Acoustic Event Recognition for the Application of Surveillance

Supriya. A. M, Vandana Shree J S, Soujanya.H.N, Manjunath, Prof. Aneesh Jain. M.V
Department of Electronics & Communication
Alva's Institute of Engineering& Technology, Mijar, Moodbidri, Mangaluru, Karnataka, India

Abstract: **From few decades many systems was designed and proposed for automatically detecting the situations on road such as accidents to safe guard quick intervention of the emergency teams. However in some situations visual data is not efficient and sufficiently reliable. So use of audio event detectors and microphones along with the existing systems improves the overall reliability and obtains efficient results in surveillance systems.**

The name Acoustic Event Recognition (AER) deals with detection, classification and recognition of unstructured or unmannered environment which may contain overlapping sound events and non-stationary noises in the background. Many sounds along with noise contribute to the context of the surrounding environment. So noises must not be degraded. As such noises are commonly useful in Automatic Speech Recognition (ASR) which are also useful for many surveillance applications.

***Keywords*-Acoustic event recognition, Mel-Frequency Cepstral Coefficients, Perceptual Linear Prediction, tire skidding, car crashes.**

I. INTRODUCTION

Even though a variety of techniques has been developed for acoustic event recognition. The most standard approaches are often based on frame features like Mel-Frequency Cepstral Coefficients and Perceptual Linear Prediction (PLP) of ASR are used to train the Support Vector Machinefor classification process. While these methods are effective in ASR for particular source speech recognition, such systems may not perform well in the motivatingunequal conditions present in many AER tasks.

The aim of acoustic part classification is to classifytest recording into one of pre-defined classes that personifies the environment in which it was recorded driving the human ability to Categorize and recognize sounds and sounds capes. The dataset for example botanical gardens, path, workplace etc. The acoustic data will include recordings from 20 frameworks are taken to understanding the perceptual processes consists of recordings from various acoustic sections, all having distinct recording locations. Each audio file is recorded 2-6 minute long. The original audio files were then splitting into segments with a length of 10 seconds. These audio segments are provided to individual files.

The MFCC features, widely used for several audio recognition tasks like speech recognition or speaker identification, are sensitive to additive noise. When the energy of an event of interest decreases, it becomes comparable with the one of the background noise and it is more difficult to discriminate such events [1].The events are modelled using a network of hidden Markov models; their size and topology is chosen based on a study of isolated events recognition. For event detection, the system performs recognition and temporal positioning of a sequence of events. An accuracy of 24% was obtained [2]. An experiment with an acoustic surveillance system comprised of a computer and microphone situated in a typical office environment. All interesting acoustic events over duration of more than two months were recorded. A number of low-level signal features are computed from the audio signal and used to classify and identify sound events. The analysis reveals interesting activities which would be difficult to find by any other means [3]. The case where a typical situations such as screams, explosions and gunshots take place in a metro station environment. Their approach is based on a two stage recognition, each one exploiting HMMs for approximating the density function of the corresponding sound class. More precisely explosion sounds, corrupted by metro station environmental noise with −5dB ratio are detected with EER of 13.2%, gunshot sounds with 24.5% and scream sounds with 28.2% [4].

Surveillance in particular the case for the class of sounds considered in this paper, sounds produced by gun shots. They specifically focused on the robustness of the detection against variable and adverse conditions and the reduction of the false rejection rate which is particularly important in surveillance applications [5].

Audio-based video surveillance system which automatically detects anomalous audio events in a public square, such as screams or gunshots, and localizes the position of the acoustic source, in such a way that a video-camera is steered consequently. This system employs two parallel GMM classifiers for discriminating screams from noise and gunshots from noise, respectively. Experimental results show that their system can detect events with a precision of 93% at a false rejection rate of 5% when the SNR is 10dB, while the source direction can be estimated with a precision of one degree [6].

Different spectral structures between the speech and acoustic events degrade the performance of the speech feature sets. They proposed quantifying the discriminative capability of each component according to the approximated Bayesian accuracy and deriving a discriminative feature set for acoustic event detection. Compared to MFCC, feature sets derived using the proposed approaches achieve about 30% relative accuracy improvement in acoustic event detection [7]. A crucial part of a speech recognizer is the acoustic feature extraction, especially when the application is intended to be used in noisy environment. The investigation of several novel front-end techniques and compare them to multiple baselines. Recognition tests were performed on studio quality wide band recordings on Hungarian as well as on narrow band telephone speech includes real-life noises [8].

Focusing on improving security in public transports (here-trains), and presents an implementation on audio-video surveillance system. Combining audio analysis, video tracking anddedicated integration of audio-video acquisition and storage equipment's, the proposed system addresses the task of providing an operator with a partially-supervised tool for tracking suspected persons along the train. Satisfying tracking performances have been achieved on simple scenarios. Camera switch according to the person displacement is correctly done. However, the tracker quickly fails when the conditions become harder [9].

For this purpose, Natural Language Processing (NLP) is a very active area of research and development in communication technology. Important applications of NLP are machine translation and automatic speech recognition. For NLP, a basic unit of speech recognition is the intermediate step of speech information around which many of the recognition process is organized for human-beings or for machines. Investigators have proposed many different categories of intermediate things. Some of the possibilities include sub phoneme units, phones with right or left context biphones, diaphones and variations, dyads and events, trip hones, demi syllables, whole words and phrases. Speech recognition involves different functions

- ➢ Speech analysis
- ➢ Feature extraction
- ➢ Acoustic modelling
- ➢ Language modelling Recognition

Discourse acknowledgment frameworks can be ordered in a few distinct classes in view of the kind of discourse articulation, sort of speaker display, kind of channel and the kind of vocabulary that they can perceive.

II. ISSUES OF PROJECT

Multilevel classification for big audio data: More than one acoustic event occur simultaneously in the same interval of time

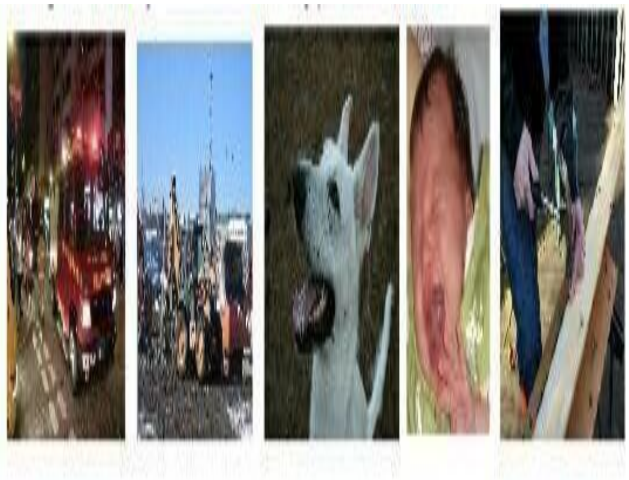


Figure A: Various events causing dissimilar audio frequencies.

Similar mix acoustic events: Similar category of sounds occur simultaneously.

Figure B: Similar events causing dissimilar audio frequencies

Recognition in highly noisy environments: Mixed mode acoustic events occurring at the same time.

Figure C: Example of noisy environment

A. Objective

- *Isolated Event Recognition:* Acoustic events occur in different interval of time.

  For example: Extracting fan sound orrefrigerator soundfrom the ambient sound.

Figure D: Example of isolated event

- *Overlapping Acoustic Event Recognition:* One or more acoustic event occurs at same interval of time instant.

For example: tires skidding, glass breaking and gun shoot occurs at same time.

Figure E: Example of for overlapping acoustic event

III. PROPOSED SYSTEM

Machine learning is a method in which the computers give the ability to learn without being explicitly programmed. Machine learning explores the study and creation of algorithms that can learn and make predictions on data such algorithms overcome following strictly static program instructions by making data-driven decisions, through building a model from sample inputs. It is employed in a range of computing tasks where designing and programmed explicitly with good performance is difficult. Example applications include electronic mail filtering, recognition of network intruder or malicious insiders working towards a data breach, Optical Character Recognition (OCR), learning towards rank and computer vision.

Machine learning is closely related to computational statistics, which focuses on prediction by the use of computers. It has strong ties to mathematical optimization, which delivers theory and application domains to the field. Machine learning is sometimes focus with data mining, then latter it focuses more on exploratory data analysis and is known as unsupervised Acoustic Event Recognition for Surveillance system.

Applications of speech processing

- Speech recognition for legal and forensic purposes of national security also for personalized services.
- Speech Enhancement for use in boisterous conditions, to take out resounds to adjust voices to video fragments. Possibly to enhance ability and expectation of discourse.
- Language translation to convert pronounced words in one language to another simplified ordinary language interchanges between people speaking in different languages i.e. travellers, business people

B. Need for Surveillance System

Crime Detection - This is the greatest and the most evident advantage of introducing surveillance cameras. When they are set, you will have the capacity to see their impact on individuals very quickly. Regardless of whether they are set tactfully, you will begin feeling a conviction that all is good, which is inestimable.

Monitor Consequences and Events -It is to a great degree simple to work with surveillance camera frameworks as they can be put anyplace insofar as there is a power source close by. They come in all shapes, sizes and some of them are small enough to be hidden in images, photo frames, etc. Based on our requirement we can buy either hidden cameras or mountable ones.

Gather Evidence - Having cameras installed in strategic places comes in handy when you need to monitor actions during an event. Modern security cameras are not only equipped with high-quality video capabilities, but also audio as well.

IV. METHODOLOGY

The methodology of the proposed system can be categorized mainly into two process.

1) Training process and 2) Prediction process.

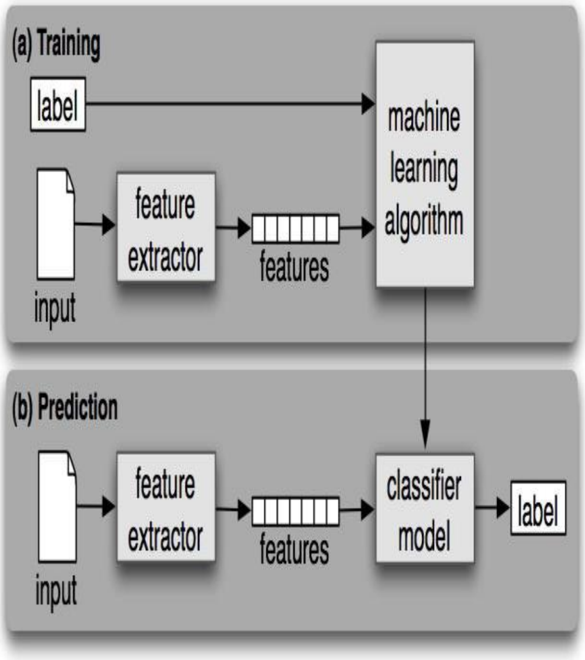


Figure E: Block diagram of supervised model

In training process, the audio file is given to feature extractor where the required key features are extracted and given to machine learning algorithm. On the other side label is given to the set of extracted features. This is all done during training period of the machine.

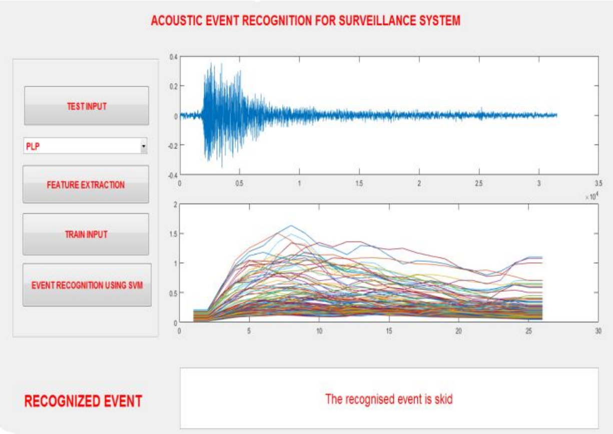
During prediction period or result period the input is given to feature extractor where the features are extracted and given to the trained classifier model where the machine algorithm is saved. After running the algorithm the label trained for the features during training period should be obtained as the label. This model is called as supervised learning model.

The supervised model uses two types of features.

1. Temporal features (Time domain features): which are simple to extract the features and have easy physical interpretation, like energy of the signal, zero crossing rate of the signal, maximum amplitude and minimum amplitude, minutest energy, etc.

2. Spectral features (Frequency based features): which are acquired by converting the time based audio signal into the frequency domain using the Fourier transformer. It extracts a features like fundamental frequency, frequency components of the signal, spectral centroid, and spectral flux.

V. RESULT ANALYSIS



ACOUSTIC EVENT RECOGNITION FOR SURVEILLANCE SYSTEM

PLP

RECOGNIZED EVENT

The recognised event is skid

Steps in analysing the results:

- In first real time audio is used as input.
- Based on feature extraction PLP or MFCC is used.
- Feature will be extracted.
- Data trained will be loaded to SVM classifier model.
- Real input and trained input will be compared in the event recognition using SVM.

PLP features are reported to be more robust when there is an acoustic mismatch between training and test data. In experiments we found that under clean conditions and when there is no significant mismatch, MFCC features lead to a performance that is slightly superior to PLP. Based on this advisable we decide for each task which one of the two feature types would be more appropriate in order to extract the feature. However, in many applications the acoustic conditions do not remain constant over the whole data set for instance and segments with clean speech are intermixed with segments that contain background music or noisy environment. In order to achieve optimal performance it is desirable to have a feature extraction that is well-suited both for clean and adverse acoustic conditions. Thus, the favourable properties of PLP and MFCC have to be combined

VI. APPLICATION

- Detect the road accidents by analysing the audio data in order to identify harmful situations such as tire skidding, glass breaking and car crashes.
- Detection of the unusual situations in the noisy surroundings.
- Military surveillance application.

Criminal event detection such as gun shoot.

VII. CONCLUSION

This system proposed a detection of harmful situations on roads by analysing the features of the audio data acquire by using surveillance microphones and apply for SVM classifier in order to detect the which harmful situation is happens in the road. This paper introduces a framework for acoustic occasion discovery in accounts from genuine conditions. This investigation utilizes a standard highlights, for example, Mel-Frequency Cepstral Coefficients (MFCC) and classifiers, for example, Perceptual Linear Prediction (PLP). At long last, this paper exhibits a determined assessment of SVM-based occasion location and order framework utilizing recording of the distinctive common habitats, for example, gunfire, vehicles floating sound and vehicle crash and sound occasions identified with human nearness, for example, discourse, cackling or hacking. It gets greatest precision of 80% for completely associated PLPs and MFCCs.

REFERENCES

[1]. Audio Surveillance of Roads: A System for Detecting Anomalous Sounds Pasquale Foggia, Nicolai Petkov, Alessia Saggese, Nicola Strisciuglio, and Mario Vento.

[2]. Annamaria Mesaros, Heittala, Eronen, virtanen- "Acoustic Event Detection in Real Life Recordings", 18th European Signal Processing Conference(EUSIPCO2010).JantoSkowronek,

[3]. Harma, Martin F, McKinney- "Automatic Surveillance of the Acoustic Activity in Our Living In Our Environment",digital Signal Processing Group, Philips Research.

[4]. Nikos Fakotakis, Ilyas, Stavros – "On Acoustic Surveillance of Hazardous Situations",Department of Electrical and Computer Engineering University of Patras

[5]. G.Richard, C.Clavel, T.Ehrette –"Events Detection for an Audio-Based Surveillance System", Thales Research and Technology, France.

[6]. L.Gerosa, G.Valenzise, A.Sarti- "Scream and Gunshot Detection and Localization for AudioSurveillance System", Department of Electronics and Communication – Politecnico di Milano.