Drone Detection and Classification Using an Acoustic Camera

Sanja Grubeša[1], Jasna Stamać[2], Nikša Orlić[2], Tomislav Grubeša[2]
[1] University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Electroacoustics,
Zagreb, CROATIA, sanja.grubesa@fer.hr
[2] Geolux d.o.o., Ljudevita Gaja 62, Samobor, CROATIA
geolux@geolux.hr

ABSTRACT

In our research, as part of the 4D Acoustic Camera project, an acoustic detector, i.e. a prototype of an acoustic camera, was developed. In order to achieve our goal which is to design a robust yet small acoustic camera, which can be used in different security systems, 72 MEMS microphones were used to form a microphone array in the shape of a hemisphere.The acoustic camera prototype records the sound in a protected area and classifies it. The classification is carried out by comparing the recorded sound from the protected area with the sounds existing in a database, and in this way the acoustic camera either eliminates or confirms the sound source as the actual target. Thus, an acoustic camera would eliminate the expected sounds typical of human movement, such as walking through low vegetation, as a potential target because it eliminates the normal sounds present in the environment. The first step of classification was to create a database of different sounds. We divided the samples into the following four categories: noise, walking, speech, and drone flight. For each of the sounds in the database it is necessary to obtain a spectrogram. When recording sounds, i.e. obtaining spectrograms with our acoustic camera prototype, each sound is recorded multi-channel (72-channel). The acoustic camera prototype uses the "Delay and Sum" (DAS) algorithm which enables it to maximize the array's sensitivity to incoming sound waves coming from a particular direction. In this way, attenuation of sound from directions which are not of interest is achieved, making it is possible to reduce the influence of noise, and thus it is easier to extract a useful signal.Such spectrograms from the database represent a training set for a convolutional neural network that we will use as a classification tool, with which the sounds from a protected area will be classified. Since convolutional neural networks are often used to classify images, we can imagine the input to the convolutional network as a grayscale image which represents the spectrogram. Preliminary classification results show a high classification of the samples, which shows that the method and architecture of the selected neural network are appropriate.

Index Terms *- drone, acoustic camera, MEMS microphones, microphones array, classification, convolutional neural networks*

1.0 INTRODUCTION

The primary idea as part of the 4D Acoustic Camera project is to develop an acoustic detector based on the correlation of sound obtained from a large number of microphones (i.e. Acoustic Camera). This type of correlation principle could precisely locate sound in space, and thus create a map of amplitude and frequency distribution in space and time. Combining a large number of microphones will increase the detection range and the precision of localization and classification i.e. it will significantly increase the probability of threat detection while reducing the number of false alarms.

Acoustic cameras present a relatively new approach and only a few finished products are commercially available. All available products are designed for application in noise measurements and acoustic characterization of different noise sources with small spatial spacing [1-4]. Therefore, it can be concluded that these acoustic cameras are primarily developed for laboratory or controlled conditions i.e. long outdoor exposure and use without proper protection could damage the microphones and reduce, or completely disable, their performance.

Keeping in mind our primary goal, which is to design a broadband acoustic camera using micro-electromechanical system (MEMS) microphones [5-7], an optimization of the microphone array has been performed in a way that the gain in the desired direction and the attenuation of side lobes is maximized at a frequency up to 4 kHz. In our previous research [8, 9] several simulations were performed considering square, circular and hemisphere shaped MEMS arrays with varying number of microphones and varying spacing between the microphones.

In order to achieve our goal which is to design a robust yet small acoustic camera, which can be used in different security systems, 72 MEMS microphones were used to form a microphone array in the shape of a hemisphere, and this represents the prototype of the Acoustic camera designed and developed as part of the 4D Acoustic Camera project, Figure 1-1.

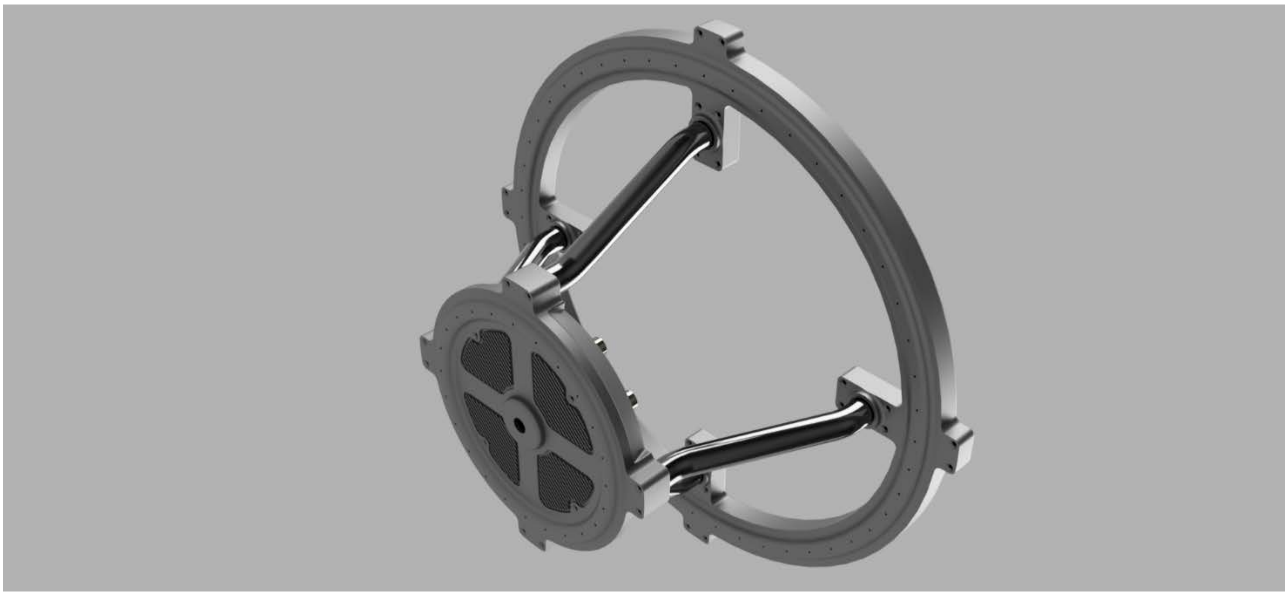

Figure 1-1: The Designed Prototype of a Hemisphere Shaped Acoustic Camera.

2.0 CLASSIFICATION OF SOUNDS RECORDED BY AN ACOUSTIC CAMERA

The acoustic camera prototype records the sound in a protected area and classifies it. The classification is carried out by comparing the recorded sound from the protected area with the sounds existing in a database, and in this way the acoustic camera either eliminates or confirms the sound source as the actual target. Thus, an acoustic camera would eliminate the expected sounds typical of human movement, such as walking through low vegetation, as a potential target because it eliminates the normal sounds present in the environment.

The first step of classification was to create a database of different sounds. We divided the samples into the following four categories: noise, walking, speech, and drone flight. For each of the sounds in the database it is necessary to obtain a spectrogram. An example of a spectrogram can be seen in Figure 2-1.

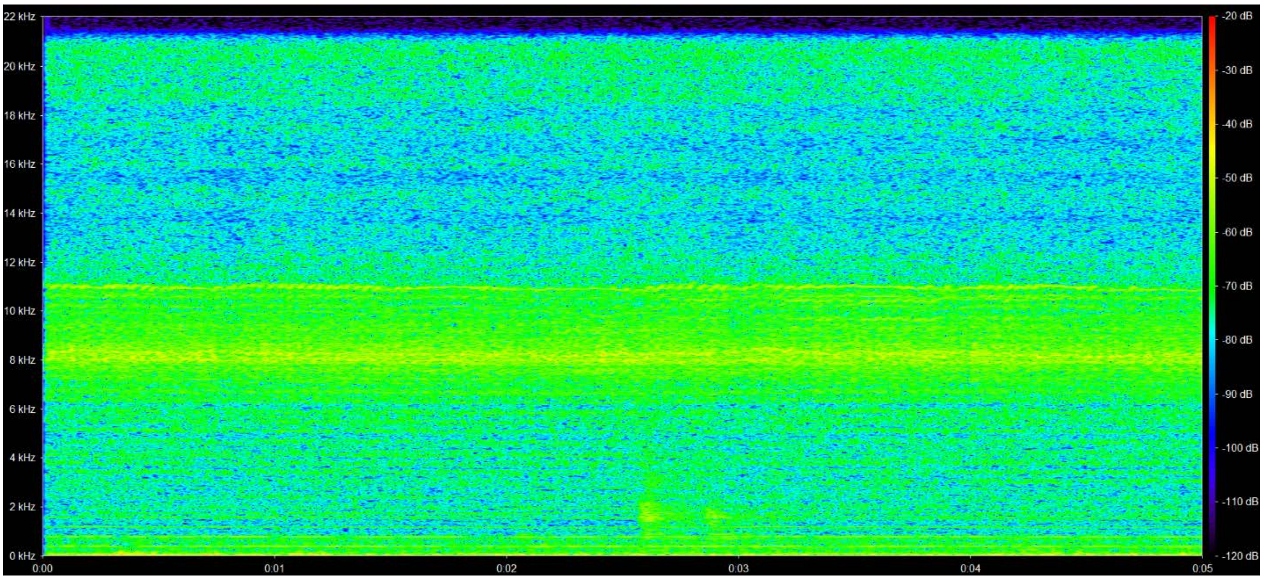

Figure 2-1: Example of a Spectrogram.

When recording sounds, i.e. obtaining spectrograms with our acoustic camera prototype, each sound is recorded multi-channel (72-channel). Spectrograms shown in Figure 2-2 and Figure 2-3 were obtained by processing the data recorded with our acoustic camera prototype. The acoustic camera prototype uses the "Delay and Sum" (DAS) algorithm which enables it to maximize the array's sensitivity to incoming sound waves coming from a particular direction. In this way, attenuation of sound from directions which are not of interest is achieved, making it is possible to reduce the influence of noise, and thus it is easier to extract a useful signal. The algorithm is easy to apply because it delays the signal recorded by each microphone with respect to the incoming direction. The signals are then added together, which results in the amplification of signals coming from desired direction, while signals from unwanted directions are attenuated. The spectrogram is obtained using the resulting audio signal. Figure 2-2 shows an example of a sound spectrogram generated by a drone 5 m away from the acoustic camera prototype, and Figure 2-3 shows an example of a sound spectrogram generated by a drone 50 m away from the acoustic camera prototype.

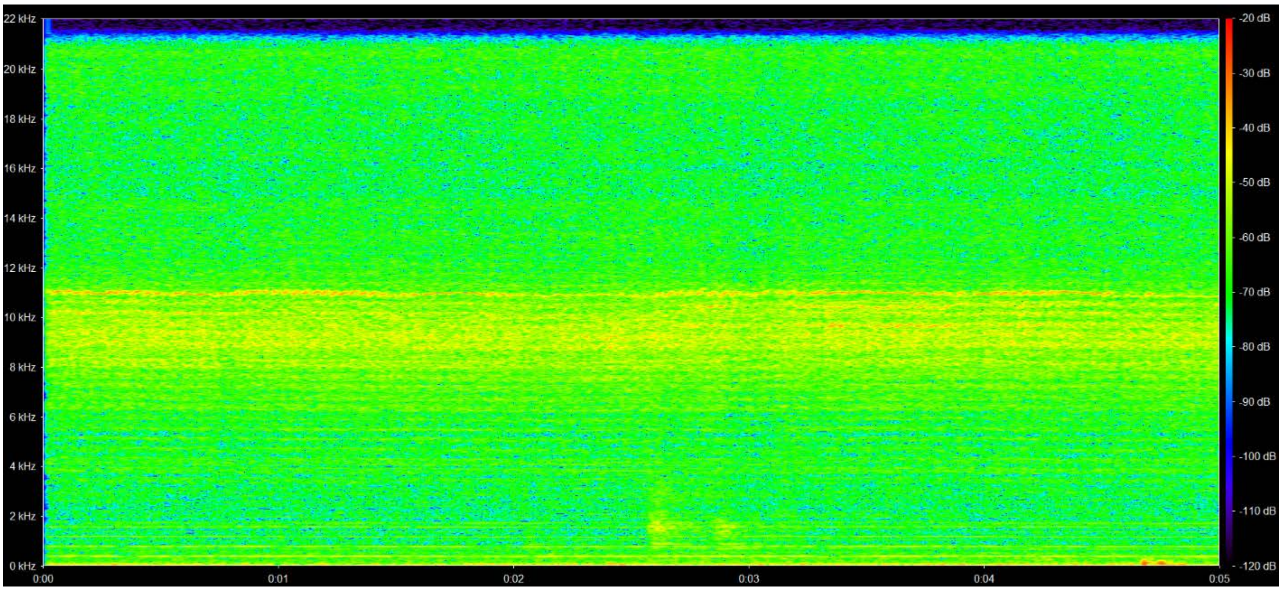

Figure 2-1. Example of a Sound Spectrogram Generated by a Drone 5 m Away From the Acoustic Camera Prototype.

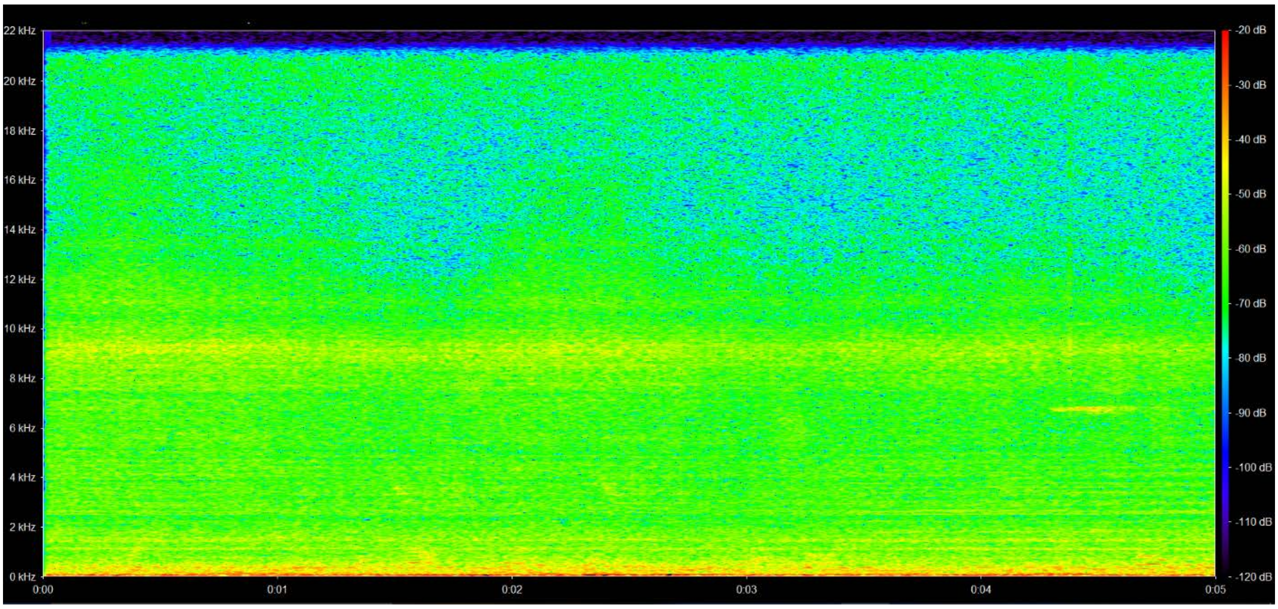

Figure 2-3. Example of a Sound Spectrogram Generated by a Drone 50 m Away From The Acoustic Camera Prototype.

Such spectrograms from the database represent a training set for a convolutional neural network that we will use as a classification tool, with which the sounds from a protected area will be classified. Since convolutional neural networks are often used to classify images, we can imagine the input to the convolutional network as a grayscale image which represents the spectrogram.

3.0 CONVOLUTIONAL NEURAL NETWORK

A Convolutional Neural Network is an algorithm which can take in an input image, assign importance to various objects in said image and is able to differentiate one object from another. The name “convolutional neural network” is derived from the fact that the network uses a mathematical operation called convolution, [10].

The traditional Convolutional Neural Network structure is shown in Figure 3-1. In general, the basic structure of Convolutional Neural Network consists of two layers. The first layer is the feature extraction layer which is usually composed of a convolutional layer and a pooling layer. The convolutional layer consists of several convolution units, and the parameters of each convolution unit are optimized by the backpropagation algorithm. The pooling layer uses down-sampling to reduce the size of the feature map output of the convolution layer. The second layer is the feature mapping layer which usually refers to the fully connected layer. In a fully connected layer each neuron is connected to all the neurons of the previous layer to integrate the features from multiple feature maps extracted by the convolution layer, [11].

The architecture of convolutional neural networks has proven to be extremely effective in working with images and recognizing features from them. Therefore, a convolutional neural network imposes itself as a very good solution for classifying sounds from a protected area, provided, of course, that these sounds are presented as a grayscale spectrogram.

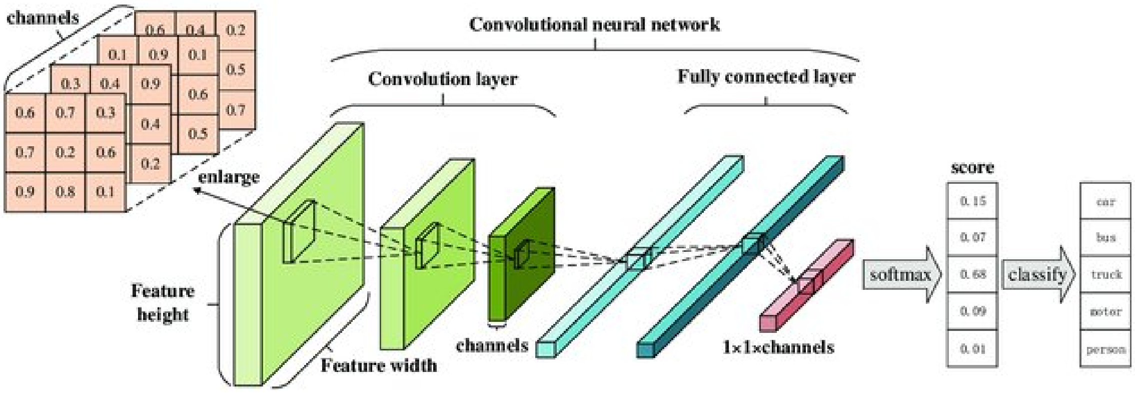

Figure 3-1. Architecture of a Convolutional Neural Network [11].

3.1. Implementation of a Convolutional Neural Network for Acoustic Camera

Today, there are several software tools that allow easy and fast work with convolutional neural networks. The two best known are PyTorch and TensorFlow, and both are based on the Python programming language. In the acoustic camera development project, we chose the TensorFlow tool. Using these tools, it is very easy to experiment with different architectures of convolutional neural networks, and with different sets of recorded samples (training sets) to select the optimal network architecture and check if the training set contains enough recordings. Neural network training with samples from the training set is performed on a personal computer, and the weighting factors obtained in the training phase are then used to classify the samples in real time.

In the classification of acoustic samples, the input to the first layer of the neural network is a 2D spectrogram matrix, where each element of the matrix contains the intensity of the spectrogram at a given point. Since convolutional neural networks are frequently used to classify images, we can imagine the input to a convolutional network as a grayscale image depicting a spectrogram.

After a few experiments, we chose a neural network architecture that looks as follows:

- 1. Input layer
- 2. Max-pooling layer
- 3. 2D convolution 3x3, 64 elements
- 4. Max-pooling layer
- 5. Dropout 50%
- 6. 2D convolution 2x2, 64
- 7. Max-pooling layer
- 8. Dropout 50%
- 9. 2D convolution 2x2, 64
- 10. Flatten layer, 128 elements
- 11. Dropout 50%

We divided the samples into the following four categories:

- noise
- walking
- speech
- and drone flight.

4.0 RESULTS

From the sample set, 10% of the samples from each category were randomly selected for validation, and the remaining 90% of the samples were used to train the neural network. The final results on the validation set are shown in Table 4-1. As expected, the worst results were obtained for the speech. Best results were obtained for drone flight, and this shows that drones make recognizable noise prints or noise spectrograms.

Table 4-1. Results of Validation of Audio Signals.

		Real samples for validation			
		Noise	Walking	Speech	Drone flight
Classification results	Noise	87%	7%	7%	4%
	Walking	2%	89%	5%	1%
	Speech	3%	3%	85%	3%
	Drone flight	8%	1%	3%	92%

5.0 CONCLUSION

The architecture of convolutional neural networks has proven to be extremely effective in working with images and recognizing features from them. In the classification of acoustic samples, the input to the neural network is a 2D spectrogram matrix, where each element of the matrix contains the intensity of the spectrogram at a given point. Since convolutional neural networks are frequently used to classify images, we can imagine the input to the convolutional network as a grayscale image which represents the spectrogram. Preliminary classification results show a high classification accuracy of the samples, i.e. spectrograms, which shows that the method and architecture of the selected neural network are appropriate. Although the number of collected samples used as the training set for the convolutional neural network was relatively small, the achieved classification accuracy proved to be high, hence it will not be necessary to modify the neural network architecture. In order to achieve even higher classification accuracy, a much larger set of neural network test samples will need to be made in the future.

REFERENCE

[1] Norsonic acoustic camera, https://web2.norsonic.com/product_single/acoustic-camera/

[2] GFAI Tech Gmbh – Acoustic Camera, https://www.acoustic-camera.com/en/products.html

[3] B&K Acoustic Camera Type 9712-W-FEN, https://www.bksv.com/-/media/literature/Product-Data/bp2534.ashx

[4] K. Tontiwattnakul, J. Hongweing, P. Trakulsatjawat and P. Noimai, "Design and build of a planar acoustic camera using digital microphones," 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), 2019, pp. 1-4, doi: 10.1109/ICEAST.2019.8802537.

[5] R. Bauer, Y. Zhang, J. C. Jackson, W. M. Whitmer, W. O. Brimijoin, M. A. Akeroyd, D. Uttamchandani, and J. F. C. Windmill, “Influence of Microphone Housing on the Directional Response of Piezoelectric MEMS Microphones Inspired by Ormia Ochracea,” IEEE Sensors Journal, vol. 17, pp. 5529- 5536 , September 2017.

[6] Application note AN4426, “Tutorial for MEMS microphones”.

[7] S. Walser, C. Siegel, M. Winter, G. Feiertag, M. Loibl and A.Leidl, “MEMS microphones with narrow sensitivity distribution”, Sensors and Actuators A: Physical, 247, 663-670, 2016.

[8] J. Stamac, S. Grubesa and A. Petosic, “Designing the Acoustic Camera using MATLAB with respect to different types of microphone arrays”, Second International Colloquium on Smart Grid Metrology, SMAGRIMET 2019.

[9] S. Grubesa, J. Stamac and M. Suhanek, “Acoustic Camera Design with Different Types of MEMS Microphone Arrays”, American Journal of Environmental Science and Engineering 3(4):88-93, 2019, DOI: 10.11648/j.ajese.20190304.14

[10] I. Goodfellow, Y. Bengio and A. Courville, “Deep Learning”, MIT Press, 2016., http://www.deeplearningbook.org

[11] X. Kang, B. Song and F. Sun, “A Deep Similarity Metric Method Based on Incomplete Data for Traffic Anomaly Detection in IoT”, January 2019 Applied Sciences 9(1):135, DOI:10.3390/app901013