Acoustic Event Classification using MFCC and MP Algorithm

[1]Gunavathi, [2]Geethesh, [3]Mohammad Rizwan Khader, [4]Leora D'souza, [5]Sunil B.N

[1,2,3,4]Students, [5]Assistant Professor
Department of Computer Science Engineering,
Sahyadri College of Engineering and Management, Mangaluru, Karnataka, India

***Abstract*—This paper presents our experiments on acoustic scene classification. Classification of acoustic scene to extract the audio features is a tough task because of the diversity in their nature. Extracting the appropriate features is an important task for increasing the efficiency of the classification. Using MFCC features alone will not achieve a maximum efficiency in characterization of acoustic scenes. Here we use MFCC algorithm along with MP Algorithm to classify the acoustic scenes. Here we use Support Vector Machine Algorithm to train the machine. Classifier model is built using libSVM for training samples and the machine is then able to recognize various acoustic Scenes. Results shows that Acoustic scene classification achieves better accuracy when MP algorithm is used along with MFCC.**

***IndexTerms*—MFCC, MP algorithm, classification, feature extraction**

I. INTRODUCTION

Machine Learning is a branch in artificial intelligence that is rapidly gaining attention these days. The machines are trained based on previous experiences in order to build a machine learning model and it is able to predict the future events based on the trained data. The study of acoustic scene classification is gradually receiving interest since last few decades. Acoustic is a branch of physics. Scene is nothing but the environment. Acoustic scene classification is the identification of environmental sound. There are three types of environment: indoor, outdoor and transportation. The execution of acoustic scene distinguishment framework may be straightforwardly proportional of the caliber of the dataset.

The major issues in acoustic scene classification is the diversity in its nature which makes grouping and classification of the individual events from the complex mixture of sound patterns tedious, identification of suitable features from the audio scenes is one of the issues as well. Based on the feature extracted from the environmental sound audio clips we need to train the machine. Once the machine is trained, it should be able to automatically identify the environment. For example, mobile automatically switched off when a person enters a meeting room by identifying the environment. For the effective SVM preparing and classification, extraction of apt features is vital. These features can be either time domain or frequency domain. Some of the features are Energy, Pitch, Amplitude, MFCC etc.

The MFCC features are utilized for Classification. It represents the short-term power spectrum of Sound, that is based on cosine transform of log power spectrum on a nonlinear mel scale of frequency. Maximum accuracy cannot be achieved by MFCC algorithm alone. Hence we utilize MP algorithm along with MFCC. The MP algorithm breaks down an audio signal into a straight extension of waveforms that are chosen from the predefined dictionary which is an accumulation of parameterized waveforms. Each waveform is called an atom. The molecule is described based on the type of its dictionary (Gabor functions, Haar wavelets, Fourier functions), different parameters (time, frequency, phase, window type, length) are likewise used to portray the atoms [1]. Here, we used Gabor functions with MP algorithm which gives both Time and frequency features. MP algorithm is an iterative algorithm. In each iteration, the MP algorithm chooses the atom that best matches with signal structure.

Gabor dictionary with MP algorithm gives Time-Frequency (TF) features. Once the extracted features are obtained, the machine is trained using Support Vector Algorithm. The main purpose behind usage of SVM classifier is that it takes the voice samples as input and classifies them according to their features. During the training phase, each feature is labelled which identifies the acoustic environment like home, bus etc. The kernel functions of support vector machine are used in the training process of SVM. A total of 20 voice samples of 24 different classes are used for the purpose of training the SVM. After the SVM is trained with various inputs from different acoustic scenes, it is ready to identify a new audio input. SVM is known to have high generalization capability due to properties like structural risk minimization oriented training. Once the accuracy of the classification is obtained, the classification result is displayed to the user. In order to increase the accuracy of the classification, new features can be added during the training phase.

II. PREVIOUS WORK

Over the past few years, scene classification has turned out to be one of the significant fields in audio signal processing. Almost all works on scene classification is based on feature extraction algorithm which is different for many acoustic scenes [5]. This is one of the toughest tasks regarding computational auditory scene analysis. Acoustic scene classification consists of collection of datasets, feature extraction and categorization of several acoustic scenes. Acoustic scene classification can be a

complicated task since the audio sample associated to a particular location is capable of containing large quantity of a single sound event while only some of these events may give useful details about the scene of the audio recording. Precisely, an audio scene is related to audio sample recorded at a specific location is expected to produce some acoustic events that differentiates it from other audio scenes. These selective acoustic events can be generated on various aspects and they may differ. This distinction is because of physical environment or human activities. Majority of the works on scene classification was to design the new methods and algorithms in order to obtain those features while trying to avoid this distinction.

Environmental scene audio samples are unstructured [2]. Sequence of acoustic scene data contains adequate information which facilitates the system to record a richer scene. The study of acoustic scene classification is increasingly gaining popularity since last few decades. But it is relatively less popular when compared to the study of speech recognition [7]. There are very few frameworks or systems that work on raw audio.It is necessary to choose the appropriate features in order to make the classification successful. Mel-frequency Cepstral Coefficients (MFCC) features are commonly used to extract feature out of audio clips. MFCC features were combined with Recurrence Quantification Analysis (RQA) features in order to obtain highly organized set of features. RQA features are used to improve the temporal representation of the frame level.MFCCs works well for structured sounds such as speech and music, but its accuracy reduces in the presence of noise [2]. MFCCs are not effective in analyzing noise-like signals that have a flat spectrum. Environmental audio contains a wide variety of sounds, including those with strong temporal domain signatures, such as chirpings of insects and sounds of rain that are typically noise-like with a broad flat spectrum that may not be effectively modeled by MFCCs.

Due to the randomness, high variance, and other difficulties in working with environmental sounds, the recognition accuracy reduces rapidly with growing number of classes. The analysis of sound environments in Peltonen's thesis which is closest to our works, presented two classification schemes. The first scheme was based on averaging the band-energy ratio as features and classifying them using a K-nearest neighborhood (kNN) classifier [6]. The second uses MFCC as features and a Gaussian mixture model (GMM) classifier. Peltonen noticed the limitations of MFCCs for environmental sounds and proposed using the band-energy ratio as a way to represent sounds occurring in different frequency ranges. The two experiments consisted of categorizing 13 different contexts or classes. The classifiers and types of features were comparatively similar to our experiments, however the actual types of classes were different. In a subsequent paper by Peltonen et al. they extended the investigation to audio-based context recognition by proposing a system that classifies 24 individual contexts. They sub-divided 24 contexts into six higher-level categories, with each category consisting of four to six contexts. Peltonen et al. also performed a listening test and reported the findings in [6]. Subjects were presented with 34 samples each one minute in duration, for the first experiment and 20 samples of three minutes each, in the second experiment. The tests were mostly conducted in a specialized listening room.

In Matching Persuit (MP) algorithm, the audio signals are broken into linear expansion of waveforms. It seperates signal structures that are consistent with respect to the given dictionary [4]. MP algorithm can be used in case of unstructured audio scenes [1]. Since using MFCC alone for unstructured acoustic scenes will not achieve a better efficiency we can make use of MP algorithm along with MFCC features.

III. ARCHITECTURAL MODEL

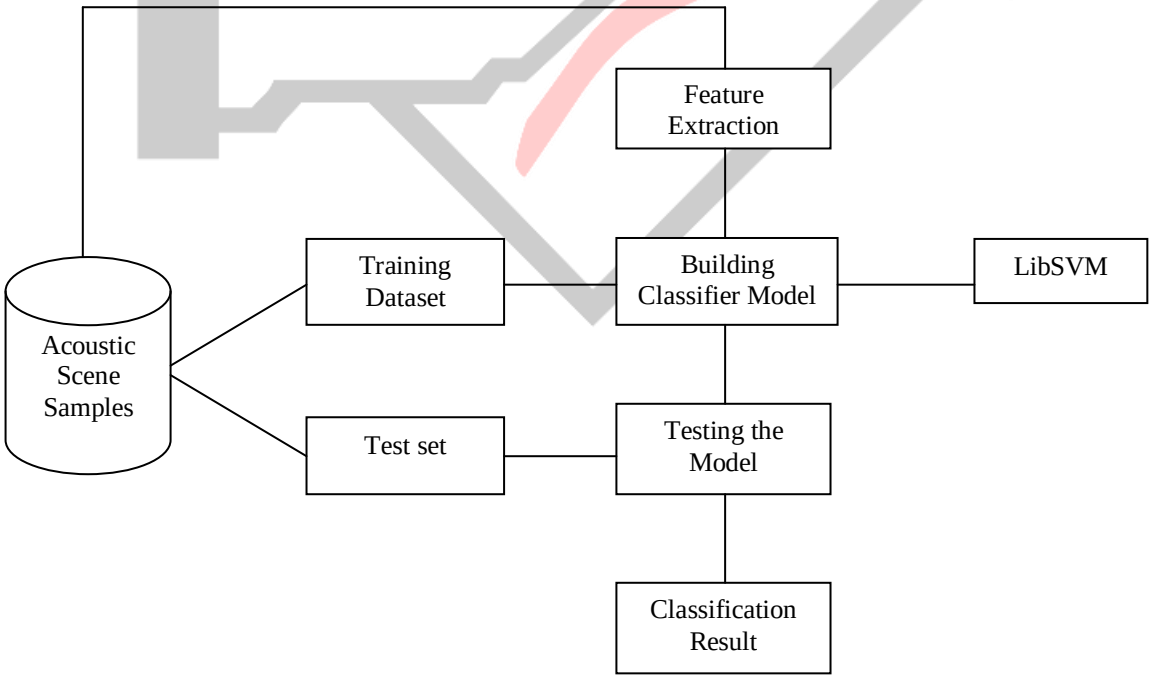

Fig.1 Architectural Model of ASC

Figure.1 shows the abstract version of the Acoustic Scene Classification. The acoustic scene database contains various acoustic scenes data. Acoustic features are extracted from the acoustic scenes for classification. Initially this dataset is classified into groups.

The datasets are trained and tested for better outcome. The extracted features from the acoustic database and the trained dataset are used for building the classifier model in order to build LibSVM library. Using the tested data, the classifier is tested. The classification result can be analyzed and evaluated.

IV. **RESULTS**

To carry out the Experiments on acoustic scene classification, we have created an acoustic scene database comprising of 600 audio files of 10 different acoustic scenes. Out of 600 audio files, 400 audio files are used for training and the remaining for testing i.e. There are a total of 40 audio samples of each class were used and total of 20 audio samples were used for training.

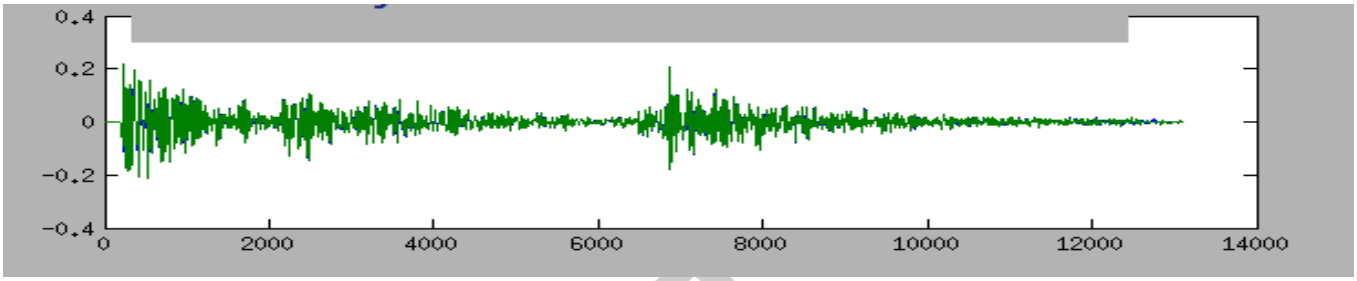

Fig.2 Plot of Short event

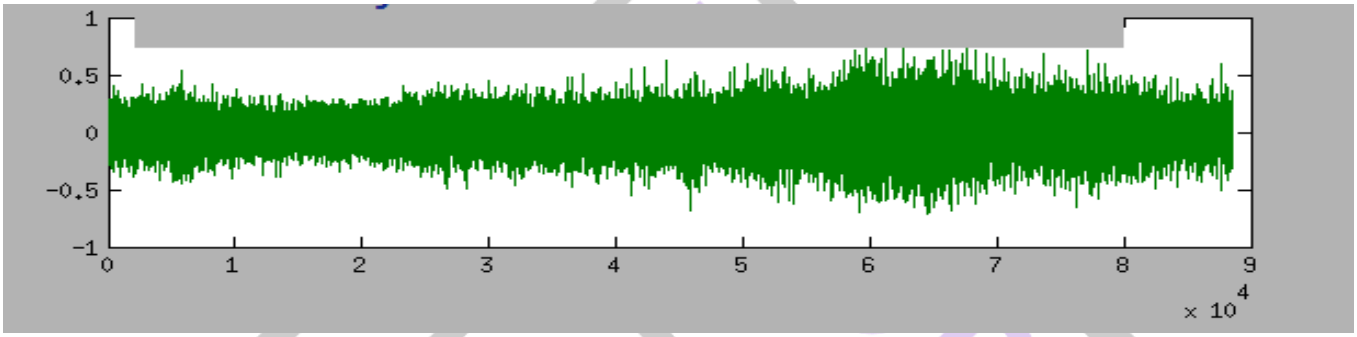

Fig.3 Plot of Long event

Figure.2 and figure.3 shows the graph of short events and long events respectively. This graph was plotted by taking time along X-axis and audio signal vectors along Y-axis.

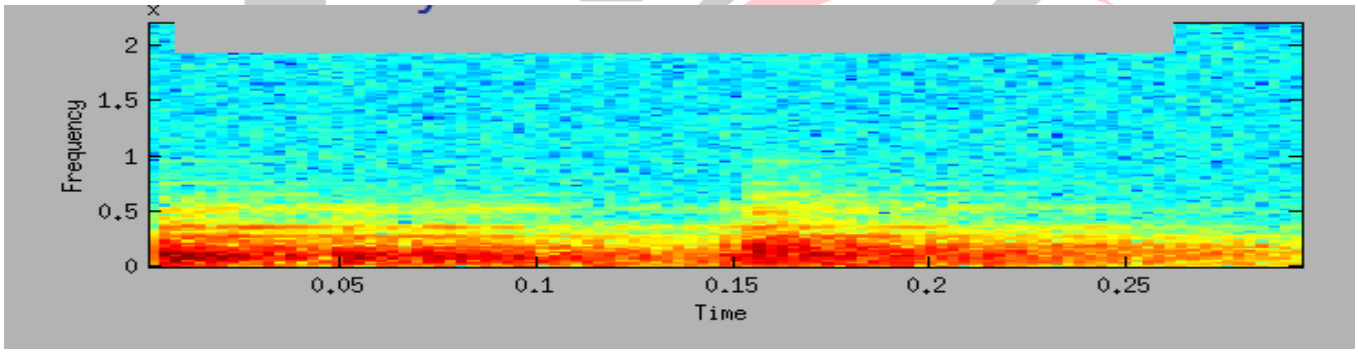

Fig.4 Spectrogram of Short event

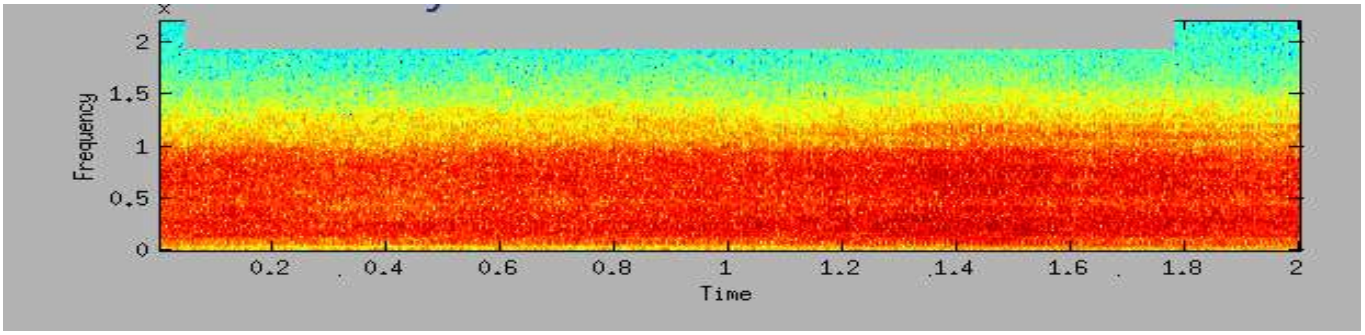

Fig.5 Spectrogram of Long event

Figure.4 and figure.5 shows the specification of short and long events respectively. Spectrogram represents the spectrum of frequencies in audio signal which vary with time. This graph is plotted by taking time along X-axis and frequency along Y-axis.

For the classification of acoustic scenes, we have used SVM algorithm. Since MFCC alone is inefficient in classification of acoustic scenes, we make use of MP algorithm along with the MFCC algorithm. Finally we achieve the accuracy of 72% when MFCC algorithm was used with MP algorithm. Experiment was also conducted by taking only MFCC features. In this case, accuracy was found to be 67.5%.

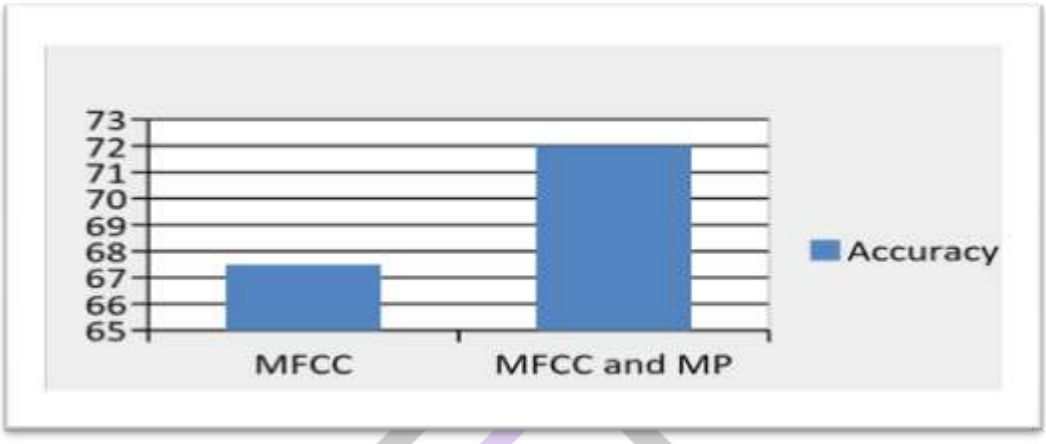

Fig.6 Accuracy comparison graph

Figure.6 depicts the comparison between the accuracy obtained with MFCC alone and MFCC used with MP algorithm. The overall accuracy of the classification of acoustic scene was found to be 72%.

V. CONCLUSION

This paper gives the overview of Acoustic scene classification. We have learnt from literature survey that MFCC algorithm has been used mainly in structured sound patterns which are not accustomed to noise. Hence we have implemented MP algorithm along with MFCC algorithm to process unstructured sound patterns. It has helped to give a better accuracy and classification of scenes with the usage of SVM model.

VI. FUTURE SCOPE

This paper gives the overview of Acoustic scene classification. We have learnt from literature survey that MFCC algorithm has been used mainly in structured sound patterns which are not accustomed to noise. Hence we have implemented MP algorithm along with MFCC algorithm to process unstructured sound patterns. It has helped to give a better accuracy and classification of scenes with the usage of SVM model.

VII. ACKNOWLEDGMENT

The authors thank Prof. Manjunath Mullimani for helping us develop the software. We would also like to extend our gratitude to Asst. Prof. Shailesh Shetty and Prof. Sudheer Shetty for their insightful comments and feedback.

REFERENCES

[1] Manjunath Mulimani, Shashidhar G.Koolagudi, "Acoustic scene classification using MFCC and MP features" in Detection and Classification of Acoustic Scenes and Events 2016, Budapest, Hungary,2016.J.ClerkMaxwell,ATreatiseonElectricityandMagnetism,3rded.,vol.2.Oxford:Clarendon,1892,pp.68–73.

[2] S. Chu and C.-C. J. Kuo, "Environmental sound recognition with time–frequency audio features," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 6, pp. 1142–1158, 2009.

[3] A. Rakotomamonjy and G. Gasso, "Histogram of gradients of time-frequency representations for audio scene detection," arXiv preprint arXiv:1508.04909, 2015.

[4] S. Mallat and Z. Zhang, "Matching pursuits with time–frequency dictionaries,"IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415,Dec. 1993.M.Young,TheTechnicalWriter'sHandbook.MillValley,CA:UniversityScience,1989.

[5] D. P. W. Ellis, "Prediction-driven computational auditory scene analysis,"Ph.D. dissertation, Dept. of Elect. Eng. and Comput. Sci., Mass.Inst. of Technol., Cambridge, MA, Jun. 1996.

[6] V. Peltonen, "Computational auditory scene recognition," M.S. thesis, Tampere Univ. of Technol., Tampere, Finland, 2001. A. Eronen, V. Peltonen, J. Tuomi, A. Klapuri, S. Fagerlund, T. Sorsa.

[7] G. Lorho, and J. Huopaniemi, "Audio-based context recognition,"IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 1, pp.321–329, Jan. 2006.