REVIEW ON OVERLAPPING ACOUSTIC EVENT CLASSIFICATION

Chaithanya S P[1],Haripriya R[2], Likhitha P[3],Preethika J[4], Prof. Shankar B B[5]

[1]*Chaithanya S P Student, ECE Dept, AIET Mijar, Karnataka, India*
[2]*Haripriya R Student, ECE Dept,AIET Mijar, Karnataka, India*
[3]*Likhitha P Student, ECE Dept, AIET Mijar, Karnataka, India*
[4]*Preethika J Student, ECE Dept, AIET Mijar, Karnataka, India*
[5]*Shankar B B professor, ECE Dept, AIET Mijar, Karnataka, India*

ABSTRACT

It is important to know not only speech and music, which has been researched but also common sound in day to day environment each time sound signal contain a combination of information as a mixture of noise, clean sound and noise like characteristics with flat spectrum have extract audio event from audio signals .In this review have propose an approach Mel-Frequency Cepstral Coefficients (MFCC) feature extraction technique, for the classification process To classifies overlapping sound events, we use a support vector machine (SVM) to feature extraction using the statistics that mainly contains Mel spectra where the most relevant feature frame based classification using SVM is a algorithm that analyses the data for classification and recognition it is a important machine learning technique.

Keyword— *Mel-frequency Cepstral Coefficients (MFCCs), support vector machine (SVM), Acoustic event classification (AEC), Acoustic event detection (AED),Non negative matrix factorization (NMF), Hidden markov Module (HMM)*

---

1. INTRODUCTION

In background each surrounding environment has its own set of sounds. For instance sounds generated from vehicles, horn, refrigerator running, ringing phone belong to the background. The sounds present in a particular environment are called acoustic events.

AEC aim to categorizes the audio elements inside an audio clip classifies the overlapped acoustic events in audio can be used in various uses, including indoor environment recognition, surveillance systems and automatic audio indexing Overlapping AEC is a much more puzzling difficult due to the combination of acoustic sources and is measured to be more significant because acoustic events frequently overlap with each other in actual lifetime records spectral features used for speech gratitude may not be suitable for AEC. Hence finding of new features specific to AEC is a interesting task in that environment the human activity is reflected in a rich variety of acoustic events either produced by the human body or by objects handled by humans. Consequently detection or classification of acoustic events may help to detect and describe the human and social activity that takes place in the room. For example clapping or laughter inside a speech discourse, a strong yawn in the middle of a lecture, a chair moving or door noise when the meeting has just started, Acoustic event is a segment of environmental audio that easily occur in human life, such as coughing, phone ringing, clash sound and so on. AEC and AED aim to recognize the audio elements inside an audio clip. Recognizing acoustic events in audio can be utilized in various applications including indoor environment recognition, surveillance systems and automatic audio indexing recently, as the interest in this area increases, huge datasets were released and challenges such as the challenge have been held. Research on AED can be separated into two main scenarios, overlapping and non-overlapping. Overlapping AED is a much more challenging problem due to the mixture of acoustic sources and is considered to be more important because acoustic events often overlap with each other in real life recordings. Automatic scene analysis includes several tasks that target at the acoustic sources segregation, localization, identification previously reported works have considered the problem of segmenting audio streams using a small number of categories or detecting a given acoustic event Several other published papers aim at classifying acoustic events, each one focusing on a given environment or a type of sounds e.g. telemedicine, sports, animals, etc. The topic of sound event recognition covers the detection and classification of sound events in unstructured environments, which may contain multiple overlapping sound sources and non-stationary background noise. Many sounds contribute to the understanding and context of the surrounding environment, and therefore should not be regarded simply as noise, as is common in automatic speech recognition (ASR), automatic noise source recognition, and intelligent audio-based personal archives and so on. Temporal, spectral features used for speech/speaker recognition may not be suitable for AEC. Hence, identification of new features, specific to AEC is a challenging task. Some of the reported works in literature show that combined Time-Frequency (TF) features from Time-Frequency Representations (TFR's) perform better for AEC than Mel-frequency Cepstral coefficients (MFCCs), TF features are obtained using Matching Pursuit(MP) algorithm, Nonnegative Matrix Factorization (NMF) is used to extract TF from MP features TFD, Time-Frequency characteristics of events are also extracted from spectrogram. Local spectrogram features are evaluated from keypoints of the event in the spectrogram with high computational cost, Acoustic scene classification aims to characterize the acoustic environment of an audio stream by selecting a semantic label for it. It can be considered as a machine-learning task within the widespread single-label classification paradigm, in which a set of class labels is provided and the system must select exactly one for any given input.

2. LITERATURE REVIEW

Acoustic event classification and Detection and various sound events is not a simple task, it is one of the challenging task the machine will lesion different kind of audio signal and recognition each event.

According to the author Manjunath Mulimani [1] in his paper a graph signal is generated from spectrogram and features are investigated from graph signal for Acoustic Event Classification (AEC).Three different noises are selected from NOISEX'92 database and added to test samples at different noise conditions separately. The logarithmic spectrogram of a signal that contains event and non-event subspaces is generated. One-dimensional graph signal is obtained from logarithmic spectrogram. High energy spectral components belonging to the events are extracted from graph signal and considered as features. The SVM is used for AEC. The robustness of the features is tested in three different noise conditions and compare the results with MFCC baseline system,

The study of Andrey Temko [2] according to him, this paper focusing on acoustic events that may take place in meeting-rooms or classrooms and on the preliminary task of classifying isolated sounds. The number of sounds encountered in such environments may be large but in this initial work have chosen 16 different acoustic events including speech, music and a database has been defined for training and testing. While in the authors looked at the AEC problem from the point of view of speech recognition, applying the usual automatic speech recognition strategy, in their work they have consider to develop and compare several feature sets and classification techniques, aiming at finding the ones which are most appropriate for the problem are tackling,

Don stowell *et.al* [3] proposed methods for efficient feature representation and building proper machine learning algorithm using MFCCs and SVM for detecting audio events and taken 16 different audio events from the IEEE Audio and Acoustic Signal Processing Challenge Dataset, namely alert, clear throat, cough, door slam, drawer, keyboard, keys, knock, laughter, mouse, page turn, pen drop, phone, printer, speech, and switch that are collected from office live environments are utilized in the evaluations,, they performed three empirical studies. In the first phase performed MFCC feature extraction with different number of coefficients and in the second phase they performed MFCC feature extraction with different window and hop sizes to investigate how it affects the recognition rate. In the third phase optimize SVM parameters using 5-fold cross validation method and using grid search algorithm in order to increase evaluation results. The analysis results of performance with different number of MFCC coefficients 20 particular coefficients with the highest probability to give the most accurate results are selected. In order to define the optimal window and hop sizes and also studied different values to achieve best results using MFCC feature they obtain measure value of 55%. Since the precision difference is minor window size and hop size in terms of performance complexity are chosen to be 30ms and 10ms, respectively. In the SVM parameter optimization phase start with the default SVM parameters.

Carla Lopes *et.al* [4] proposed a speech event detector that segments speech signals in terms of four broad acoustic phonetic classes of events. Frame-based detection was carried out using Support Vector Machines (SVM). Non Negative Matrix De-convolution (NMD) was used in order to switch from a frame-based detection to a segment based detection. Results obtained using the TIMIT corpus are reported and compared to a broad class detector based on hidden Markov models (HMM) with a MFCC front-end. It was found that the proposed SVM/NMD system outperforms the HMM system in what concerns to accuracy and also to the quality of he detected boundaries.

J. Dennis [5] In this paper, we address the challenging task of simultaneous recognition of overlapping sound events from single channel audio. Conventional frame-based methods are not well suited to the problem, as each time frame contains a mixture of information from multiple sources, we propose an approach based on Local Spectrogram Features (LSFs) which represent local spectral information that is extracted from the two-dimensional region surrounding ‘‘key points’’ detected in spectrogram. LSF method performs consistently well, maintaining a low FA, whereas the performance of the baseline methods drops and the FA increases. Comparing the baselines, Overlap-SVM performs relatively well compared to Mix Max- GMM. Proposes an audio-event detection system. Their system consists of two- layered hidden Markov Model as back-end classifier. The optimization applied on the dataset returns F-measure value of 0.73 on frame-based measure. Present environmental sound and auditory scene detection system in their study. They use local discriminate bases (LDB) technique for time-frequency subspace distinction. In the study, they use 21 distinct audio events and they obtain average accuracy rate of 81% for this dataset. However, when a sort of dataset that involves background noise is used, overall accuracy decreases to 28.6%. Whenever thunderstorm and traffic events are taken into consideration, accuracy decreases to 52.0% and 66.7%, respectively since these two events are the subsets of the others. Presents some methods for their tasks such as feature extraction and classification. They use different audio features, namely MFCC, ZCR and SF with hierarchical HMM classifier For event based evaluations, the results on the development dataset gives Precision, Recall and F-measure value of 55.65%, 42.21% , and 47.53%, respectively.

Table-1: contains different methods of classification and efficiency

Methods of classification	Efficiency
HMM	55.65%
GMM	42.2%
NMF	49.3%
SVM	64%
DNN	69%

3. CONCLUSION

This review is based on the classification of sound events in challenging overlapping and noisy conditions System proposed a detection of overlap situations on sound events by analyzing the features of the audio data acquire by using microphones, in the first phase they perform MFCC feature extraction with different number of coefficients and in the second phase they performed MFCC feature extraction with different window and hop sizes to investigate how it affects the recognition rate. In the third phase they optimize SVM parameters using 5-fold cross validation method and using grid search algorithm in order to increase evaluation results and apply for SVM classifier this project utilizes a standard highlights, like MFCC and GMM for overlapping and SVM and DNN for classification, in order framework Utilizing recording of the distinctive common habitats.It is mentioned in the above table about the different method of classification used and there efficiency they get.

4. REFERENCES

[1] Manjunath Mulimani, Jahnavi U.P, Shashidhar G.Koolagudi Prof. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017 “Acoustic Event Classification using Graph Signals”

[2] Victor Bisot, Slim Essid, Gael Richard, ”HOG and Subband Power Distribution Image Features for Acoustic Scenes Classification”Institute Mines-Telecom, TelecomParisTech, CNRS LTCI, 37-39 rue Dareau2015 23rd European Signal Processing Conference.

[3] Gregoire Lafay, Mathieu Lagrange, Mathias Rossignol, Emmanouil Benetos,"A Morphological Model for Simulating Acoustic Scenes and Its Application to Sound Event Detection" ,IEEE/ACM Transaction on Audio, Speech, and Language Processing, Vol. 24, NO. 10, October 2016.

[4] EDahl Geiger. Geoffrey Hinton, Li Deng, Dong Yu, Abdel rahman Mohamed, Navdeep Jaitly, Andrew Senior and Brian Kingsbur. "Deep Nueral Networks for Acoustic Events", VOL. 18, NO. 8, October 2015

[5] Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, NavdeepJaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury, "Deep Neural Networks for Acoustic Events" , IEEE Transaction on Multimedia VOL. 18, NO. 8, October 2017.

[6] Jens Schröder, Stefan Goetze "Acoustic Scene Classification using Time-Delay Neural Networks and Amplitude Modulation Filter Bank Features" Detection and Classification of Acoustic Scenes and Events 2016 3September 2016, Budapest, Hungary.

[7] Philip Eichinski**, "** Clustering and visualization of long-duration audio recordings for rapid exploration avian surveys ", 2017 IEEE 13th International Conference on eScience.

[8] Jort F. Gemmeke1, Lode Vuegen1,2,3, Peter Karsmakers2,3,4, Bart Vanrumste2,3,4, Hugo Van hamme1, "An Exempler-Based NMF Approch to Audio Event Detection", 2016 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics October 20-23, 2016.

[9] Achyut Mani Tripathi,"AcousticEvent Classification using Cauchy Non-Negative Matrix Factorization and Fuzzy Rule-based Classifier", India Email: t.achyut@iitg.ernet.in 2015.

[10] Emmanouil Benetos , GregoireLafay , Mathieu Lagrange, and Mark D. "Detection of Overlapping Acoustic Events using A Temporarily –Constrained Probabilistic Model" Centrale de Nantes, France ´ 3 Centre for Vision, Speech and Signal Processing, University of Surrey, UK 2015.

[11] Dan clayton, Dimitrios Giannoulis, Emmanouil Benetos, Member, IEEE, Mathieu Lagrange, and Mark D. Plumbley, Fellow, IEEE "Detection and Classification of Acoustic Scenes and Events "IEEE Transaction on MultimediaVOL. 17, NO. 10, OCTOBER 2015.

[12] Kazumasa Yamamoto "Detection of Overlapping Acoustic Events Based on NMF with Shared Basis Vectors" 017 IEEE 6th Global Conference on Consumer Electronics (CGCA2017).

[13] Annamaria Mesaros, Toni Heittola, Tuomas Virtanen Department of Signal Processing Tampere University of Technology Tampere, Finland 2015 "TUT Database for Acoustic Scene Classification and Sound Event Detection"

[14] Jozef VAVREK, Matúš PLEVA, Jozef JUHÁR Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic vavrekjozef@centrum.sk, matus.pleva@tuke.sk " Acoustic events detection with Support Vector Machines"

[15] Inkyu Choi, Soo Hyun Bae, Sung Jun Cheon, Won Ik Cho, and Nam Soo Kim Dept. of Electrical and Computer Engineering and INMC, Seoul National University, "Seoul Acoustic Event Detection using Local Detector and Global Classifier"