Automatic Classification of Multiple Acoustic Events Using Artificial Neural Networks

Dineth Egodage

188015M

Thesis submitted in partial fulfillment of the requirements for the
Degree of Master of Science (Research) in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa
Sri Lanka

January 2021

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature of the candidate: Date:

The above candidate has carried out research for the MSc thesis under my supervision.

Name of the supervisor:

Signature of the supervisor: Date :

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude towards the Department of Computer Science and Engineering offering the Master's Degree programme to perform this kind of work.

Secondly, I would like to express my heartiest gratitude towards my research supervisor Dr. Sulochana Sooriyaarachchi for the guidance and immense support provided in both knowledge and resource-wise to make this research a success.

I would also like to thank Dr. Navinda Kottege and Dr. Chandana Gamage for guiding and advising us to stay on the correct path at the initial stages of this research. The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, should also be mentioned thankfully for providing this research idea and Dr. Kottege as a resource person for us.

Furthermore, I would like to convey my gratitude towards Dr. Sampath Seneviratne from Animal Communications, Genetics, Evolutionary Biology section of the University of Colombo for sharing his expertise domain knowledge and spend his valuable time to guide me through the research.

Besides, I am grateful to the progress panel members, Dr. Ranga Rodgrigo and Dr. Charith Chithrarajan, for giving their valuable feedback on the work carried out.

Finally, I would like to express my gratitude towards Dr. Charith Chithrarajan as the Research Coordinator and all other staff members for the knowledge given throughout the last two years.

Finally, I would like to express my heartiest gratitude towards my friends and family for the immense support given throughout the research. Special thanks to my wife, who remained supportive until the end.

ABSTRACT

There are numerous scenarios where similar acoustic events occur multiple times. Acoustic monitoring of migratory birds is an ideal example. Birds make a type of call known as flight calls during migration. A flight call can be considered as *an acoustic event* because it is a short-term, intuitively distinct sound. It is challenging to identify multiple occurrences of extremely short-range acoustic events such as flight calls in real-world recordings using classification techniques that require more computational power. It is mainly due to background noise and complex acoustic environments. This research aims at developing a classification model that reduces the effect of background noise, extract ROIs from continuous recordings, extract suitable features of flight calls and detect multiple occurrences of flight calls. An improved algorithm that can extract features has been developed in this research—by combining a well known Maximally Stable Extremal Regions (MSER) technique with state of the art traditional techniques. Namely Spectral and Temporal Features(SATF) and a combination of SATF and Spectrogram-based Image Frequency Statistics(SIFS). We name this novel algorithm as Spectrogram-based Maximally Stable Extremal Regions (SMSER). Three distinct feature sets have formed such that Featureset-1 created using SATF. Featureset-2 is a blend of SATF and SIFS. Featureset-3 is a combination of SATF, SIFS, and SMSER. The kNN, RF, SVM, and DNN classification techniques evaluated a real-world dataset using the extracted feature sets. Research carried out several tests to find out the best performing combination of classification model and feature set. The results showed that the flight calls' detection accuracy increased when the number of features increased, although high computational power requirement is a disadvantage. The performance of SMSER feature set was the best among almost every classification technique above. It should be because the SMSER Feature set has the highest number of features. Classification of the SMSER feature set from the DNN classifier showed the highest accuracy of 87.67%.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
DNN	Deep Neural Network
DFT	Discrete Fourier Transformation
DWTC	Discrete Wavelet Transform Coefficients
FFT	Fast Fourier Transformation
GMM	Gaussian Mixture Models
GPU	Graphics Processing Unit
HMM	Hidden Markov Models
k-NN	k-Nearest Neighbor
MFCC	Mel Frequency Cepstral Coefficients
PC	Personal Computer
PCA	Principal Component Analysis
RF	Random Forest Algorithm
RFE	Recursive Feature Elimination
RMSE	Root Mean Square Energy
SATF	Spectral and Temporal Features
SIFS	Spectrogram-based Image Frequency Statistics
SMSER	Spectrogram-based Maximally Stable Extremal Regions
SVM	Support Vector Machine
ZCR	Zero-Crossing Rate

TABLE OF CONTENTS