2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin).

Audio-based early warning system of vehicle approaching event for improving pedestrian's safety

Chih-Jung Lee, Yu-Hao Tseng, and Pao-Chi Chang
Department of Communication Engineering
National Central University
Taoyuan City, Taiwan

***Abstract*— This paper presents an audio-based early warning system of vehicle approaching event for improving pedestrian's safety. Sound signals were collected by an external directional microphone connected to a smart phone. Multiple feature techniques, such as root mean square, zero crossings, spectral centroid, and spectral rolloff, were applied to short-time frames of audio samples. Multiple machine learning classifiers, such as K nearest neighbor, multi-layer perceptron, decision tree, and random forest, were applied to classify the audio frames to effectively detect vehicle approaching sound. The experimental results showed that the accuracy of the system is as high as 97% for two class (NoVehicle and Vehicle) classification .**

***Keywords*—vehicle approaching warning, audio detection, machine learning**

I. Introduction

More and more people tend to bring smart phones and communication headsets with them during their outdoor activity. They play the music from their phones and listen to the music via the headsets. This behavior could bring down the user's awareness on the surrounding events and increase the risk for causing accidents from rear approaching vehicles [1].

Modern technology has changed people's life style significantly and brought some side-effects such as the above example. However, the technology can also be used to reduce the risks. The computing power of the smart phones has been improved a lot in last decades. Complex signal processing and machine learning algorithms can be run on smart phones. Directional microphones, which help to reduce the noise coming from unwanted directions, can be integrated into a modern headset. Smart phones and headsets can be utilized and integrated into a system to increase the safety.

Nowadays, environmental sound recognition applications have become popular on mobile platforms. Angelos and his colleagues presented the realization of real-time environmental sound recognition on Android operation system [2].

Detecting the sound event of vehicle approaching has been presented in several works. Vancheswaran presented the system for improving bikers' safety [3]. Paulraj presented the methods to classify 4 different kinds of vehicles running in 4 different zones on the same road [4].

In this paper, we propose an early warning system for detecting vehicle approaching events. It is designed to be a cost-effective and feasible approach to detect vehicle approaching sound in the suburb area using a smart phone with an additional directional microphone. Our goal is to alert the user regarding the vehicle approaching in a short period ahead, e.g., 3 seconds, before the vehicle passing by the user. The alert should be triggered ahead of this 3-second period.

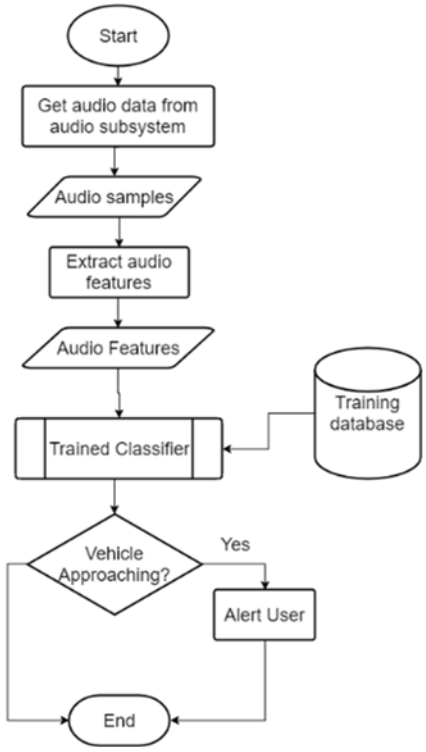

Fig. 1. System work flow of the proposed early warning system

II. System Approach

The proposed system is the combination of a directional microphone and a smart phone. Audio signal is collected from the directional microphone and passed through the audio processing subsystem of the smart phone's operating system. The vehicle approaching detection application running on the smart phone obtains the audio data from the subsystem. Whether the directional microphone is wired or wireless, the operating system will be able to handle the connection and the

978-1-5386-6095-9/18/$31.00 ©2018 IEEE

audio data path. For the simplicity of the research purpose, we use a wired directional microphone; the prediction model which can also be used on smart phones is evaluated on PC.

The working flow of the system is shown in Fig. 1. Each major block is described as follows. Single channel audio input with 44100 Hz sampling rate and 16-bit floating-point resolution is applied in our research as the specification is guaranteed to work with all the devices of the mainstream operating system on the shelf. The short-time frame size is 8192 samples that is about 0.185 seconds. The detection is performed in every single frame.

III. Feature Extraction

Feature extraction function generates a set of features that represent the characteristics of the sound signal in the short-time frame.

In this work, we apply several analysis method which are commonly used by speech and environmental sound recognition research. We worked with the jAduio[5] audio feature extraction program to obtain four features mentioned below.

Root mean square (RMS) is a measure of the energy of the frame by computing the root-mean square value of the samples during the frame [6].

Zero crossings (ZCS) is a measure of the number of times the signal value crosses the zero axe [6].

Spectral centroid (SPC) is the center of 'gravity' of the spectrum. It is calculated as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as weights [6].

Spectral rolloff point (SPR) is a measure of the amount of the right-skewedness of the power spectrum. It is calculated as the fraction of bins in the power spectrum at which 85% of the power is at lower frequencies [6].

IV. Classification Algorithms

Weka Classification Library [7] was used in this work. It is a collection of machine learning algorithms, which provides a number of classification models.

The extracted features were sent to four classifiers for detecting the vehicle approaching event residing the audio frames: *k* Nearest Neighbor (kNN) [8], Multi-layer Perceptron (MLP) [9], Decision Tree [10], and Random Forest [11].

V. Data Collection

As there was no open standard database for this research, we collected data with a directional microphone connected to a smart phone in the suburb area. To simulate the real situation that users will encounter, the height we placed the microphone from the ground was about 140~165 cm, and sounds were recoded on straight and crooked sections.

The applicable environment of this work is restricted to two-lane paved roads with speed limited to 50 km/hr and background sound level is around 40~50 dBA[12]

The collection includes three classes of recordings: 83 sedan car approaching, 52 scooter approaching and 63 background sounds. Each vehicle approaching recording is cut into the same length by a Python script of taking the 3-sec samples before the largest sample value.

VI. Evaluation Results

We use 60% of our data collection as the training set and 40% of it as the testing set. The objective of the classification task is to classify two labels: NoVehicle and Vehicle.

The first evaluation is done with our own dataset. The accuracy of each classifier is quite high, at least 97% and up, as shown in Fig. 2.

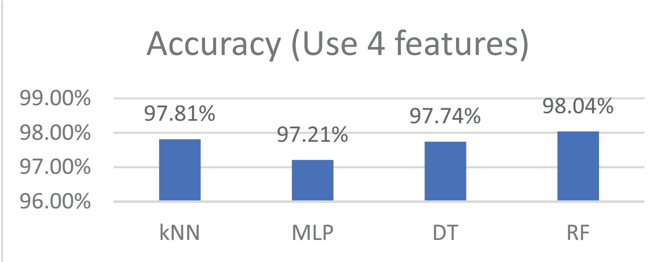

Fig.2. The accuracy of four classifiers of two categories

The second evaluation the Vehicle label is subdivided into three labels to stand for different zones of time before the vehicle passing the microphone. VehicleZ1, VehicleZ2 and VehicleZ3 stand for frames from far to near, shown as in TABLE I.

TABLE I.
Classification Label Denfinitons For Different Distances

Frames	2~3 secs before passing	1~2 secs before passing	1 sec before passing
Label	VehicleZ1	VehicleZ2	VehicleZ3

By examining the confusion matrix, shown as in TABLE II, we note that Decision Tree and Random Forest are more reliable to detect the frames at the close time zone, in which no VehicleZ3 is classified as NoVehicle. This low false negative rate ensures the safety of the pedestrian.

VII. Conclusion and Futurework

The evaluation results showed that the proposed system utilizing devices on the shelf is feasible for doing the task of detecting the vehicle approaching and giving early-warning. The experimental results show that the accuracy of the vehicle detection is high. This can be applied to realtime smart phone applications. More environmental sound events need to be taken into consideration for reducing false alert.

TABLE II.
COMFUSION MATRIX FOR CLASSIFYING DIFFERENT TIME ZONE

KNN (using 4 features) k=1				
Classified as =>	NoVehicle	VehicleZ1	VehicleZ2	VehicleZ3
NoVehicle	377(96.41%)	8(2.04%)	6(1.53%)	0(0.00%)
VehicleZ1	15(5.63%)	171(64.28%)	68(25.56%)	12(4.51%)
VehicleZ2	5(1.50%)	91(27.32%)	184(55.25%)	53(15.91%)
VehicleZ3	1(0.29%)	7(2.08%)	56(16.66%)	272(80.95%)
MLP (using 4 features)				
Classified as =>	NoVehicle	VehicleZ1	VehicleZ2	VehicleZ3
NoVehicle	386(98.72%)	5(1.27%)	0(0.00%)	0(0.00%)
VehicleZ1	25(9.39%)	181(68.04%)	53(19.92%)	7(2.63%)
VehicleZ2	7(2.10%)	76(22.82%)	209(62.76%)	41(12.31%)
VehicleZ3	7(2.08%)	4(1.19%)	58(17.26%)	267(79.46%)
Decision Tree (using 4 features)				
Classified as =>	NoVehicle	VehicleZ1	VehicleZ2	VehicleZ3
NoVehicle	378(96.67%)	10(2.55%)	3(0.76%)	0(0.00%)
VehicleZ1	16(6.01%)	199(74.81%)	48(18.04%)	3(1.12%)
VehicleZ2	5(1.50%)	106(31.83%)	169(50.75%)	53(15.91%)
VehicleZ3	0(0.00%)	10(2.97%)	40(11.90%)	286(85.11%)
Random Forest (using 4 features)				
Classified as =>	NoVehicle	VehicleZ1	VehicleZ2	VehicleZ3
NoVehicle	387(98.97%)	3(00.76%)	1(00.25%)	0(0.00%)
VehicleZ1	20(7.51%)	186(69.92%)	57(21.42%)	3(1.12%)
VehicleZ2	5(1.50%)	88(26.42%)	198(59.45%)	42(12.61%)
VehicleZ3	0(0.00%)	10(2.97%)	56(16.66%)	270(80.35%)

REFERENCES

[1] R. Lichenstein, D. Smith, J. Ambrose, L. Moody, "Headphone use and pedestrian injury and death in the United States: 2004-2011", *Injury prevention*, vol. 18, pp. 287-290, 2012.

[2] A. Pillos, K. Alghamidi, N. Alzamel, V. Pavlov, and S. Machanavajhala, "A real-time environmental sound recognition system for the Android OS," in Proceedings of Detection and Classification of Acoustic Scenes and Events, September 2016, Budapest.

[3] VK Ananthanarayanan, "Audio based detection of rear approaching vehicles on a bicycle", New Brunswick, New Jersey, 2012.

[4] M. P. Paulraj, and et al. "Moving Vehicle Recognition and Classification Based on Time Domain Approach" Procedia Engineering, Volume 53, 2013.

[5] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle. "jaudio: A feature extraction library", In Proc. of ISMIR 2005, pages 600--603, 2005.

[6] G. Tzanetakis, P. Cook, "Multifeature audio segmentation for browsing and annotation", *Proc. Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1999.

[7] M. Hall et al., "The WEKA data mining software: An update", *SIGKDD Explor.*, vol. 11, no. 1, pp. 10-18, 2009.

[8] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule", *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, pp. 325-327, Apr. 1976.

[9] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning Representations by Back-Propagating Errors", *Nature*, vol. 323, pp. 533-536, 1986.

[10] J.R. Quinlan, "Induction of Decision Trees", *Machine Learning*, vol. 1, pp. 81-106, 1986.

[11] C. Chen, A. Liaw, L. Breiman, *Using random forest to learn imbalanced data*, 2004.

[12] M. E. Nilsson, "A-weighted sound pressure level as an indicator of perceived loudness and annoyance of road-traffic sound", *Journal of Sound and Vibration*, vol. 302, pp. 197-207, 2007.