ISSN: 2321-7782 (Online)
Volume 2, Issue 11, November 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Predictive Sound Recognization System

Ajay R.Kadam[1]
Department of Computer Engineering
Dr. D. Y. Patil SOET
Lohegaon, Pune.

Ramesh Kagalkar[2]
Assistant Professor Department of Computer Engineering
Dr. D. Y. Patil SOET
Lohegaon, Pune.

Abstract: ***The proposed research goal is to develop a system for automatic detection of sound. In this system the major task is to identify any input sound stream analyse it & predict the possibility of different sounds appear in it. To develop and commercially deployed a flexible sound search engine. The algorithm is noise and distortion resistant, computationally efficient, and massively scalable, capable of quickly identifying a short segment of sound stream captured through a cell phone microphone in the presence of foreground voices and other dominant noise, and through voice codec compression, out of a database of over available tracks. The algorithm uses a combinatorial hashed time-frequency constellation analysis of the sound, yielding unusual properties such as transparency, in which multiple tracks mixed together may each be identified.***

Keywords: ***Acoustic event detection and classification, voice activity detection, voice interface, continuously listening environment***

I. INTRODUCTION

We examine user-friendly voice interface that requires the hands-free sound acquisition in the Sample sound streams. The traditional voice activity detection (VAD) algorithms cannot successfully identify potential acoustic event sounds from sound. This makes the sound recognition system frequently or incorrectly activated. In this paper, we propose a novel voice activity detection technique that consists of two major modules: 1) classification and 2) detection module. In the classification module, we label the successive sound segments based on the training models. Then, in the detection module, we remove the acoustic event sounds and make decision of the explicit utterance boundary from the input sound stream. As a result, we proposed technique enables the efficient operation of sound recognition in the Sample sound streams without any touch and/or key input. Experiments in a real-world environment and performance comparison with state-of-the-art techniques are conducted to demonstrate the effectiveness of the proposed technique.

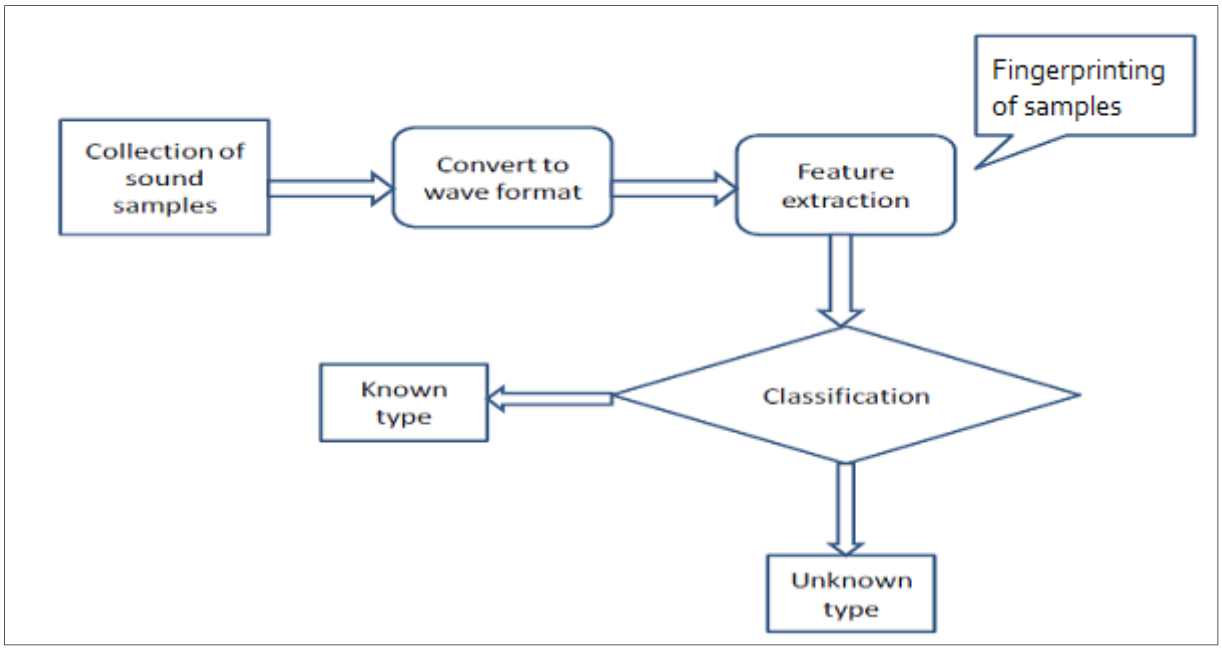
In general, we have seen two types of applications, offline and online processing. An offline application processes sound data in batches, such as in media search or transcription, while an online application typically detects and classifies short events when a real-time response is required, such as in security surveillance and hearing aids. Online applications are typically more challenging than offline ones because we have to make decisions based on short sound signals. As part of an ongoing project for sound surveillance, we are particularly interested in accurately classifying short sound events. This paper is focused on the techniques required for offline applications. Despite its importance is sound event recognition is still relatively new when compared to related topics such as speech recognition, speaker recognition, or music classification. It was also found that the majority of previous works focused on very narrow and specific problems of sound recognition. For example, recognition of bird sounds was studied. Environmental sound recognition through mobile devices is difficult because of background noise, unseen sound events, and changes in sound channel characteristics due to the phone's context, e.g., whether the phone is in the user's pocket or in his hand. We propose a crowd sourcing framework that models the combination of scene, event, and phone context to overcome these issues.

© 2014, IJARCSMS All Rights Reserved

II. TECHNOLOGY VISION MISSION

Working Principles

The proposed system gives hands on technical tool to elaborate and to find the classes of different from its sound pattern of a sample space. This is possible by our proposed approach. In this approach data base is created, which consists of .wave form of different activity, event, scenes sounds and its information of each sample model of various types. Once the sound sample query is given to the system first it converts into .wave format and extract features. Each sound file is "fingerprinted," a process in which reproducible hash tokens are extracted. Both "database" and "sample" sound files are subjected to the same analysis. The fingerprints from the unknown sample are matched against a large set of fingerprints derived from the sample database. The candidate matches are subsequently evaluated for correctness of match. Some guiding principles for the attributes to use as fingerprints are that they should be temporally localized, translation-invariant, robust, and sufficiently entropic. The temporal locality guideline suggests that each fingerprint hash is calculated using sound samples near a corresponding point in time, so that distant events do not affect the hash. The translation invariant aspect means that fingerprint hashes derived from corresponding matching content are reproducible independent of position within an sound file, as long as the Temporal locality containing the data from which the hash is computed is contained within the file. This makes sense as an unknown sample could come from any portion of the original sound track. Robustness means that hashes generated from the original clean database track should be reproducible from a degraded copy of the sound. Furthermore, the fingerprint tokens should have sufficiently high entropy in order to minimize the probability of false token matches at non-corresponding locations between the unknown sample and tracks within the database. Insufficient entropy leads to excessive and spurious matches at non-corresponding locations, requiring more processing power to cull the results, and too much entropy usually leads to fragility and non-reproducibility of fingerprint tokens in the presence of noise and distortion.

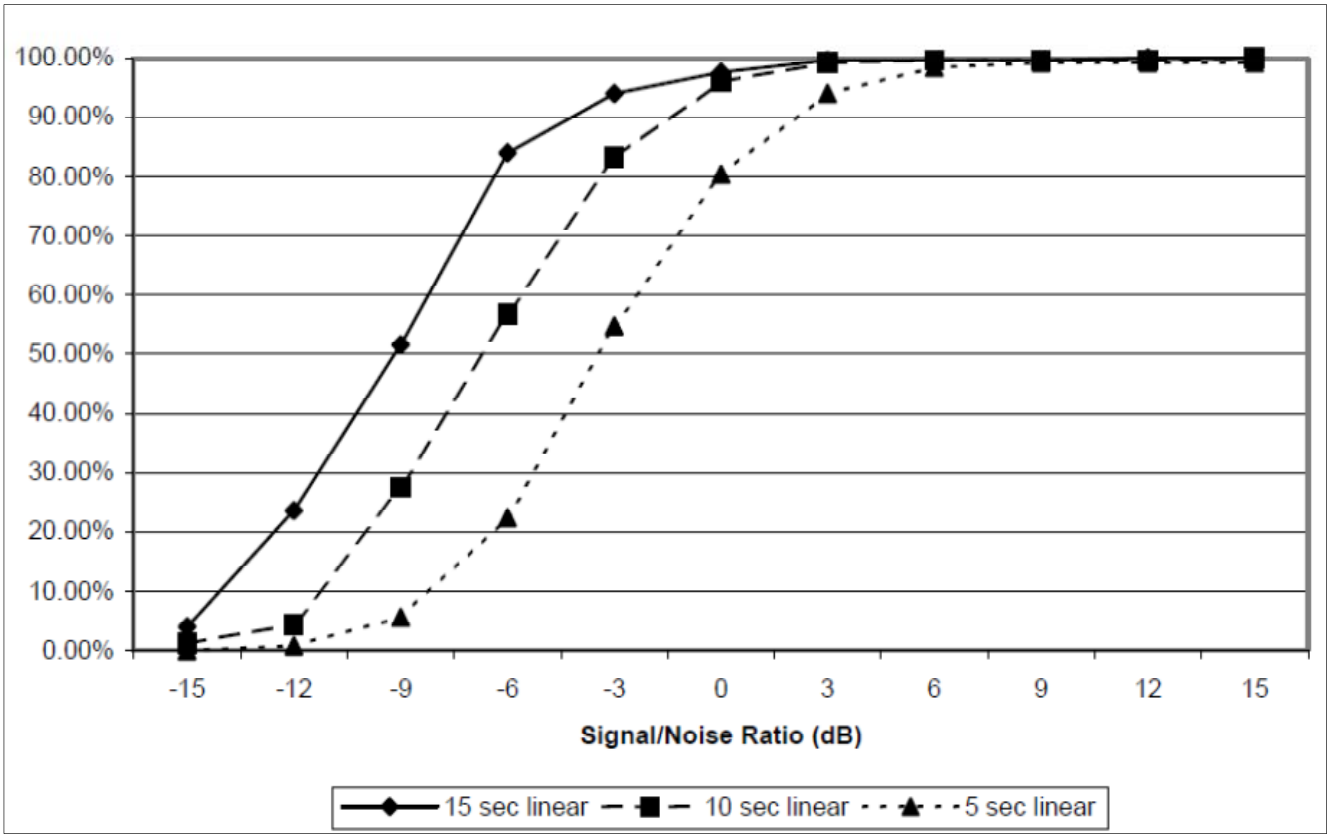


Methodology & Analysis

In order to address the identification in the presence of highly significant noise and distortion, we experimented with a variety of candidate features that could survive GSM encoding in the presence of noise. We settled on spectrogram peaks, due to their robustness in the presence of noise and approximate linear superposability. A time-frequency point is a candidate peak if it has higher energy content than all its neighbors in a region centered on the point. Candidate peaks are chosen according to a density criterion in order to assure that the time-frequency strip for the audio file has reasonably uniform coverage. The peaks in each time-frequency locality are also chosen according amplitude, with the justification that the highest amplitude peaks are

© 2014, IJARCSMS All Rights Reserved

most likely to survive the distortions listed above. Thus, a complicated spectrogram, as illustrated in Figure. A may be reduced to a sparse set of coordinates, as illustrated in Figure. Notice that at this point the amplitude component has been eliminated. This reduction has the advantage of being fairly insensitive to EQ, as generally a peak in the spectrum is still a peak with the same coordinates in a filtered spectrum (assuming that the derivative of the filter transfer function is reasonably small peaks in the vicinity of a sharp transition in the transfer function are slightly frequency-shifted). We term the sparse coordinate lists "constellation maps" since the coordinate scatter plots often resemble a star field. The pattern of dots should be the same for matching segments of audio. If you put the constellation map of a database song on a strip chart, and the constellation map of a short matching audio sample of a few seconds length on a transparent piece of plastic, then slide the latter over the former, at some point a significant number of points will coincide when the proper time offset is located and the two constellation maps are aligned in register. The number of matching points will be significant in the presence of spurious peaks injected due to noise, as peak positions are relatively independent; further, the number of matches can also be significant even if many of the correct points have been deleted. Registration of constellation maps is thus a powerful way of matching in the presence of noise and/or deletion of features. This procedure reduces the search problem to a kind of "astronavigation," in which a small patch of time-frequency constellation points must be quickly located within a large universe of points in a strip-chart universe with dimensions of band limited frequency versus nearly a billion seconds in the database.



Fingerprint hashes are formed from the constellation map, in which pairs of time-frequency points are combinatorial associated. Anchor points are chosen, each anchor point having a target zone associated with it. Each anchor point is sequentially paired with points within its target zone, each pair yielding two frequency components plus the time difference between the points. These hashes are quite reproducible, even in the presence of noise and voice codec compression. Furthermore, each hash can be packed into a 32-bit unsigned integer. Each hash is also associated with the time offset from the beginning of the respective file to its anchor point, though the absolute time is not a part of the hash itself. To create a database index, the above operation is carried out on each track in a database to generate a corresponding list of hashes and their associated offset times. Track IDs may also be appended to the small data struts, yielding an aggregate 64-bit strut, 32 bits for the hash and 32 bits for the time offset and track ID. To facilitate fast processing, the 64-bit struts are sorted according to hash

© 2014, IJARCSMS All Rights Reserved

token value. The number of hashes per second of audio recording being processed is approximately equal to the density of constellation points per second times the fan-out factor into the target zone. For example, if each constellation point is taken to be an anchor point, and if the target zone has a fan-out of size F=10, then the number of hashes is approximately equal to F=10 times the number of constellation points extracted from the file. By limiting the number of points chosen in each target zone, we seek to limit the combinatorial explosion of pairs. The fan-out factor leads directly to a cost factor in terms of storage space. By forming pairs instead of searching for matches against individual constellation points we gain a tremendous acceleration in the search process.

III. Literature Review

Unlike other audio or speech signals, sound events have a relatively short time span. They are usually distinguished by their unique spectra-temporal signature. This paper proposes a novel classification method based on probabilistic distance support vector machines (SVMs). We study a parametric approach to characterizing sound signals using the distribution of the sub band temporal envelope (STE), and kernel techniques for the sub band probabilistic distance (SPD) under the framework of SVM. We show that generalized gamma modeling is well devised for sound characterization and that the probabilistic distance kernel provides a closed form solution to the calculation of divergence distance, which tremendously reduces computational cost. We conducted experiments on a database of ten types of sound events. The results show that the proposed classification method significantly outperforms conventional SVM classifiers with Mel-frequency campestral coefficients (MFCCs). The rapid computation of probabilistic distance also makes the proposed method an obvious choice for online sound event recognition [1].

In this paper, we study the spectral and temporal periodicity representations that can be used to describe the characteristics of the rhythm of a music audio signal. A continuous-valued energy-function representing the onset positions over time is first extracted from the audio signal. From this function we compute at each time a vector which represents the characteristics of the local rhythm. Four feature sets are studied for this vector. They are derived from the amplitude of the discrete Fourier transform (DFT), the auto-correlation function (ACF), the product of the DFT and the ACF interpolated on a hybrid lag/frequency axis and the concatenated DFT and ACF coefficients. Then the vectors are sampled at some specific frequencies, which represent various ratios of the local tempo. The ability of these periodicity representations to describe the rhythm characteristics of an audio item is evaluated through a classification task. In this, we test the use of the periodicity representations alone, combined with tempo information and combined with a proposed set of rhythm features. The evaluation is performed using annotated and estimated tempo. We show that using such simple periodicity representations allows achieving high recognition rates at least comparable to previously published results[2].

A music piece can be considered as a sequence of sound events which represent both short-term and long-term temporal information. However, in the task of automatic music genre classification, most of text-categorization-based approaches could only capture temporal local dependencies (e.g., unigram and bigram-based occurrence statistics) to represent music contents. In this paper, we propose the use of time-constrained sequential patterns (TSPs) as effective features for music genre classification. First of all, an automatic language identification technique is performed to tokenize each music piece into a sequence of hidden Markov model indices. Then TSP mining is applied to discover genre-specific TSPs, followed by the computation of occurrence

Frequencies of TSPs in each music piece. Finally, support vector machine classifiers are employed based on these occurrence frequencies to perform the classification task. Experiments conducted on two widely used datasets for music genre classification, GTZAN and ISMIR2004Genre, show that the proposed method can discover more discriminative temporal structures and achieve a better recognition accuracy than the unigram and bigram-based statistical approach [3].

Environmental audio recognition through mobile devices is difficult because of background noise, unseen audio events, and changes in audio channel characteristics due to the phone's context, e.g., whether the phone is in the user's pocket or in his hand. We propose a crowd sourcing framework that models the combination of scene, event, and phone context to overcome these issues. The framework gathers audio data from many people and shares user-generated models through a cloud server to

© 2014, IJARCSMS All Rights Reserved

accurately classify unseen audio data. A Gaussian histogram is used to represent an audio clip with a small number of parameters, and a k-nearest classifier allows the easy incorporation of new training data into the system [4].

IV. Research Objectives

The research aims towards the implementation of sound recognization in the following points

1. Deep analysis of sampled sounds
2. Classification of sound into various categories.
3. FINGERPRINTING the sampled sound streams: - a process in which reproducible hash tokens are extracted.
4. Exact matching of sample space with input sound stream.
5. Prediction of exact results.

V. Conclusion

We examined technical barriers in the user-friendly voice interface thus want to develop a system which will identify and predict sound into different classes. Since the conventional VAD algorithms could not successfully identify the potential acoustic event sounds from speech, we proposed analysis and methodology for the same. The algorithm uses a combinatorial hashed time-frequency constellation analysis of the audio, yielding unusual properties such as transparency, in which multiple sounds mixed together may each be identified. The algorithm was designed specifically to target recognition of sound files that are already present in the database. It is not expected to generalize to live recordings.

Acknowledgement

The authors would like to thank Chairman Groups and Management and the Director/Principal Dr. Uttam Kalwane, Colleague of the Department of Computer Networking and Colleagues of the various Department the D. Y. Patil School of Engineering and Technology, Pune Dist. Pune Maharashtra, India, for their support, suggestions and encouragement.

References

1. Huy Dat Tran, Member, IEEE, and Haizhou Li, Senior Member, IEEE "SOUND EVENT RECOGNITION WITH PROBABILISTIC DISTANCE SVMS" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 6, AUGUST 2011.
2. Geoffroy Peeters "SPECTRAL AND TEMPORAL PERIODICITY REPRESENTATIONS OF RHYTHM" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 5, JULY 2011.
3. Jia-Min Ren, Student Member, IEEE, and Jyh-Shing Roger Jang, Member, IEEE "Discovering Time-Constrained Sequential Patterns for Music Genre Classification" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 4, MAY 2012.
4. Kyuwoong Hwang and Soo-Young Lee, Member, IEEE "Environmental Audio Scene and Activity Recognition through Mobile-based Crowdsourcing" IEEE Transactions on Consumer Electronics, Vol. 58, No. 2, May 2012.
5. Namgook Cho, Member, IEEE and Eun-Kyoung Kim "Enhanced Voice Activity Detection Using Acoustic Event Detection and Classification" IEEE Transactions on Consumer Electronics, Vol. 57, No. 1, February 2011.

Author(s) Profile

Ajay Kadam received the B.E. degree in Computer Science & Engineering in 2012 and now pursuing M.E. degree in Computer Networking from Dr. D. Y. Patil School of Engineering and Technology in current academic year 2014-15. He is now studying for the domain Sound Processing as research purpose on Predictive sound recognization.

Prof. Ramesh Kagalkar (P. Hd. Scholar) now assistant professor in department of computer engineering Dr. D. Y. Patil School of Engineering and Technology Lohegaon , Pune. He is now in the research field of area of Image Processing domain.

© 2014, IJARCSMS All Rights Reserved