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

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.

Table 1. Enviromental Sound Categories and Classes

Categories	Sound Class	Total of Sounds
Social places	restaurant, casino, and playgroud	15
Street	street traffic, street with ambulance and train	15
Countryside	nature at day time and nature at night time	10
Water	ocean and river	10

3. Data Collection

The data set used in this paper was collected from an open web collaborative online database called *freesound* [1]. We select this audio database because has been widely used as a dataset in several research works of audio recognition [7]. This is due to the fact that it consist of more than 160,000 audio samples, which are heterogeneous, available in large quantities and all sounds are moderated by a group of users that check that description are correct and that the sounds are not illegal.

Our data set is conformed by Complex Environmental Sounds (CES) that belong to different places The audio files are in *wav* format with a sample rate of 44.1 kHz. A total of 50 sounds was considered for the experimentation.

Environmental sounds were associated with a place with tags; each of these sounds will be called *class*. A total of 10 classes was conformed, each one consisting of 5 sounds of a place that have a different time duration in a range of 10 seconds to 25 minutes. We propose categorizing these classes in categories considering sounds as is shown in table 1. The audio data set is available in a sound repository in our web page [2].

4. Environmental Complex Sound Classification Scheme

In this section, we present our proposed environmental complex sound classification scheme. This is composed of three phases as is shown in figure 1. These phases are described in the following sections.

4.1. Fingerprint Extraction

This phase consists of two tasks: *feature extraction* and *data normalization*. The feature extraction (FE) process was carried out using the programming environment R, a tool for statistical analysis and graphics [8]. FE consist in to extract 62 features from the first 10 seconds of CES selected in data collection, these are shown in table 2. These were categorized in temporal, frequency and statistical features. To extract the audio signal features, we use our approach of feature extraction presented in our previous work [3], which consists of extract frequency features through applying a Fast Fourier Transform (FFT) to the original signal; and to extract temporal and statistical features from the original signal, analyzing it without any preprocessing.

These features were used to conform an *audio fingerprint*, it refers to an small feature vector that contain a significative information that enable us to represent the audio signal behavior.

Once we have the audio fingerprint, we apply *Z-normalization* to obtain a feature vector with zero mean and a *percentile rank* to keep data in values of 0 to 1.

4.2. Fingerprint Reduction

Since we are considering the development of an environmental sound classification scheme that can be deployed in mobile devices (e.g. Smart phone), in this paper, we propose to apply a feature selection process, which consists in

[1] http://www.freesound.org
[2] http://aaami.mty.itesm.mx

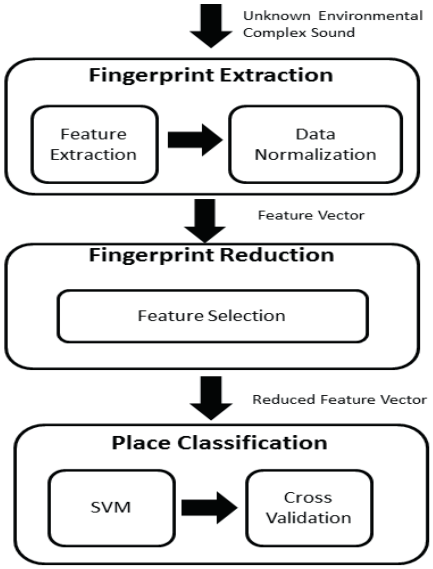

Fig. 1. Block Diagram of Environmental Complex Sound Classification Scheme

Table 2. Features Extracted

	Temporal Features	Frequency Features	Statistical Features
Short-Time Average Zero-Crossing Rate	*		
Logarithmic Short-Term Energy	*		
Squared Short-Term Energy	*		
Absolute Short-Term Energy	*		
Spectral Flux		*	
Spectral Roll Off		*	
Spectral Centroid		*	
Spectral Flatness		*	
Shannon Entropy		*	
Slope		*	
Maximum		*	*
Minimum		*	*
Mean		*	*
Median		*	*
Standard Deviation		*	*
Variance		*	*
Coefficient of Variation		*	*
Inverse Coefficient of variation		*	*
Interquartile Range		*	*
Trimmed Mean		*	*
Skewness		*	*
Kurtosis		*	*
percentile 1, 5, 10, 25, 50, 75, 90, 95, 99		*	*
10 higher frequencies.		*	

Table 3. Feature Selection

	Frequency Features	Statistical Features
Spectral Roll off	*	
Slope	*	
Minimum	*	*
Median		*
Coefficient of Variation		*
Inverse Coefficient of Variation		*
Trimmed Mean		*
Skewness		*
Kurtosis	*	*
Percentile 1, 75, 95 and 99		*

selecting a subset of the most significative features from original audio feature vector that enable us to modeling the audio signal behavior.

Several techniques have been used for feature selection, for instance, filter, wrapper and embedded methods[9]. We propose using a *filter method*, since it has the advantage that is independent of the classifier and is better computational complexity than wrapper methods and some disadvantages ignores the interaction with the classifier. While, *wrapper methods* have the advantage of interacting with the classifier, but is computationally extensive and have a high risk of overfitting. Some advantages of *embedded methods* are that interacts with the classifier, better computational complexity than wrapper method and have as disadvantage that the feature selection depend of the classifier.

Considering that the filter methods enable us to obtain a set of features independent of the classifier and with a low cost complexity, we propose to use a common filter method called *Chi-Squared Filter*[10]. This filter method evaluates the data as a function of a calculated weight; features with a low weight are removed based on the chi squared statistics[11]. After that, this subset of features is used as input of the classification method.

4.3. Place Classification

To carry out a place classification we proposed using a Support Vector Machine (SVM) binary classifier[12,13]. Given a set of points in a space, the data with which we train the system would be a set of tagged vectors, in which the tag is the class to which a vector belongs. An SVM classifier seeks a separating hyperplane that divides the space into two regions. The separating hyperplane searched by the algorithm is such that it maximizes the distance between the two different classes of the problem. In this paper is considered using a SVM with a method called one-versus-one[14], which enables us to calculate how many binary classifiers we need for our experiments that can be calculated by (1); where k is the number of classes.

$$NumberOfBinaryClassifier = \frac{k(k-1)}{2} \tag{1}$$

In our experiments, the SVM creates 6 binary classifiers to classify our categorized sounds. Once that the binary classifiers were created, a class is assigned through the largest number of votes.

5. Results

In this section, we evaluate the classification scheme of complex environmental sound (CES). First discussion is about the feature selection process results and finally, the quality of the inference made by the SVM classifier.

Table 4. Environmental sound classification

	SocialPlaces	SoundOfcountryside	Street	Water
SocialPlaces	13	1	3	0
SoundOfcountryside	0	8	0	0
Street	2	0	12	1
Water	0	1	0	9

5.1. Feature Selection Process Results

After the feature extraction process, a vector that is composed of 62 audio features for each sound was obtanied. These audio features were used to conform an *audio fingerprint*. In this research work, a total of 50 audio fingerprints were obtained.

All audio fingerprints were used to construct a feature vector (50x62). The vector was used as input for Chi-Squared feature selection method. The Classification And REgression Training package was used to perform the analysis (CARET, version 6.0-24 for R language[11]). Of a total of 62 audio features from the original vector, Chi-Square provides us 15 features with a high weight, which provide the information needed to model the original vector, these audio features are shown in table 3. Most of these audio features belong to statistical features, whether we compare table 2 and table 3, we can see that feature selection process maintains about of a 52 % (11/21) of original statistical features and about of a 10 % (4/37) of the original frequency features. However, no time related features survive the process.

5.2. Classification Performace

Support Vector Machine (SVM) with a radial basis kernel was used as classifier method, this analysis was performed using the *e1071* R package[15]. For the SVM clasification, each row of table 3 was used as an outcome variable with a 10 fold cross validation strategy for train and test subsets of data.

In table 5 we present the confusion matrix of each class that conform the 4 categories. Our results indicate that SVM is able to classify correctly 45 of the 50 CES, this represents a 90 % of accuracy and 10 % of error rate. Regarding the categories we identify that SVM was able to classify correctly 42 of 50 classes using the set of features obtain by Chi-Squared, this represent a 16 % of classification error rate and a 84 % of correct classification. These results are shown in table 4.

We identify that the most missclasified CES are presented in the classification of categories like social places and street categories. These categories were classified with a 86.66% and 90% of accuracy respectively. It means that in Social places category were classified correctly 13 of 15 sounds, and in Water category 9 of 10 sounds. We can infer that the classes in both categories share some similar simple sound sources like music and speech. We identify that adding the feature selection process decrease the accurracy of scheme in 1.42 % for class classification and 7.42 % dividing those classes in categories as is shown in table 6. Although accuracy decrease, we are able to obtain a set of 15 features that represent about 24 % of the information with an accuracy 84 %.

6. Conclusions and Future Work

From our results, we conclude that the feature selection phase allows us to reduce the feature vector from 62 to 15 features. However, the accuracy of the classification scheme decreased a 1.42 % for class classification and 7.42 % dividing those classes in categories in comparison with the scheme that uses only feature extraction.This result suggests the possibility to deploy the classification of the CES scheme in a mobile device, since processing less information required less computational cost.

Regarding the set of features, we identify that temporal features do not survive the feature selection process. Therefore, the reduced feature vector is composed of statistical (11) and frequency (4) features that can be used to classify CES with an accuracy of 84 % and an error rate of 16 %.

Table 5. Class Classification

Nat1=Nature Day time, Nat2=Nature Night time, Rest=Restaurant, Ambulance=Street with ambulance, Traffic=Street with traffic,

	Casino	Nat1	Nat2	Ocean	Playground	Rest	river	Ambulance	Traffic	Train
Casino	5	0	0	0	0	0	0	1	0	0
Nature_DT	0	5	0	0	0	0	0	0	0	0
Nature_NT	0	0	4	0	0	0	0	0	0	0
Ocean	0	0	0	5	0	0	0	0	0	0
Playground	0	0	0	0	4	0	1	0	0	0
restaurant	0	0	0	0	0	4	0	0	0	0
river	0	0	1	0	0	0	4	0	0	0
Street_Amb	0	0	0	0	0	0	0	4	0	0
street-traffic	0	0	0	0	1	1	0	0	5	0
Train	0	0	0	0	0	0	0	0	0	5

Table 6. Comparison Scheme

	accuracy class	error rate class	accuracy categories	error rate categories
Feature Extraction Scheme	91.42 %	8.58 %	-	-
Feature Selection Scheme	90 %	10 %	84 %	16 %

Finally, our future work comprises two aspects: (i) The implementation of our model in a mobile device to operate in real environments instead of the online sound base; and (ii) to evaluate the performance of the scheme in real environments.

References

1. Okuyucu, C., Sert, M., YAZICI, A.. Audio feature and classifier analysis for efficient recognition of environmental sounds. In: *Multimedia (ISM), 2013 IEEE International Symposium on*. 2013, p. 125–132. doi:10.1109/ISM.2013.29.
2. Zhang, R., Li, B., Peng, T.. Audio classification based on svm-ubm. In: *Signal Processing, 2008. ICSP 2008. 9th International Conference on*. 2008, p. 1586–1589. doi:10.1109/ICOSP.2008.4697438.
3. Ruben Delgado-Contreras, J., Garcia-Vazquez, J.P., Brena, R.F.. Classification of environmental audio signals using statistical time and frequency features. In: *Electronics, Communications and Computers (CONIELECOMP), 2014 International Conference on*. 2014, p. 212–216. doi:10.1109/CONIELECOMP.2014.6808593.
4. Rossi, M., Troster, G., Amft, O.. Recognizing daily life context using web-collected audio data. In: *Wearable Computers (ISWC), 2012 16th International Symposium on*. 2012, p. 25–28. doi:10.1109/ISWC.2012.12.
5. Su, F., Yang, L., Lu, T., Wang, G.. Environmental sound classification for scene recognition using local discriminant bases and hmm. In: *Proceedings of the 19th ACM International Conference on Multimedia*; MM '11. New York, NY, USA: ACM. ISBN 978-1-4503-0616-4; 2011, p. 1389–1392. URL: `http://doi.acm.org/10.1145/2072298.2072022`. doi:10.1145/2072298.2072022.
6. Eronen, A., Peltonen, V., Tuomi, J., Klapuri, A., Fagerlund, S., Sorsa, T., et al. Audio-based context recognition. *Audio, Speech, and Language Processing, IEEE Transactions on* 2006;**14**(1):321–329. doi:10.1109/TSA.2005.854103.
7. Font, F., Roma, G., Serra, X.. Freesound technical demo. In: *Proceedings of the 21st ACM International Conference on Multimedia*; MM '13. New York, NY, USA: ACM. ISBN 978-1-4503-2404-5; 2013, p. 411–412. URL: `http://doi.acm.org/10.1145/2502081.2502245`. doi:10.1145/2502081.2502245.
8. Dessau, R., Pipper, C.B.. ' r project for statistical computing. *Ugeskrift for laeger* 2008;**170**(5):328.
9. Saeys, Y., Inza, I., Larrañaga, P.. A review of feature selection techniques in bioinformatics. *bioinformatics* 2007;**23**(19):2507–2517.
10. Liu, H., Setiono, R.. Chi2: Feature selection and discretization of numeric attributes. In: *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*. IEEE Computer Society; 1995, p. 388–388.

11. from Jed Wing, M.K.C., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., et al. *caret: Classification and Regression Training*; 2014. URL: `http://CRAN.R-project.org/package=caret`; r package version 6.0-24.
12. Cristianini, N., Shawe-Taylor, J.. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press; 2000.
13. Wang, J.C., Wang, J.F., He, K.W., Hsu, C.S.. Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio low-level descriptor. In: *International Joint Conference on Neural Networks*. IEEE; 2006, p. 1731–1735.
14. Duan, K.B., Keerthi, S.S.. Which is the best multiclass svm method? an empirical study. In: *Multiple Classifier Systems*. Springer; 2005, p. 278–285.
15. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F.. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*; 2014. URL: `http://CRAN.R-project.org/package=e1071`; r package version 1.6-2.