System of audio mining for obtaining traffic flow characteristics

A.S. Privalov
Department of Information System and Technologies
Samara National Research University
Samara, Russia
e-mail: asprivalov96@gmail.com

O.K. Golovnin
Department of Information System and Technologies
Samara National Research University
Samara, Russia
e-mail: golovnin@ssau.ru

Abstract[1]

The paper describes traffic flow analysing system based on audio recording. It describes the system architecture, specifies implementation features, and also provides options for using the system to improve the efficiency of traffic flow analysis.

Keywords: audio mining; traffic analyzing; audio recording; audio signals.

1. Introduction

Traffic flow characteristics analysis is the most important part in creation of an intelligent transport system infrastructure [1].

It is possible to organize effective management of transport flow only if the necessary relevant information of traffic flow available [2].

Aggregation of information of the current status of the transport flow is carried out by using various technical devices like: loop sensors, video cameras, detectors. Such an approach is not without drawbacks, since such an implementation requires significant costs for installation, operation and support [3]. Methods for analysing traffic are shown in Figure 1.

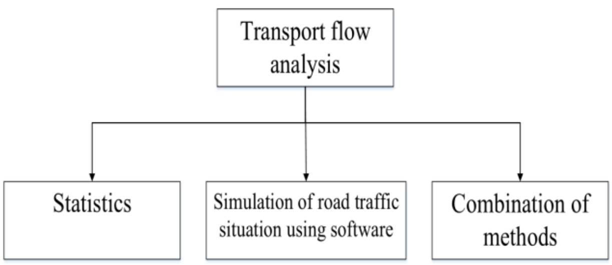

Fig 1. Traffic flow analysing methods

The main purpose of this paper is to investigate the possibility of deep machine learning to extract information about intensity of traffic flow from audio recording.

2. Methods and technologies

The main idea of the system is to extract attributes from audio recording so called audio events. We propose the audio events recognition schema (shown on Figure 1).

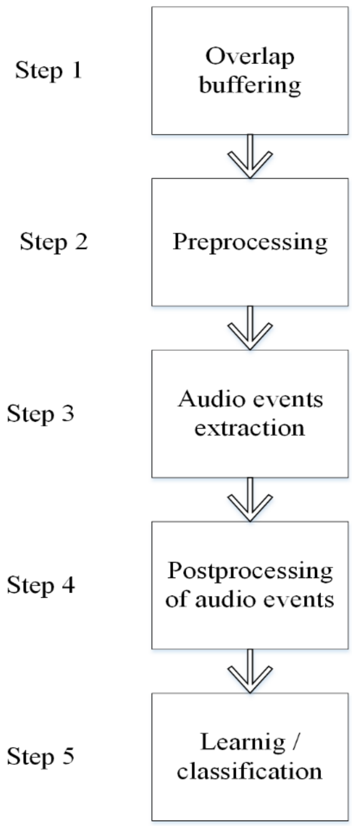

Fig 2. Audio events recognition schema

At the first step, the original audio signal is converted into an overlapping frame set. The Figure 3 shows the

Proceedings of the 7th All-Russian Scientific Conference "Information Technologies for Intelligent Decision Making Support", May 28-30, Ufa – Stavropol- Khanty-Mansiysk, Russia, 2019

schema of the audio signal represented as list of overlapped frames.

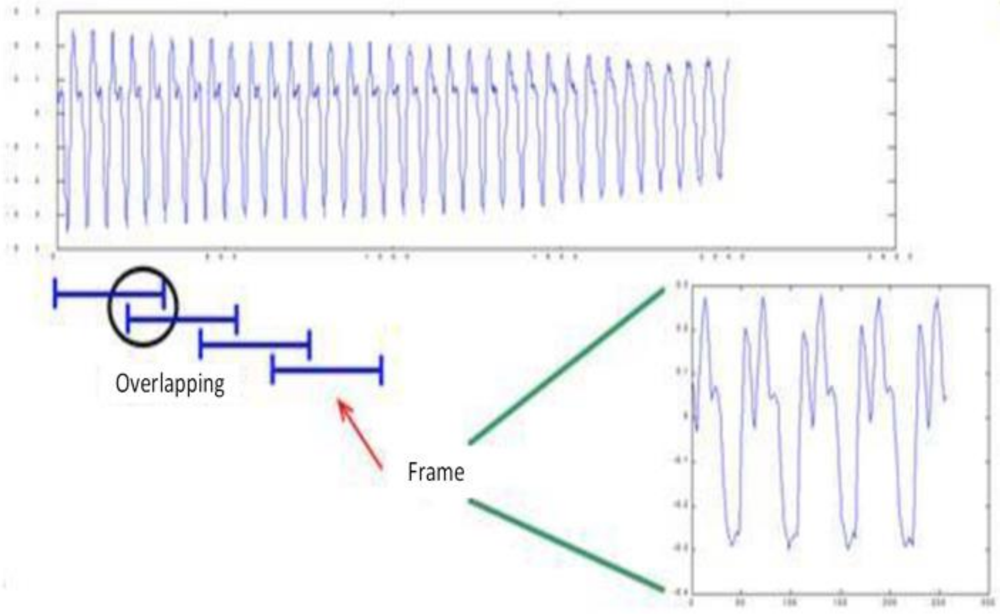

Fig 3. Overlapping frame set

The next step is preprocessing. At this step, various filters are applied to reduce the noise (random signals) generated during signal processing.

The third step is audio events extraction. The main goal is to reduce signal redundancy and highlight the most relevant information, and, at the same time, discard irrelevant information. Typically, features that describe an audio signal from different points of view are combined into one feature vector, on the basis of which the learning process takes place and then classified using the selected learned model.

Postprocessing of audios events is fourth step in audio event. In general, this is normalization of a signal.

The final step is classifier learning. In our system as a classifier have been chosen convolutional neural network (CNN).

3. Implementation

3.1. Data processing

Initially, the analogue electrical signal is digitized. Since an analogue signal mathematically consists of an infinite, continuous in time set of amplitude point-values, in the process of measurement only a finite series of values are separated from it at discrete points in time.

The first step in obtaining information from an audio file is the Fourier transform. The fast Fourier transform is used to speed up the conversion process [4]. Example of the processed file is shown in Figure 4.

After the signal conversion and its processing, it is necessary to detect an audio event. CNN allows to detect predefined patterns to further determine the intensity of movement [5]. The result of processing with CNN is a detected vehicle. An example of a detected vehicle is shown in Fig. 4.

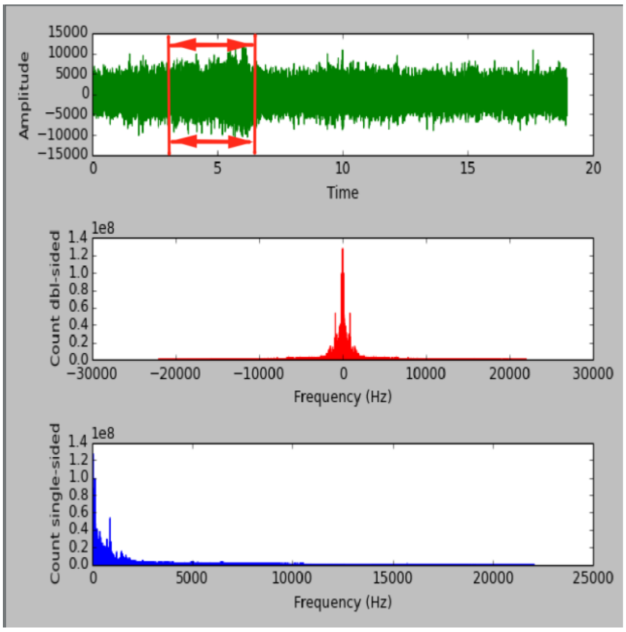

Fig 4. Processed signal

CNN model training is performed on the natural sounds dataset from the urbansound8k dataset [6]. This dataset contains 8732 labeled sound excerpts (≤4s) of city sounds from 10 classes: car and its sounds, air conditioner, children playing, dog bark, drilling, engine idling, gun shot, jack-hammer, siren, and street music. These data are divided into different folds.

The main source of patterns we use self-recorded data as well as data from open source data storages.

3.1. System architecture

During the process of a system development has been proposed a system architecture (see Fig. 5).

System consists of:

- The client device;
- Web application;
- Virtual machine working as remote server with deployed application;
- A database.

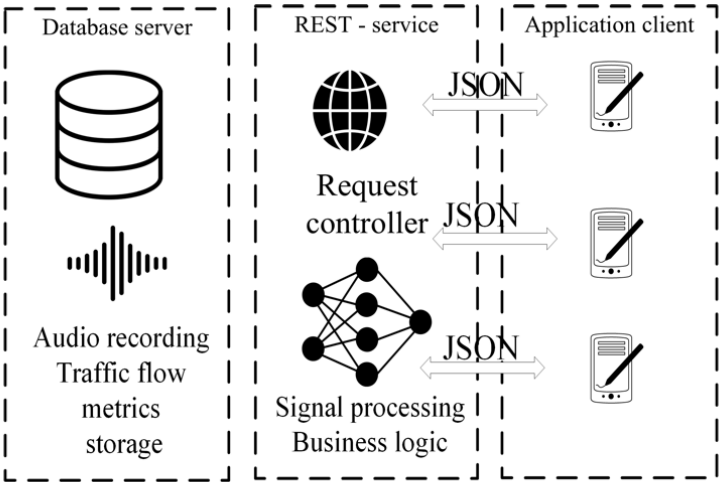

Fig 5. System architecture

The client device, which is a smartphone, records audio with a voice recorder. Further, if there is a connection to the Internet, the recorded audio is transferred to the remote server. The file contains metadata such as the geolocation of recording, as well as the time when this record was made. After the data is transferred to the server, the received signal is processed, and the necessary information is extracted to compile the characteristics of the traffic flow. After that, the received information is persist to the database.

The system is implemented on the basis of a client-server architecture. The server part of the system is implemented by using Java and Python languages. The choice of these languages is determined by the fact that Java is one of the easiest languages to create and configure servers, and Python allows you to quickly convert audio files, as well as to extract the necessary characteristics from the processed data. PostgreSQL is used as a database management system because it allows geo-referencing of events using PostGIS spatial extensions. The client part of the system is developed in the Swift programming language for work on mobile devices running iOS; further development of a mobile application for Android devices will be performed.

4. Results

To test developed system have been collected data from open audio file storages [7] and using iPhone App on the street. Collected data presented on the storage service Dropbox [8].

The result of the work of the CNN by detecting vehicles is shown in Table 1. First five files were given from open data storages and last five files were received manually.

After training of CNN and applying received data was received 62.5% accuracy. However, some files are not long enough (<5s) to be representative data. Also, dataset that trained CNN don’t have water sounds (personal records were made nearby puddles).

Table 1. Results

File name	Number of passed vehicles	Calculated number of vehicles	Percentage of vehicles, %
Traffic sound 1	8	6	75
Traffic sound 2	3	3	100
Traffic sound 3	54	37	69
Traffic sound 4	22	19	86
Traffic sound 5	13	10	77
Personal-record-1	3	0	0
Personal-record-2	4	1	25
Personal-record-3	5	4	80
Personal-record-4	8	5	63
Personal-record-5	2	1	50

5. Conclusion

The main results are listed below:

- Proposed method of analysis of traffic flow;
- Method is implemented as software in programming languages like Swift, Java, Python and SQL;
- Developed convolutional neural network and trained on datasets for pattern detection in audio recordings;
- Achieved effectivity in 62.5 %.

References

1. Golovnin O.K. Decision Support in Intelligent Transport Systems / O.K. Golovnin, A.A. Fedoseev, A.V. Sidorov // Proc. of the 3rd International Conference “Intelligent Technologies for Intelligent Decision Making Support”, Ufa, Russia, 2015. P. 127-130.
2. Trajectory data and flow characteristics of mixed traffic / Kanagaraj V. et al. // Transportation Research Record: Journal of the Transportation Research Board. 2015. No. 2491. P. 1-11.
3. Golovnin O., Stolbova A., Ostroglazov N. An Analysis of Road Traffic Flow Characteristics Using Wavelet Transform // Studies in Systems, Decision and Control, Springer, Cham, 2019. Vol. 199. P. 433-445.
4. Hågenvik H.O., Malema M.E., Skaar J. Fourier theory of linear gain media // Physical Review A. 2015. Vol. 91. No. 4. P. 043826.
5. Rothmann D. What’s wrong with CNNs and spectrograms for audio processing? URL: towardsdatascience.com/whats-wrong-with-spectrograms-and-cnns-for-audio-processing-311377d7ccd.
6. Urban Sound Data Set. URL: urbansounddataset.weebly.com/urbansound8k.html.
7. Free Sound Data Set. URL: https://freesound.org/.
8. Authors’ Sound Data Set. URL: https://www.dropbox.com/sh/bwu8vnek1tlgduc/AACh76tcFD3tpFpyMUBowNEaa?dl=0