ANKARA YILDIRIM BEYAZIT UNIVERSITY GRADUATE
SCHOOL OF NATURAL AND APPLIED SCIENCES

ENVIRONMENTAL SOUND RECOGNITION WITH VARIOUS FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES

M.Sc. Thesis by

Yaseminhan ARPACI

Department of Electrical and Electronics Engineering

July, 2019
ANKARA

ENVIRONMENTAL SOUND RECOGNITION WITH VARIOUS FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES

A Thesis Submitted to

The Graduate School of Natural and Applied Sciences of

Ankara Yıldırım Beyazıt University

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical and Electronics Engineering, Department of Electrical and Electronics Engineering

by

Yaseminhan ARPACI

July, 2019

ANKARA

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"ENVIRONMENTAL SOUND RECOGNITION WITH VARIOUS FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES"** completed by **YASEMİNHAN ARPACI** under the supervision of **PROF. DR. HÜSEYIN CANBOLAT** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Hüseyin Canbolat

Supervisor

Assoc. Prof. Dr. Asaf Behzat ŞAHİN

Jury Member

Prof. Dr. Murat YÜCEL

Jury Member

Prof. Dr. Ergün ERASLAN

Director

Graduate School of Natural and Applied Sciences

ETHICAL DECLARATION

I hereby declare that, in this thesis which has been prepared in accordance with the Thesis Writing Manual of Graduate School of Natural and Applied Sciences,

- All data, information and documents are obtained in the framework of academic and ethical rules,

- All information, documents and assessments are presented in accordance with scientific ethics and morals,

- All the materials that have been utilized are fully cited and referenced,

- No change has been made on the utilized materials,

- All the works presented are original,

and in any contrary case of above statements, I accept to renounce all my legal rights.

Date: **Signature:**

Name & Surname: Yaseminhan ARPACI

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Hüseyin CANBOLAT for his enormous support and motivation during my study. His knowledge and precious recommendations constituted the milestones of this study. His guidance assisted me all the time of my research and while writing this.

2019, 08 July **Yaseminhan ARPACI**

ENVIRONMENTAL SOUND RECOGNITION WITH VARIOUS FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES

ABSTRACT

This thesis proposes “An Environmental Sound Recognition with various feature extraction and classification techniques” for environmental sound recognition. Study in Environmental Sound Recognition (ESR) has taken attention in recent years. In the past decade, research on the Environmental Sound Recognition (ESR) area has accelerated. ESR has important role on intelligent computer systems and robots for the purpose of identification, recognition and discrimination. In this survey, I will put forward a survey on which various feature extraction and classification techniques is better to recognize environmental sounds. Survey includes these parts: environmental sound recognition system processing, feature extraction techniques, classification techniques, and performance comparison of selected techniques. At long last, finishing up comments and future innovative work slants in the ESR field will be given.

Keywords: Environmental sound recognition, feature extraction techniques, classification techniques

ÇEŞİTLİ ÖZNİTELİK ÇIKARTMA VE SINIFLANDIRMA METOTLARIYLA ÇEVRESEL SES TANIMA

ÖZ

Bu tezde, çevresel ses tanıma amaçlı "Öznitelik Çıkartma ve Sınıflandırma Metotları" önerilmektedir. Çevresel ses tanıma çalışması, son yıllarda popüler bir konu haline gelmiştir. Geçtiğimiz on yılda, çevresel ses tanıma alanı ile ilgili araştırmalar hızlandı. Çevresel ses tanıma, akıllı bilgisayar sistemleri ve robotlar için önemli bir role sahiptir. Bu araştırmada, çevresel sesleri tanımaya yönelik çeşitli öznitelik çıkarımı ve sınıflandırma tekniklerini kullanılarak, bu teknikler üzerinden kıyaslamalar yapılarak en verimli şekilde ses tanıma yapılma amaçlanmıştır. Araştırma üç kısmı içeriyor, bunlar: temel çevresel ses tanıma, öznitelik çıkarma teknikleri, sınıflandırma teknikleri ve seçilen tekniklerin performans karşılaştırması. Son olarak, ESR alanındaki sonuca yönelik açıklamalar ve gelecekteki araştırma ve geliştirme eğilimleri verilecektir.

Anahtar Sözcükler: Çevresel ses tanıma, özellik çıkarımı, sınıflandırma metotları

CONTENTS

NOMENCLATURE

Acronyms

ESR	Environmental Sound Recognition
CWT	Continuous Wavelet Transform
STFT	Short-Time Fourier Transform
KNN	K nearest neighborhood
SVM	Support Vector Machine
MFCC	Mel-Frequency Cepstral Coefficients

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Environmental Sound define as a product or manufacturing process that, from beginning to end, is in essential harmony with its environment and the associated ecological factors [1]. Other define is proven by objective evidence that activities taken are environmentally positive. Harmony stems from the product or manufacturing processes that improve its environment and the associated ecological factors [2]. However, it isn't so basic for an environmental sound recognition system.

Practically all research on sound-related data is done by concentrating on individual investigations on acoustic flag preparing, hearing handling and discourse correspondence. Notwithstanding, the most significant point to comprehend the sound scene is the nearby collaboration and joining of these capacities. So as to perceive a particular sound, the framework ought to limit the objective sound among an expansive number of encompassing sounds blends and center the sound. The accumulation of the sound scene information in genuine sound situations is essential for the examination of the sound scene. Sound stage database contributes to the study of the sound stage understanding. [3].

Environmental Sound Recognition (ESR) is on a very basic level established on artificial intelligence and machine learning estimations require starting a readiness database. This database must incorporate the sounds to be seen and other related sounds. An ESR framework needs the database during preparing, testing and in the generation arrange [4].

The applications of ESR can be listed as follows: In military, legal and law authorization area there are thinks about on shot discovery frameworks. In [5], a gunfire discovery framework is proposed. In [6], the shot impact is utilized to recognize the bore of the firearm. In [7] and [8] ESR is utilized for robot route. ESR can be utilized for home observing. It very well may be utilized to help older individuals living in their home alone [9, 10]. In [11], it is utilized for home automation... In [12] and [13], ESR is utilized for acknowledgment of creature sounds.

In [13], system used a large database with 1418 animal vocalizations annotated with the species and additional metadata, including recording conditions and type of vocalization for each audio file. In the observation region, it is utilized for reconnaissance of street [14], open transport [15], lift [16] and office passage [17]. There are a lot of papers in literature on environmental sound recognition [18-21]. These articles mention that during the improvement of an environmental speech recognition system, many sound clasps, a large number of these sounds, and numerous calculations that utilization these ascribes should be contemplated. When the framework is operational and prepared to be grown, new solid sorts and their new stable clasps, new characteristics and calculations that ought to be perceived should be included. Establishing a database structure that will facilitate these studies will be very useful [22].

The objective of this thesis is design environmental sound recognition system with various feature extraction and classification techniques and then compare these techniques to obtain the most successful result.

1.1 Literature Survey

Application of this thesis we use http://wiki.cnbc.cmu.edu/Sound_Databases database. There are also other databases which other researchers can use http://www.auditorylab.org is developed by Carnegie Mellon University. http://www.desra.org is a general-purpose environmental sound database. In [23] explains the purpose of this database, the design details and the scientific resources used. The database published at http://www.daresounds.org/.This database is a database where the voices we hear in the complexity of everyday life are recorded with explanations and are accessible to the scientific world.

1.2 Motivation

Environmental sounds involve a wide range of sound that are neither discourse nor music making the space almost vast in size. A lot of research has been made towards displaying and recognition of environmental sounds over the previous decade. By environmental sounds, we allude to different daily sounds, both natural and unnatural (for example sounds one experiences in day by day life other than discourse and music). Environmental Sound Recognition (ESR) has an essential influence in late endeavors to consummate machine tryout. This provides the motivation to develop and design the most efficient environmental sound recognition system.

1.3 Organization of Thesis

This thesis is organized as follows: In Chapter 2, a thorough overview of the field of environmental sound and environmental sound recognition is provided. Chapter 3 then introduces future extraction techniques. In Chapter 4, shows classification techniques. Chapter 5 then shows implement and use of future extraction and classification techniques together and compare these techniques.

CHAPTER 2

ENVIRONMENTAL SOUND RECOGNITION

2.1 Environmental Sound

The perception of environmental sounds has only recently begun to receive the level of attention that speech and music perception have enjoyed for many years. Given the prevalence of environmental sounds (defined here as all naturally occurring sounds other than speech and music) in everyday life and, importantly, throughout the evolution of the mammalian auditory system, this class of sounds certainly deserves more attention [24], Environmental Sound has different definitions. The perception of environmental sound is diversified in the literature. As in [20], the wind in the trees can be caused by the sounds of nature in the nature, such as the sound of rain or the chirping of insects. In [1], Environmental Sound defined as all naturally occurring sounds other than speech and music. In this thesis we use rolling and air event environmental sound.

2.2 Environmental Sound Recognition Overview

While research on sound recognition has traditionally focused on speech and music signals, the issue of environmental sound recognition (ESR) has received more attention in last years. ESR research has increased significantly in the last decade. [25].

Over the last decade, a significant amount of research has been done to model and identify environmental sounds. With environmental sounds, we have referred (that is, everyday sounds, except speech and music) to various voices that are natural and artificial. Environmental sound recognition (ESR) plays an important role in the near-term efforts to perfect machine audition [25]. With expanding test-based demands, for example, content-based picture and video search, ESR has started to be utilized as an instrument in productive voice search applications. [26]. ESR can be utilized for programmed labeling of sound documents with identifiers for catchphrase-based sound gathering [27]. By including ESR in the system, robot navigation can be advanced. [7, 8]. ESR can be utilized in the home observing condition to help older individuals living alone in their very own home [9, 10] or for a shrewd home [11]. ESR, alongside picture and video examination, discover applications in observation [28, 29]. ESR can likewise be adjusted for acknowledgment by particular hints of creature and winged creature species [12, 13].

Speech and music among the various types of sound signals are two categories that are examined extensively. The ESR algorithms are only a reflection of the speech and music recognition paradigms in the first period. However, due to the non-stationary nature of ambient sounds, these algorithms are insufficient for large-scale databases. [25].

The reason for ESR frameworks is to perceive a sound occasion that outcomes in an individual sound source in a ceaseless sound signal. The fundamental concern is to identify the occasion in a given concentrate of the chronicle. The exact area of the occasion, for example the begin and end times of the occasion, has an optional prefix, so location frameworks frequently have an unpleasant worldly goal. This is the place ESRs vary from other sound-based acknowledgment frameworks, for example, discourse and music acknowledgment. Past ESR considers we are normally performed for secluded sound occasions [30]. Each record has a solitary sound occasion, and the ESR is prepared to distinguish the occasion at a predefined time interim. ESR is performed on records gathered from genuine situations. These conditions have distinctive acoustical properties because of the nearness of human or nature exercises. There are numerous covering sound occasions, so the discovery of individual occasions is more testing than the confined accounts. An ESR framework fit for distinguishing various occasions in a similar circumstance is known as a polyphonic ESR framework. ESR frameworks first structure a middle of the road portrayal of the sound waveform in a record. This procedure is called highlight extraction. Since irrelevant features make it difficult to detect because it can lead to loss of valuable information, it is very important to choose the right features to extract from an audio signal. There are strong signs of time in many cases of nature, such as wind and rain. For this reason, it is somewhat necessary to extract temporal domain knowledge and spectral domain knowledge in the environmental sound recordings [31].

2.3 Basic Environmental Sound Recognition System Design

A lot of research has been made towards displaying and acknowledgment of natural sounds over the previous decade. By ecological sounds, we allude to different quotidian sounds, both regular and counterfeit (for example sounds one experiences in every day life other than discourse and music). Ecological sound acknowledgment (ESR) has an essential influence in late endeavors to immaculate machine audition[25]. With a developing interest on precedent based inquiry, for example, content-based picture and video search, ESR can be instrumental in productive sound hunt applications [26]. ESR can be utilized in programmed labeling of sound records with descriptors for watchword based sound recovery [27]. Robot route can be improved by consolidating ESR in the framework [7], [8]. ESR can be embraced in a home-observing condition, be it for helping older individuals living alone in their very own home [9, 10] or for a keen home [11]. ESR, alongside picture and video examination, discover applications in observation [28, 29]. ESR can likewise be custom-made for acknowledgment of creature and winged creature species by their particular sounds [12, 13].

Environmental Sound Recognition System Design starts with database selection. After, select the right environmental sound database firstly, feature extraction techniques are applied the environmental sounds. Then, these data which are obtained from applying feature extraction techniques use for learning. Before the learning phase we divide the data train and test. For learning, train data are used and then, learning techniques are applied these train data. System learns from train data and system aims to recognize environmental sounds with test data. Each try, system learn better to recognize environmental sounds.

2.4 Environmental Sound Recognition Application with Python

In the research, http://wiki.cnbc.cmu.edu/Sound_Databases is used as a database. Various environmental sounds have been selected from this database. Anaconda plugin Spyder, a program developed for Python, is used to create the sound recognition system to work. The reason why Python language is preferred is that its libraries are suitable for both feature extraction and classification techniques. Another reason is that it is efficient in such applications. The majority of the sounds are used for learning data and the rest are used as test data (learning data 2/3, test data 1/3). Audio files with .wav extension selected as learning data and loaded into the sound recognition system using the Python library which name is librosa. Different feature extractions (Environmental Sound Features Extraction Techniques section) are applied to the audio signals which are obtained from these environmental sounds. The features are obtained from each feature extraction techniques and combined together to be used in learning techniques. It is aimed to learn the system by applying various classification techniques (Environmental Sound Classification Techniques section). After the system learning, the sounds that are separated as test data are tested. For each combination, accuracy percentages are obtained from these test results.

CHAPTER 3

ENVIRONMENTAL SOUND FEATURE EXTRACTION TECHNIQUES

3.1 Sound Feature Extractions Overview

Each sound signal contains numerous features. Nonetheless, we should separate the attributes that are important to the issue we are attempting to tackle. The way which extract sound attributes to utilize them for investigation is called feature extraction [33]. Feature extractions is the principle part of the environmental sound recognition system. The work of this is to extract those features from the input signal that help the system in identifying the sound. Feature extraction compresses the magnitude of the input signal without causing any harm to the power of sound signal.

Each sound is fundamentally comprised of two significant things, the sample rate, and the sample data and with the help of the sample rate and the sample data, one can perform several transformations on it to extract valuable features out of it.

3.2 Sound Feature Extractions Techniques and Application

There are many feature extraction techniques. In this thesis, few of the features are used like a zero-crossing rate, spectral centroid, mel-frequency cepstral coefficients, chroma frequencies, continuous wavelet transform.

Firstly, the sound is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, a few libraries are used to extract the signal features. Librosa and signal libraries are used for feature extraction properties. Figure 3.1 shows the sample code part.

```
import librosa
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

y0, sr0 = librosa.load('sound/train/rolling/rollingbigmarbleonmetal-1.wav')

zcr0 = librosa.feature.zero_crossing_rate(y=y0)
```

Figure 3.1: General feature extraction sample code part

3.2.1 Zero Crossing Rate

The Zero Crossing Rate is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to negative or back. This feature has been used heavily in both speech recognition and music information retrieval [33].

Zero crossing rate of any signal frame is the rate at which a signal changes its sign during the frame. It denotes the number of times the signal changes value, from positive to negative and vice versa, divided by the total length of the frame [37].

Zero crossings are a basic property of an audio signal that is often employed in audio classification. Zero crossings allow for a rough estimation of dominant frequency and the spectral centroid (SC).

In this part, the music is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, librosa which is the python library is used to extract the signal features. Then, feature extraction property Zero Crossing Rate is used. Figure 3.2 shows the Zero Crossing Rate sample code part.

```
import librosa

y0, sr0 = librosa.load('sound/train/air/airOutofBaloon1.wav')
zcr0 = librosa.feature.zero_crossing_rate(y=y0)
```

Figure 3.2: Zero Crossing Rate feature extraction sample code part

3.2.2 Spectral Centroid

The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It indicates where the "center of mass" of the spectrum is located. Perceptually, it has a robust connection with the impression of "brightness" of a sound [34].

It is calculated as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as the weights: [35]

$$\mathrm{Centroid} = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

Figure 3.3: Centroid Formula

where x(n) represents the weighted frequency value, or magnitude, of bin number n, and f(n) represents the center frequency of that bin.

The spectral centroid and the spectral spread are two simple measures of spectral position and shape. The spectral centroid is the center of ‘gravity’ of the spectrum.

In this part, the music is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, librosa which is the python library is used to extract the signal features. Then, feature extraction property Spectral Centroid is used. Figure 3.4 shows the Spectral Centroid sample code part.

```
import librosa

y0, sr0 = librosa.load('sound/train/air/airOutofBaloon1.wav')
spec_cent0 = librosa.feature.spectral_centroid(y=y0, sr=sr0)
```

Figure 3.4: Spectral Centroid feature extraction sample code part

3.2.3 Mel-Frequency Cepstral Coefficients

In audio processing, Mel-Frequency Cepstral Coefficients (MFCC) is a representation of the short-term power spectrum of a sound based on a linear cosine transformation of a log power spectrum on a non-linear frequency scale.

The MFCC feature is employed to represent audio signals. The idea of the MFCC feature is motivated by perceptual or computational considerations. As the feature captures some of the crucial properties used in human hearing, it is ideal for general audio discrimination. The MFCC feature has been successfully applied to speech recognition

In this part, the music is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, librosa which is the python library is used to extract the signal features. Then, feature extraction property Mel-Frequency Cepstral Coefficients is used. Figure 3.5 shows the Mel-Frequency Cepstral Coefficients sample code part.

```
import librosa

y0, sr0 = librosa.load('sound/train/air/airOutofBaloon1.wav')
mfcc_r10 = librosa.feature.mfcc(y=y0, sr=sr0, n_mfcc=40)
```

Figure 3.5: Mel-Frequency Cepstral Coefficients feature extraction sample code part

3.2.4 Chroma Frequencies

Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave [33].

In this part, the music is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, librosa which is the python library is used to extract the signal features. Then, feature extraction property Chroma Frequencies is used. Figure 3.6 shows the Chroma Frequencies sample code part.

```
import librosa

y0, sr0 = librosa.load('sound/train/air/airOutofBaloon1.wav')
c0 = librosa.feature.chroma_stft(y=y0, sr=sr0,hop_length=hop_length)
```

Figure 3.6: Chroma Frequencies feature extraction sample code part

3.2.5 Continuous Wavelet Transform

The conventional method of producing a time-frequency map using the short time Fourier transform (STFT) limits time-frequency resolution by a predefined window length. In contrast, the Continuous Wavelet Transform (CWT) method does not require preselecting a window length and does not have a fixed time-frequency resolution over the time-frequency space. CWT uses dilation and translation of a wavelet to produce a time-scale map. A single scale encompasses a frequency band and is inversely proportional to the time support of the dilated wavelet [34].

In this part, the music is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, librosa which is the python library is used to extract the signal features. Then, feature extraction property Continuous Wavelet Transform is used. Figure 3.7 shows the Continuous Wavelet Transform sample code part.

```
import librosa
import scipy.signal as signal

sig0, s0 = librosa.load('sound/train/rolling/rollingbigmarbleonmetal-1.wav')
cwt0 = signal.cwt(sig0, wavelet, widths)
```

Figure 3.7: Continuous Wavelet Transform feature extraction sample code part

CHAPTER 4

ENVIRONMENTAL SOUND CLASSIFICATION TECHNIQUES

4.1 Classification Techniques Overview

The concept of classification is simply to distribute the data between the various classes defined on a data set. Classification algorithms learn this distribution from the given training set and then try to classify them correctly when the test data is not specified.

Machine Learning (ML) is a vast interdisciplinary field which builds upon concepts from computer science, statistics, cognitive science, engineering, optimization theory and many other disciplines of mathematics and science [38]. There are numerous applications for machine learning but data mining is most significant among all [39]. Machine learning can mainly classify into two broad categories include supervised machine learning and unsupervised machine learning.

Unsupervised machine learning used to draw conclusions from datasets consisting of input data without labeled responses [40] or we can say in unsupervised learning desired output is not given. Supervised machine learning techniques attempt to find out the relationship between input attributes (independent variables) and a target attribute (dependent variable) [41]. Supervised techniques can further classify into two main categories; classification and regression. In regression output variable takes continuous values while in classification output variable takes class labels [41]. Classification is a data mining (machine learning) approach that used to forecast group membership for data instances [43]. Although there are variety of available techniques for machine learning but classification is most widely used technique [44]. Classification is an admired task in machine learning especially in future plan and knowledge discovery classification is categorized as one of the supreme studied problems by researchers of the machine learning and data mining fields [45].

4.2 Classification Techniques and Application

There are many feature extraction techniques. In this thesis, we use few of the classification techniques like a K nearest neighborhood (KNN), Support Vector Machine (SVM), Decision Tree Classifier.

Firstly, the sound is loaded as an audio .wav file with python code in Spyder which is environment for python. Then, a few libraries are used to extract the signal features. Librosa, signal libraries are used for feature extraction properties. After that, these features are used for learning and classification techniques. System learn from these data and try to recognize and classify test data which environmental sound type are they.

4.2.1 K nearest neighborhood

According to this algorithm used in the classification, the characteristics removed during the classification are looking at the similarity of the new individual from previous individuals to k.

In this part, system learn from features which are obtained from feature extraction techniques. After that K nearest neighborhood is used to recognize and classify test data. Figure 4.1 shows the K nearest neighborhood sample code part.

```
#train the model
neighEnv = KNeighborsClassifier(n_neighbors=3)

neighEnv.fit(X, y)
```

Figure 4.1: K nearest neighborhood sample code part

4.2.2 Support Vector Machine

It is one of the most effective and simple methods used in classification. For classification, it is possible to separate two groups by drawing a border between two groups in a plane. The place where this limit will be drawn should be the most distant from the members of both groups. Here SVM determines how this limit is drawn.

In this part, system learn from features which are obtained from feature extraction techniques. After that Support Vector Machine is used to recognize and classify test data. Figure 4.2 shows the Support Vector Machine sample code part.

```
#train the model
svc = svm.SVC(kernel='linear', gamma=2)
svc.fit(X, y),
```

Figure 4.2: Support Vector Machine sample code part

4.2.3 Decision Tree Classifier

In the decision tree learning, a tree structure is formed and the class labels on the leaf level of the tree and the handles that go to these leaves and with the arms coming from the beginning are expressed.

In this part, system learn from features which are obtained from feature extraction techniques. After that Decision Tree Classifier is used to recognize and classify test data. Figure 4.3 shows the Decision Tree Classifier sample code part.

```
#DecisionTreeClassifier
dtc = DecisionTreeClassifier(random_state=0)
dtc.fit(X, y)
```

Figure 4.3: Decision Tree Classifier sample code part

CHAPTER 5

PERFORMANCE COMPARISON OF SELECTED TECHNIQUES

Comparisons are made from the results and the best combinations are determined. These are MFCC-SVM 95%, CWT-SVM 95%, CWT-DecisionTreeClassifier 95% as shown in Table 5.1. Mel-Frequency Cepstral Coefficients (MFCC) and Continuous Wavelet Transform (CWT) are the most effective feature extraction techniques. As a classification technique, the most effective techniques are Support Vector Machine (SVM) and Decision Tree Classification.

Table 5.1: Performance comparison result

Combinations	KNN	SVM	Decision Tree Classifier
MFCC	%75	%95	%75
CWT	%50	%95	%95
Chroma	%50	%30	%50
Zero Crossing Rate	%40	%60	%70
Spectral Centroid	%75	%65	%80

CHAPTER 6

CONCLUSION

The aim of this paper is to make environmental sound recognition in the most efficient way. It is aimed to find the combination of the most successful feature extraction and classification for the most efficient environmental sound recognition. In this, various learning techniques are applied to the features which are obtained after time-frequency measured feature extraction techniques and for using different feature extraction techniques and learning techniques together various combinations are obtained to find best sound recognizing system. Comparisons are obtained from the test data result and in this way, the best combinations are determined. According to the comparison result MFCC-SVM 95%, CWT-SVM 95%, CWT-DecisionTreeClassifier 95% rate is obtained. Further elaboration on these techniques can be achieved by obtaining better classification rates.

The field of artificial intelligence, which is constantly evolving, enables the testing of new types of feature extraction and classification techniques, and paves the way for environmental sound recognition studies.

REFERENCES

[1] Environmentally sound. Retrieved May 26, 2019, from http://www.businessdictionary.com/definition/environmentally-sound.html

[2] What is the environmentally sound. Retrieved June 2, 2019, from https://thelawdictionary.org/environmentally-sound/

[3] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, T. Yamada, "Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition," Proceedings of 2nd ICLRE, pp. 965-968, 2000

[4] Y. Arslan, A. Tanış and H. Canbolat, "A Relational Database Model and Tools for Environmental Sound Recognition", Advances in Science, Technology and Engineering Systems Journal (ASTESJ), vol. 2, no. 6, pp. 145-150, 2017

[5] T. Ahmed, M. Uppal and A. Muhammad, "Improving Efficiency and Realibility of Gunshot Detection Systems", IEEE, ICASSP 2013

[6] P. Thumwarin, T. Matsuura and K. Yakoompai, "Audio forensics from gunshot for firearm identification", Proc. IEEE 4th Joint International Conference on Information and Communication Technology Electronic and Electrical Engineering Tailand, pp. 1-4, 2014

[7] S. Chu, S. Narayanan, C.-C. J. Kuo, and M. J. Mataric, "Where am I? Scene recognition for mobile robots using audio features," in Multimedia and Expo, 2006 IEEE International Conference on. IEEE, pp. 885–888, 2006

[8] N. Yamakawa, T. Takahashi, T. Kitahara, T. Ogata, and H. G. Okuno, "Environmental sound recognition for robot audition using Matching- Pursuit," in Modern Approaches in Applied Intelligence. Springer, pp. 1–10, 2011

[9] J. Chen, A. H. Kam, J. Zhang, N. Liu, and L. Shue, "Bathroom activity monitoring based on sound," in Pervasive Computing. Springer, pp. 47–61, 2005

[10] M. Vacher, F. Portet, A. Fleury, and N. Noury, “Challenges in the processing of audio channels for ambient assisted living,” in e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on. IEEE, pp. 330–337, 2010

[11] J.-C. Wang, H.-P. Lee, J.-F. Wang, and C.-B. Lin, “Robust environmental sound recognition for home automation,” IEEE Trans. Autom. Sci. Eng., vol. 5, no. 1, pp. 25–31, 2008

[12] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt “Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring,” Pattern Recognition Letters, vol. 31, no. 12, pp. 1524–1534, 2010

[13] F. Weninger and B. Schuller, “Audio recognition in the wild: static and dynamic classification on a real-world database of animal vocalizations,” in acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on. IEEE, pp. 337–340, 2011

[14] P. Foggia, A. Saggese, N. Strisciuglio, M. Vento, and N. Petkov. Carcrashes detection by audio analysis in crowded roads. In Advanced Video and Signal Based Surveillance e (AVSS), 2015 12th IEEE International Conference on, pages 1-6, Aug 2015

[15] J. Rouas, J. Louradour, and S. Ambellouis, "Audio Events Detection in Public Transport Vehicle," Proc. of the 9th International IEEE Conference on Intelligent Transportation Systems, 2006

[16] R. Radhakrishnan and A. Divakaran, “Systematic acquisition of audio classes for elevator surveillance,” in Image and Video Communications and Processing 2005, vol. 5685 of Proceedings of SPIE, pp. 64–71, March 2005

[17] P. K. Atrey, N. C. Maddage, and M. S. Kankanhalli, “Audio based event detection for multimedia surveillance,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’06), vol. 5, pp. 813–816, Toulouse, France, May 2006

[18] K. Lupatka, P. Zwan, A. Czyzewski, “Dangerous sound event recognition using support vector machine classifiers”, Advances in Multimedia and Network Information System Technologies Volume 80 of the series Advances in Intelligent and Soft Computing pp 49-57, 2010

[19] A. Okamoto, K. Hayashida, M. Nakayama, T. Nishiura, “A detection of dangerousness based on variable-state hidden Markov models”, Signal Processing in Acoustics, 2013

[20] S. Chu, S. Narayanan, and C.-C. J. Kuo, “Environmental sound recognition with time– frequency audio features,” IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 6, pp. 1142–1158, 2009

[21] M. Lojka, M. Pleva, E. Kiktova, J. Juhar, A. Cizmar. "Efficient acoustic detector of gunshots and glass breaking." Multimedia Tools and Applications, 75(17):10441–10469, 2016

[22] Yüksel Arslan and Hüseyin Canbolat, A sound database development for environmental sound recognition, 2017 25th Signal Processing and Communications Applications Conference (SIU) Year: 2017

[23] Brian Gygi and Valeriy Shafiro, Development of the Database for Environmental Sound Research and Application (DESRA): Design, Functionality and Retrieval Considerations, EURASIP Journal on Audio, Speech, and Music Processing Volume 2010.

[24] Brian Gygi and Gary R. Kidd and Charles S. Watson, Similarity and categorization of environmental sounds, Perception & Psychophysics 2007, 69 (6), 839-855.

[25] S. Chachada and C.-C. J. Kuo. Environmental sound recognition: A survey. APSIPA Transactions on Signal and Information Processing, 3: e14, 2014.

[26] T. Virtanen and M. Helen, “Probabilistic model-based similarity measures for audio query-by-example,” in Applications of Signal Processing to Audio and Acoustics, IEEE Workshop on. IEEE, 2007, pp. 82– 85, 2007

[27] S. Duan, J. Zhang, P. Roe, and M. Towsey, “A survey of tagging techniques for music, speech and environmental sound,” Artificial Intelligence Review, pp. 1–25, 2012

[28] M. Cristani, M. Bicego, and V. Murino, “Audio-visual event recognition in surveillance video sequences,” IEEE Trans. Multimedia, vol. 9, no. 2, pp. 257–267, 2007

[29] R. Sitte and L. Willets, “Non-speech environmental sound identification for surveillance using self-organizing-maps,” in Proceedings of the Fourth conference on IASTED International Conference: Signal Processing, Pattern Recognition, and Applications, ser. SPPR’07. Anaheim, CA, USA: ACTA Press, pp. 281–286, 2007

[30] O. Gencoglu, T. Virtanen, H. Huttunen. (2014). "Recognition of Acoustic Events Using Deep Neural Networks," in European Signal Processing Conference (EUSIPCO 2014).

[31] Cakir, E. Sound Multilabel Sound Event Classification with Neural Networks. Unpublished master’s thesis, Tampere University of Technology, 2014

[32] Gebremedhin, T. A. An Environmental Audio–Based Context Recognition System Using Smartphones. Unpublished master’s thesis, University of Twente, 2014.

[33] https://towardsdatascience.com/music-genre-classification-with-python

[34] Satish Sinha, Partha S. Routh, Phil D. Anno, and John P. Castagna. Spectral decomposition of seismic data with continuous-wavelet transform. Geophysics Volume 70, Issue 6.

[35] A Large Set of Audio Features for Sound Description - technical report published by IRCAM in 2003. Section 6.1.1 describes the spectral centroid.

[36] Wavelet toolbox. Retrieved June 6, 2019, from https://www.mathworks.com/help/wavelet/gs/continuous-wavelet-transform-and-scale-based-analysis

[37] Nilesh Kulkarni, Vinayak Bairagi. Chapter Four - Use of Complexity Features for Diagnosis of Alzheimer Disease. 20 April 2018

[38] Ghahramani Z. "Unsupervised learning," in Advanced lectures on machine learning, ed: Springer, 2004; pp. 72-112.https://doi.org/10.1007/978-3-540-28650-9_5

[39] Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. ed, 2007

[40] Zhang D, Nunamaker JF. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information Systems Frontiers 2003; 5: 207-218.https://doi.org/10.1023/A:1022609809036

[41] Maimon O, Rokach L. Introduction to supervised methods, in Data Mining and Knowledge Discovery Handbook, ed: Springer, 2005 pp. 149-164.

[42] Ng A. "CS229 Lecture notes."

[43] Kesavaraj G, Sukumaran S. A study on classification techniques in data mining. in Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on, 2013; pp. 1-7.

[44] Singh M, Sharma S, Kaur A. Performance Analysis of Decision Trees. International Journal of Computer Applications 2013; 71.

[45] Baradwaj BK, Pal S. Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417, 2012

CURRICULUM VITAE

PERSONAL INFORMATION

Name Surname: Yaseminhan ARPACI

Date of Birth: 25/10/1990

E-mail: arpacyasemin@gmail.com

EDUCATION

High School: Eskisehir Fatih Anadolu Lisesi

Bachelor: Eskisehir Osmangazi University Electrical and Electronical Engineering

WORK EXPERIENCE

Avionics Software Development Engineer: ASELSAN - ENGINSOFT TURKEY – (2016 – ongoing)

TOPICS OF INTEREST

- Programming

- Machine Learning