Classification of Acoustic Scenes using Background and Foreground

A THESIS

submitted by

Dhanunjaya Varma D
(Roll No : S18023)

for the award of the degree of

Master of Science
(by Research)

School of Computing and Electrical Engineering
Indian Institute of Technology Mandi
India - 175005

November 2021

"Creativity is seeing the same thing but thinking differently."

– A. P. J. Abdul Kalam

Dr. A. P. J. Abdul Kalam served as the 11th President of India from 2002 to 2007. The Government of India honoured him with the Padma Bhushan in 1981, the Padma Vibhushan in 1990 and the Bharat Ratna in 1997.

I dedicate this thesis:

To my Father,

To my Mother,

To my Teachers,

To my Siblings,

To my Friends.

Declaration

I hereby declare that this work incorporated in this thesis is the outcome of the studies accomplished by me in the **School of Computing and Electrical Engineering, Indian Institute of Technology Mandi**, under the supervision of **Dr. Padmanabhan Rajan**. This work has not been submitted elsewhere for any degree or diploma. In keeping with the general practice, due acknowledgments have been made wherever the work described is based on the finding of other investigators. In addition, I certify that no part of this work will, in future, be used for submission in my name, for the award of any other degree at any university.

Place: Mandi **Dhanunjaya Varma D**

Date:

Thesis Certificate

This is to certify that the thesis titled **"Classification of Acoustic Scenes using Background and Foreground"**, submitted by **Dhanunjaya Varma D**, at Indian Institute of Technology Mandi for the award of Master of Science (by research) is a bonafide record of the research work carried out by him under my supervision. The content of this thesis, in full or in parts, have not been submitted to any other institute or university for the award of any degree or diploma.

Dr. Padmanabhan Rajan
(Thesis Supervisor)

Acknowledgement

First and foremost, I would like to express my heartfelt gratitude toward my thesis supervisor Dr. Padmanabhan Rajan. It is because of his firm belief in me, continuous support and patience that I have been able to complete this journey of M.S. He was always available to not only share his immense knowledge and expertise but also consistently motivated and guided me in the right direction. I could not have asked for a better advisor, and I will miss just walking in to his office to talk about research and many other things.

Apart from my supervisor, I would also like to express my sincere gratitude to my APC committee members: Dr. Aditya Nigam, Dr. Rameshwar Pratap, Dr. Manoj Thakur and Dr. Samar Agnihotri for their encouragement, valuable feedbacks related to my research work, co-operation and motivation.

I would also like to thank Dr. Dileep A.D. for his kind support and for all the meaningful discussions and suggestions during the Audio & Speech group meetings.

I express special thanks to staff members of school, IIT Mandi for their readiness to always help in official works. I am greatly thankful to IIT Mandi as an institution for providing well equipped research labs and infrastructure.

I would like to extend my sincere thanks to Dr. Gaurav Bhutani and HPC team for providing such a great service and computing power.

I would further like to thank my fellow group members Arshdeep Boparai, Muralikrishna H and Akansha Tyagi for helping in my work as well as in other things.

I would like to further extend my thanks to my friends Muneeswaran P, Dinesh Kumar B and Dalchand Ahirwar for keeping me motivated all the time and making this journey beautiful and smoother.

On the top of everything, I owe my life and the debt of gratitude to my *Father*,

Mother, *Brother*, and *Sister* who are my backbone. This journey would not have been possible without them and their unceasing support.

Dhanunjaya Varma D

ABSTRACT

The objective of acoustic scene classification is to classify environments based on the sound events they produce. Acoustic scene classification has been used in a variety of applications, which include audio surveillance, assistive technologies like hearing aids and context-aware services. ASC is a challenging task due to the presence of similar sound events across acoustic scenes, causing high inter-class similarity. In this thesis, we approach this problem by providing a mechanism that helps in deriving discriminative features by suppressing certain sound events.

An acoustic scene can be viewed as a combination of background sound events and foreground sound events. Often, either the background or the foreground carries beneficial information in identifying the acoustic scenes uniquely. We propose to handle these similar sound events by utilizing a combination of methods that include robust principal component analysis (RPCA), subspace projection techniques and a self-attention network. These methods help in separating the background and the foreground sound events, and in partially removing the background (or foreground) sound events.

We employ the framework of RPCA to decompose the given acoustic scene into the background and the foreground sound events. RPCA decomposes a given data matrix into a low-rank and a sparse matrix. In the context of data describing an acoustic scene, the low-rank matrix represents the slow-changing background, and the sparse matrix represents the occasional foreground sound events. Further, we utilize a subspace projection technique named nuisance attribute projection (NAP) to reduce the inter-class similarity. NAP helps in partially removing the background (or the foreground) sound events by treating either the background (or the foreground) as nuisance variations. The nuisance basis for applying NAP are learned from the background and foreground separated data obtained post RPCA. These background-suppressed and the foreground-suppressed representations are combined using fusion techniques to improve classification accuracy. We also present an approach to incorporate the label information in the subspace projections by learning class-specific nuisance bases. Further, projection using these bases in combination with an attention mechanism is used for effective suppression, leading to better

discrimination. Our results on standard datasets indicate that the proposed methods that use RPCA and subspace projections are indeed helpful in improving the classification accuracy.

Contents