourier transformation for Mel-spectrogram – 1024, the amount of mels – 220, minimal frequency - 20 Hz, maximal – 20000 Hz.

The majority of quadcopters showed the activity between 0 and 1000 mels frequencies, but as we don`t know, how other quadcopters will react, all frequencies, which microphone can read were displayed.

2.3. Neural Network Training

All signals (quadcopters and noise) were cut into parts by the length one second, then they were divided into training, validation and testing sets.

As a result, 1440 second-parts of flying quadcopter soundtracks, 511 validation and 1221 testing were used for training. For noise signal we used 3603 for training, 451 for validation and 1144 for testing.

For sound augmentation the following procedures were performed:

- Sound deceleration/precipitation for 5-10%
- Pitch shift for 3-5%
- Poor signal records multiplication by means of quadratic transformation
- Quadcopter sound amplification in relation to noise and mush by power transformation with ½ exponent
- Signal mirroring by time

The Resnet-50 neural network to detect a flying quadcopter was trained on the following data:

- Spectrograms with 2 classes: quadcopters and noises
- Spectrograms with 4 classes: quadcopter, city noises, natural noises, speech
- Mel Spectrograms with 4 classes

After that an extra training on the larger amount of data, received by link [6] was performed.

This data set is of higher quality (all these sounds were recorded with professional equipment). Each sound was classified according to quadcopter model.

For training set 8400 second-parts of quadcopter recordings were used, for validation 1994, for testing 5044. For noises 7369 in training, 1840 in validation, 27358 in testing.

This training was performed on 4 classes with Mel Spectrograms only. For the augmentation of training data set the same actions were performed.

2.4. Results of experiments

As Table 1 shows, the method of increasing of the amount of classes proved to be most effective, because we are not interested in accuracy of classes "city noises", "natural noises" and "speech" detection, if during the classification neural network doesn`t relate these data to the quadcopter. The usage of Mel Spectrograms also shows significant improvement.

Table 1. The results of quadcopter detection according to classes and representations

Amount of classes	2	4	4
Training set	Sp-s	Sp-s	Mel-sp-s
Validation: TP, % FP, %	 73.6 25.3	 97.2 3.3	 99.8 0.2
Testing: TP, % FP, %	 69.8 29.5	 90.8 9.2	 99.5 0.8

The sound detection efficiency of quadcopter was higher with the larger amount of data. 1 quadcopter sound wasn`t detected on validation and testing sets (~99.95% detections on validation and

~99.98% in testing) as well as 1 sound was false detected as quadcopter on validation and 2 on testing set (~0.05% and ~0.007% accordingly).

Resnet-50 neural network was trained by means of Adam optimizer with 10^{-5} learning rate.

2.5. Detection errors

The quadcopter was not detected in situations when the sound of quadcopter was in extremely noisy environment.

On the larger amount of data the amount of false positive detections decreased. Only 1 extremely noisy sound was false positively detected as quadcopter.

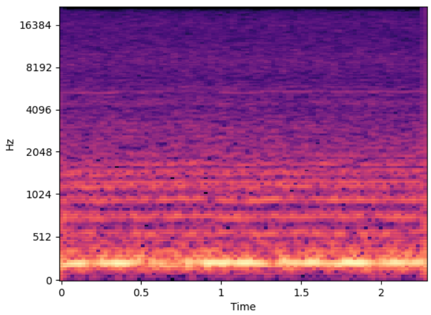

Figure 3. Quadcopter Mel-Spectrograms in noisy environment

The quadcopter was false positively detected on some natural noises (contained in class "city noises") Mel Spectrograms.

On the larger amount of data also only 2 natural noises sounds were false positively classified as quadcopter. The traffic sounds from city noises were no longer false classified as quadcopter.

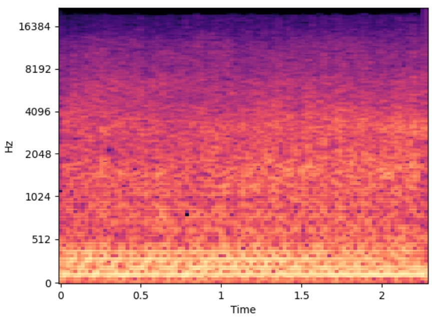

Figure 4. Natural noise Mel Spectrogram false positively detected as quadcopter

2.6. Real-time quadcopter detection software system

For real-time quadcopter detection the program was set up. It reads the sound from the microphone, each second the signal is transformed into Mel Spectrogram and is applied to the neural network which defines the detection class. The operator receives a signal, if the quadcopter is detected or not.

The above mentioned software system is planned to be tested in real conditions.

3. CONCLUSIONS

The way of quadcopter detection implied in this work can be effectively implemented.

The use of larger amount of data increases the detection efficiency

Mel Spectrograms used for training perform the best results during the testing. The adding of different noise classes results in the significant efficiency gains as well. Moreover, the confusion of the noises does not effect on the result of detection. In further research, we are going to test the software system for real-time quadcopter detection in the field, use other signal notations for neural network training, compile the new signal samplings for neural network training and testing.

4. ACKNOWLEDGMENT

The author would like to thank Prof. N. Ampilova, St. Petersburg State University, Dept. of Mathematics and Informatics for theoretical help in digital processing area and editing this paper.

References

[1] Samadzadegan, F.; Dadrass Javan, F.; Ashtari Mahini, F.; Gholamshahi, M. Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery. Aerospace 2022, 9, 31. https://doi.org/ 10.3390/aerospace9010031

[2] Baek S, Jung Y, Lee S (2021) Signal expansion method in indoor FMCW radar Systems for Improving Range Resolution[J]. Sensors 21(12):4226

[3] Djurek, Ivan & Petosić, Antonio & Grubeša, Sanja & Suhanek, Mia. (2020). Analysis of a Quadcopter's Acoustic Signature in Different Flight Regimes. IEEE Access. 8. 1-1. 10.1109/ACCESS.2020.2965177.

[4] https://www.kaggle.com/c/birdsong-recognition

[5] S. S. Stevens and J. Volkmann, E. B. Newman, "A Scale for the Measurement of the Psychological Magnitude Pitch", The Journal of the Acoustical Society of America 8, 185-190 (1937) https://doi.org/10.1121/1.1915893

[6] https://hissandroar.com/v3/soundlibrary/sd024-quadcopter-uav-drone