Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 37 (2014) 40 – 47

Procedia Computer Science

International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014)

Feature Selection for Place Classification through Environmental Sounds

Juan Rubén Delgado-Contreras[a,*], Juan Pablo García-Vázquez[a,c], Ramon F. Brena[a], Carlos E. Galván-Tejada[a,d], Jorge I. Galván-Tejada[b,d]

[a] *Tecnologico de Monterrey (ITESM), Autonomous Agents for Ambient Intelligence Research Chair, Ave. Eugenio Garza Sada 2501 Sur Col. Tecnológico C.P. 64849, Monterrey, N. L., México*

[b] *Tecnologico de Monterrey (ITESM), Bioinformatics Research Chair, Ave. Eugenio Garza Sada 2501 Sur Col. Tecnológico C.P. 64849, Monterrey, N. L., México*

[c] *School of Engineering, Autonomous University of Baja California, Mydci, Mexicali, Mexico*

[d] *Univeridad Autónoma de Zacatecas,Programa de Ingeniería de Software, Ciudad Universitaria Siglo XXI, Edificio de Ingeniería de Software e Ingeniería en Computación, C.P. 98160, Zacatecas, Zac., México.*

Abstract

In this work, an environmental audio classification scheme is proposed using a Chi squared filter as a feature selection strategy. Using feature selection (FS), the original 62 features characteristic vector can be optimized, and it can be used for environmental sound classification. These features are obtained using statistical analysis and frequency domain analysis. As a result, we obtain a reduced feature vector composed of 15 features: 11 statistical and 4 of the frequency domain. Using this reduced vector, a 10 class classification was done, using Support Vector machines (SVM) as classification method, the accuracy is higher than 90%.

Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.

Keywords: Environmental Sound; Feature Selection; Feature Extraction; Support Vector Machine; Chi-Square Filter.

1. Introduction

Humans are surrounded by several sounds that come from a variety of sources, such as living beings, objects or nature phenomena. These sounds contain information that can allow us to recognize the kind of activity is doing some individual, or enable us to be aware about the context around a person (e.g. Objects, places and events). However, recognition of an environmentally sound brings several challenges in comparison with existent recognition of music and speech techniques, because it must be considered that an environmentally sound (ES) is not structured by nature, typically contain noise and flat spectrum features [1].

As has been seen in some other works, ES classification is a pattern recognition problem, this problem commonly consists in feature extraction and classification based on these extracted features [2,1].

* Corresponding author. Tel.: +52 (81) 8358-2000
E-mail address: jr.delgado.phd.mty@itesm.mx

1877-0509
Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.
doi:10.1016/j.procs.2014.08.010

In this paper, we propose to extend an environmental sound classification scheme proposed in our previous work[3], by adding to the scheme a feature selection process based on *Chi-squared Filter* and *Support Vector Machine* (SVM) as a classifier. This strategy allows us to reduce the number of features used for a correct classification, at the same time, thus reduce the quantity of information needed, and develop an environmentally sound classification model that can be deployed in mobile devices with reduced computational capabilities (e.g. Smartphones).

There are two main contributions in this work: (1) a scheme of complex environmental sound classification and (2) a set of audio features that enable us to classify a complex environmental sound (CES). CES are sounds composed of more than one sound source; for instance, the environmental sound of a restaurant contains several sound sources, as human voices, music, lights, among others for this reason is considered a complex sound.

This paper is organized as follows. In section 2, we present schemes to classify environmental sounds. The data set of CES used in this work is described in section 3. In section 4 we described the proposed environmental audio scheme of classification. The results of experiments are presented in section 5, and finally, our conclusions and future work are presented in section 6.

2. Related Work

Several projects have been proposed schemes to recognize an ES, these could be divided in three categories:

- *Schemes for Classifying Simple Environmental Sounds*. In those projects use as an input a simple environmental sound (e.g. rain, engine). For instance, *Okuyucu* et al.[1] present an automatic recognition framework for environmental sounds by using eleven (11) audio features (MPEG-7 family, Zero Crossing Rate (ZCR), Mel Frequency Cepstral Coefficient(MFCC), and combination). Thirteen (13) environmental audio categories (e.g. car horn, explosion, wind, rain, etc.) were classified using hidden Markov Model (HMM) and Support Vector Machine (SVM). The Authors claim that using ASFCS-H with SVM yield best performance with average F-measure value of 80.6% among other stand-alone and joint features. In the same direction, *Zhang* et al.[2] proposed an algorithm of audio classification based on Support Vector Machine (SVM) and Universal Background Mixture Model (UBM) using MFCC as audio features. To evaluate the performance of the algorithm using four audio types: speech, music, speech over music and environmental sound. Regarding environmental audio the authors claim an 85.36% of accuracy.
- *Schemes for Classifiying Simple Environmental Sound with Tags* . These works use simple environmental sound and additionally sound descriptions (e.g. tags) to identify different contexts. For instance, *Rossi* et al.[4] proposed an architecture for sound context recognition, which uses web-collected audio and its crowd-sourced textual descriptions. This is based on Mahalanobis distance and Gaussian Mixture Model (GMM) as a classifier. The authors claim that their architecture can recognize 23 sound context categories in a real setting with a 51% of accuracy.
- *Schemes for Classifying Complex Environmental Sound*. These projects use as an input complex combinations of sounds (e.g. restaurant, casino) and they attempt to classify the whole given environment. For instance, *Su* et al.[5] propose an environmental sound and auditory scene recognition scheme. They use local discriminant bases (LDD) technique for feature extraction process and hidden Markov Model (HMM) as a classifier. The scheme was evaluated with audio data from internet, TV and movies. A total of 21 sound events was classified, which include SES (e.g. engine, car-braking, siren, etc.) and CES (e.g. restaurant). The authors claim an average recognition accuracy of 81% for the test set. However, whether in the scene presents several environmental audio the average of the accuracy decrease to 28.6%. Another similar work was presented in *Eronen* et al.[6], they developed a system to evaluate the recognition accuracy of several audio features (e.g. ZCR, MFCC, spectral roll-off, spectral, flux) and use as classifiers K Nearest Neighbor (KNN) and Hidden Markov Model (HMM). A total of 24 clases were tested and achieve an average recognition accuracy of 58%

The mentioned environmental sound classification approaches have in common two phases: *feature extraction* and *classification*. In this paper, we propose a classification scheme for complex environmental sounds that include an additional phase: *feature selection*. This phase enables us to get a reduced set of features so that the environmental sound classification model requires less computational resources.