Comparison of Various Techniques for Emergency Vehicle Detection using Audio Processing

Eshwar Prithvi Jonnadula and Pabitra Mohan Khilar

Department of Computer Science and Engineering
National Institute of Technology, Rourkela {714cs1038,pmkhilar}@nitrkl.ac.in

Abstract. VANETs is an important part of wireless networking. Vehicular movement is unending expanding wherever on the planet and can cause horrible activity clog. The greater part of the signals till date include a settled green signal arrangement, so the green signal timing is done without considering emergency vehicles. In this way, emergency vehicles, for example, are stuck in congested driving conditions and postponed in achieving their goal can prompt loss of property and important lives. In this paper, we do a comparative study of different methods which are used in identifying the emergency vehicle present on the road. This identification of emergency vehicles is useful in the development of smart cities. This is tried on genuine dataset got from Google Audio Dataset which had obtained and recorded in urban avenues which include distinctive activities and noises like people talking, vehicles horns etc. We found out that an artificial neural network consisting of three hidden layers give the highest accuracy which is 3% more than one hidden layer ANN.

Keywords: VANETs · ITS · Emergency Vehicles · Audio Sensing · Machine Learning

1 Introduction

The Vehicular Ad-Hoc Networks(VANETS) is a rising region of networking. It is a type of Mobile Ad-Hoc Networks. The VANETS consists of three major modes of communication such as Vehicle to Vehicle(V2V), Infrastructure to Infrastructure(I2T) and Vehicle to Infrastructure(V2I) [6]. As of late, more mishap scenarios came into existence. Because of this, roads are observed to be highly congested and occupied. With the assistance of committed short-extend correspondence (DSRC), Vehicular Ad-Hoc Networks sets up correspondence among different vehicles which are altering their course much of the time. Vehicles straightforwardly speak with various vehicles and send data in regards to automobile overloads, cautioning information with the infrastructure which is settled hardware on roads [5].

As of now, transportation has become an irreplaceable part of human life. Normally 40% of the human population spends one hour on the road on average every day. The human race has become significantly dependent on transportation as of late, due to this, transportation themselves confront a few open doors as well as a few difficulties also. To start with, the blockage has turned into an undeniable problem across the globe as the count of vehicles on the streets continues to increase. For instance, Beijing, China, had an aggregate of 4 lakh vehicles toward the start of this decade and added another 8 lakh that year. Blockage might cause an expansion in gas utilization, air contamination, what's more, difficulties in actualizing plans for open transportation [13].

The spectrum of natural sounds is large and it incorporates the range of sounds generated in various real world conditions. These sounds give us useful information about human activities. Such data is vital for applications that process sound and video streams. Cautioning sounds, for example, emergency vehicles, smoke alerts and medicinal checking alerts are of an extraordinary significance, as they are generally intended to caution individuals of perilous circumstances. Programmed recognition of such sounds may have numerous applications for clever frameworks that need to react to their acoustic surroundings.

Over the most recent couple of decades, programmed processing of sound-signs attracted extraordinary intrigue both academia and industry. In any case, most exertion has been put resources into discourse and music handling and essentially less in the emergency sounds. The greater part of works managing natural sounds are focused on the sound arrangement, e.g., blasts versus entryway pummel versus canine barks or location of unusual sound occasions. The fundamental research issue these works manage is picking an arrangement of reasonable sound highlights. Basic highlights being used for these undertakings are the Mel-frequency cepstral coefficients (MFCC), wavelet-based highlights and individual fleeting and recurrence highlights. The arrangement of sound categories is constrained significantly in these works, therefore these are generally difficult, to sum up. Along these lines, to make a dependable caution sound identifier, a more particular procedure ought to be planned.

This paper is divided into five sections as follows. In the first section, we briefly discussed VANETs and the importance of identification of emergency vehicles. In the second section, the previous works and the different techniques used by the authors are explained. The third section explains the different methods used in the identification of emergency vehicles. The fourth section gives comparative results of various methods explained in section three. The fifth section discusses the conclusion and the future work involved.

2 Related Works

The National Highway Traffic Safety Organization (NHTSA) has gathered the rescue vehicle crash information for the United States of America somewhere in the range of 1992 and 2011. There were an expected yearly average of 4500 car accidents and lot more damage crashes including an emergency vehicle. In last two decades, 662 people were executed and 52,000 people were evaluated to be harmed in such crashes, including emergency vehicle drivers, travellers, non-inhabitants and tenants of different vehicles. According to the insights of crisis vehicle mischances in the United States of America, there were about 31,600 mishaps including fire fighting vehicles brought about 645 fatalities over a multi-year time frame (2000-2009) and 300 fatalities that happen each year amid police interests [12]. From the present issue area, it tends to be comprehended that, there is a genuine requirement for an insightful movement administration framework for the compelling administration of both the typical and emergency vehicles.

In view of this understanding, a few past works have attempted to bargain particularly with the errand of caution sound discovery. A general model of alert sound cannot be built easily, therefore, most research is focused on identifying specific cautions ordinarily alarms of emergency vehicles of a particular nation. As expected a huge amount of research works do not show any performance gains since they don't show all around ok encompassing foundation sounds and for the most part don't consider moves in frequency because of the Doppler effect.
For instance, in the idea introduced in [4], the researchers attempt to identify a little arrangement of pre-chosen cautioning audio in a reenacted domain by cross-connection.

In [3] a Machine Learning technique(ANN) was utilized to recognize law enforcement vehicles in Italy.

A different approach for distinguishing alarms of emergency vehicles in Italy is introduced in [9]. The framework keeps running progressively by evaluating the pitch frequency and contrasting it with pre-characterized alarm frequencies.

Another straightforward framework for alarm identification that keeps running continuously is portrayed in [11]. In view of the periodicity of caution sounds, the aftereffects of autocorrelation are dissected by an ML-based classifier customized to acoustic crisis signals in Germany.

In [8], a strategy for distinguishing an emergency vehicle alarm sound in Taiwan in displayed. It utilizes frequency matching to find the longest basic subsequence.

3 Siren Detection

Sire detection id the process of identifying the presence of an emergency vehicle on the roads using the siren sound it produces using audio processing.
This section is divided into two parts. Part A discusses the basic siren detection

using the pitch detection method. In Part B we discuss the various machine learning based methods.

3.1 Basic Pitch Based Detection

The issue of the siren detection has been assaulted with a procedure chipping away at two levels. First MDF (Module Difference Function), a period space strategy, intends to group each bit of the sound motion as pitched or unpitched. This initial step can be isolated in MDF count and Peak Looking [10]. The Peak looking gives us the estimation of the pitch frequency. Toward the finish of the main stage we get a flag speaking to the pitch developing after some time, we call this Pitch(t).

Also, Pitch(t) is investigated with the end goal to perceive a period design of the coveted siren appearance or nonappearance of the siren.

$$MDF(w,l) = \sum_{n=0}^{ws-1} \mid y(n+l)\%ws - y(n) \mid \tag{1}$$

Where
y(n) is the audio signal at nthsample
l is the lag
N is the length of audio signal
w is the window number
ws is the window size

Using the MDF vs the lag plot, the pitch of the signal is detected.

$$pitch(w) = \frac{sr}{l} \tag{2}$$

Where
sr is the Sample Rate
l is the lag at which MDF is minimum
w is the window number

Here Pitch is the basic fundamental frequency of the sound or audio signal. The window is the small part of the audio signal. The audio signal is divided into a number of windows for the purpose of easy processing of the audio signal. The sample rate is the rate at which the analogue signal is converted to a digital audio signal. Lag is nothing but a particular instance of the sample.

Emergency signals all in all have an occasional example which rehashes over a brief length of time. Moreover, they additionally have a place with particular frequency groups which loans them the high pitch sound that they are typically connected with.

Algorithm 1 Algorithm to find Module Difference Function

```
procedure MDF($y, l, ws$)
    $r \leftarrow 0$                                         ▷ r stores the result
    $i \leftarrow 0$
    while $i \neq ws$ do
        $temp \leftarrow\mid y(i+l)\%ws - y(i) \mid$          ▷ l is the lag; ws is the window size
        $r \leftarrow r + temp$
        $i \leftarrow i + 1$
    return $r$
```

We use these qualities relating to emergency signals by utilizing a pitch identification calculation.

An Altered Pitch Discovery (MDF) calculation which is a less computationally costly form of the autocorrelation work is utilized to discover the pitch. Utilizing the MDF versus the lag plot, the pitch is distinguished. An emergency signal estimator that ascertains the division of time for which the pitch stays inside the frequency band for emergency signals predicts the nearness of emergency signals in the sound example.

Algorithm 2 Algorithm to find Pitch

```
procedure PITCH($ws$)
    $pitch\_list[\ \ ]$
    $nw \leftarrow \lfloor \frac{len(y)}{ws} \rfloor$        ▷ nw is the number of windows; ws is the window size
    $i \leftarrow 0$
    while $i \neq nw$ do
        $mdf\_list[\ \ ]$
        $y\_clip = y[i * ws : i * ws + ws]$
        $j \leftarrow 0$
        while $j \neq ws$ do
            $temp \leftarrow (MDF(y\_clip, j, ws)$
            $mdf\_list.add(temp)$
        $min\_value = min(mdf\_list)$
        $min\_index = mdf\_list.index(min\_value)$
        $pitch\_detected = \frac{sr}{min\_index}$
        $pitch\_list.add(pitch\_detected)$
    return $pitch\_list$
```

From the above two algorithms we get the pitch of each window. The window domain is converted to time domain as follows

$$time = \frac{w * ws}{sr} \tag{3}$$

Where
w is the window number

ws is the window size
sr is the sample rate

Now that we have time domain and pitch at every instance of time we draw a Pitch VS Time graph as shown in Figure 4.

The probability of presence of emergency vehicle is obtained by simply dividing the pitch obtained by the standard pitch frequency of emergency vehicle as per the standard.

$$probability = \frac{pitch}{standardpitch} \tag{4}$$

3.2 Machine Learning Techniques

Machine learning (ML) is a field of AI that utilizes statistical techniques to enable digital computers to "learn" (e.g process information without being given explicit instructions.).

In this part of subsection, we briefly discuss the Artificial Neural Network which is used in our comparative study of identification of emergency vehicles.

Artificial Neural Networks The idea of ANN is formulated upon the fact that the human mind works by making the correct networks which can be reproduced using silicon and wires in place of neurons and dendrites. The human brain is comprised of around 8600 crore nerve cells called neurons. These are interconnected to other cells with the help of structures called axons. Stimuli from various sources such as inputs from sensory organs and other external sources are processed by dendrites. These inputs transform into electrical signals which travel through the neural network quickly. A nerve cell in a network decides whether it should forward the signal to other nerve cell or not. The basic structure of ANN is shown in Figure 1.

In a similar fashion, an ANN is comprised of multiple nodes which imitate the biological nervous system, specifically the neural connections of the human brain. The neurons are interconnected and interact among themselves. The nodes in the network take data as input and perform a basic operation on it. This result is then forwarded to other neurons. The output generated at each node is termed as its activation If the output generated by the neural network is 'good or desired' then adjustment of weights is not required. On the contrary, if the neural network produces 'poor or undesired' output then the network modifies the weights such that further computations result in improved output.

In this method of machine learning, we use the feature as the input layer and based on the requirement we add hidden layers. The final output layer consists of only one node which gives the output as either 1(Emergency) or 0(Non-Emergency).

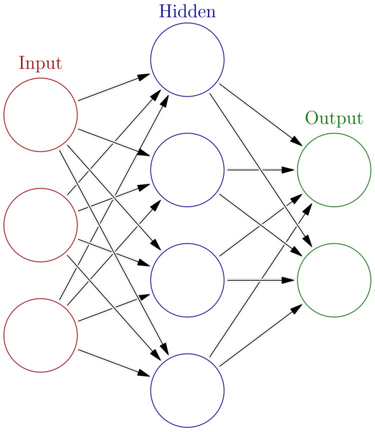

Fig. 1: Basic Structure of ANN [1]

Feature Extraction We used 34 features extracted from each audio sample. It segregate the input signal into short window frames and then computes the features for individual frame. This method produces a series of feature vectors for the entire signal.

- **Zero Crossing Rate(0):** The number of times the sign of a time series change within a frame. It approximates the frequency which is significant in a time frame.

$$ZCR = \frac{1}{2} \sum_{m=-\infty}^{+\infty} \mid sig(x(m)) - sig(x(m-1)) \mid$$

$$sig(x(n)) = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) \leq 0 \end{cases}$$

- **Energy(1):** The sum of squares of the signal values.

$$Energy = \sum_{-\infty}^{+\infty} \mid x(n) \mid^2$$

  $x(n)$ is the discrete time signal
- **Entropy of Energy(2):** The entropy of a signal is a measure of the amount of information a signal carries. It can be understood as a measure of sudden changes.

$$Entropy = -\sum_i p(x_i) ln(p(x_i))$$

  $p(x_i)$ is the probability for the signal to take values x_i

- **Spectral Centroid(3):** The spectral centroid is a measure which indicates where the "centre of mass" lies in the spectrum.

$$SpectralCentroid = \frac{\sum_{k=1}^{N} kF[k]}{\sum_{k=1}^{N} F[k]}$$

  $F[k]$ is the amplitude corresponding to bin k in DFT spectrum
- **Spectral Spread(4):** The second central moment of the spectrum.

$$SpectralSpread = \sum_{k=1}^{N} (k - SC)^2 F[k]$$

- **Spectral Entropy(5):** The entropy of the normalized spectral energies for a set of sub-frames.

$$SpectralEntropy = -\sum_{i} p(F[k]) ln(p(F[k]))$$

  $p(F[k])$ is the probability for the signal to take values $F[k]$
- **Spectral Flux(6):** Spectral flux is a proportion of how quick the power spectrum of a signal is fluctuating which is determined by comparing the previous frame of the power spectrum to the other frame.

$$SpectralFlux = \sum_{-\infty}^{+\infty} (F[k] - F[k+1])^2$$

- **Spectral Rolloff(7):** Spectral roll-off is defined as the Nth percentile of the power spectral distribution, where N is usually 85% or 95%. The roll-off point is the frequency below which the N% of the magnitude distribution is concentrated.
- **MFCCs(8-20):** Mel Frequency Cepstral Coefficients form a cepstral depiction where the frequency bands are distributed according to the Mel-scale and not linear. We use 13 Mel-frequency cepstral coefficients to describe the spectral shape of the signal. The log-energy outputs of the nonlinear Mel-scale filter-bank is taken as input and discrete cosine transform (DCT) is applied to produce the output as the coefficients.
- **Chroma Vector(21-32):** The spectral energy is represented by the twelve element buckets where each bucket represents twelve equal pitch-tempered pitch classes of western music.
- **Chroma Deviation(33):** The standard deviation of the 12 chroma coefficients.

4 Results

In this part, comparison of the various machine learning techniques for the classification of emergency vehicle detection is explained. These are implemented in python. We have taken an audio dataset provided by Google [2].

Using Algorithm 1 we find the MDF. This is used for a particular window to get the MDF VS Lag plot as shown in Figure 2.

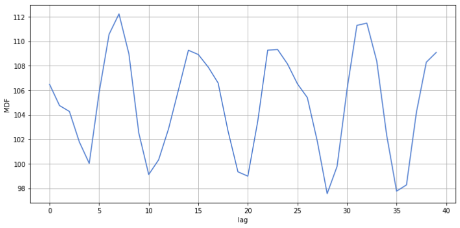

Fig. 2: The Plot of MDF VS Lag for a small interval in a window

This plot of MDF VS Lag is used to find out the pitch of a particular window. Then we easily plot the Pitch VS Time plot as shown in figure 3.

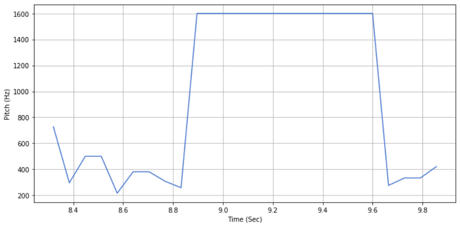

Fig. 3: The Plot of Pitch VS Time for a small interval of time

The probability of the presence of an emergency vehicle is obtained dividing the pitch by the standard pitch frequency of emergency vehicle siren.

Then we obtain the Probability VS Time plot as shown in figure 4.

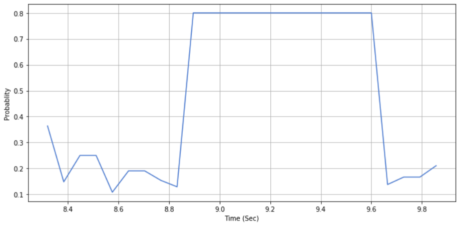

Fig. 4: The Plot of Probability VS Time for a small interval in time

The parameters [7] used for the comparisons are

1. **True Positive Rate (TPR)** $=\frac{TP}{TP+FN}$
2. **True Negative Rate (TNR)** $=\frac{TN}{FP+TN}$
3. **Positive Predictive Value (PPV)** $=\frac{TP}{TP+FP}$
4. **Negative Predictive Value (NPV)** $=\frac{TN}{TN+FN}$
5. **False Positive Rate (FPR)** $=\frac{FP}{FP+TN}$
6. **False Discovery Rate (FDR)** $=\frac{FP}{FP+TP}$
7. **False Negative Rate (FNR)** $=\frac{FN}{FN+TP}$
8. **Accuracy (ACC)** $=\frac{TP+FP}{TP+FP+TN+FN}$
9. **F1 Score (F1)** $=\frac{2TP}{2TP+FP+FN}$

We have made a comparison for Artificial Neural Networks by incrementing the number of hidden layers and increasing the features used and thereby increasing the accuracy of perdition.

The following are the architecture of the ANN's used as shown below in figures 5 and figure 6.

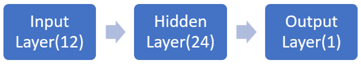

Fig. 5: Architecture of one hidden layer ANN

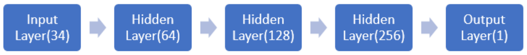

Fig. 6: Architecture of three hidden layer ANN

Using these two architectures we calculate the different parameters and compare them to find which gives the better performance in the identification of emergency vehicles. We see that the ANN with three hidden layers gives a 3% more accuracy than one hidden layer ANN in the identification of emergency vehicles as shown in figure 7.

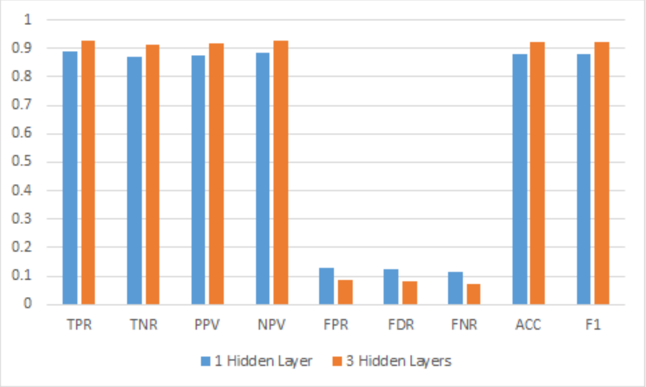

Fig. 7: Comparison of ANNs

5 Conclusion

An increased number of vehicles does not just build the reaction delay of emergency vehicles, yet it likewise increments the odds for them being engaged with mischances. The emergency vehicle entering a crossing point at a fast on a red light stances threat to activity on different streets and can cause mischances.

This paper demonstrates a comparison of several techniques for the detection of emergency vehicles using machine learning. We have also found out that the Artificial neural network with three hidden layers has the best accuracy rate i.e 3% more than the one hidden layer ANN. So this ANN can be deployed in real time to determine the existence of emergency vehicles.

Audio analysis may always not be correct as we can see in the results mentioned above in figure 7. This can be overcome by using the detection of emergency vehicles using real time detection using the traffic cameras which will be done in the future. Both audio and video detection can together decide the presence of an emergency vehicle on the road.

References

1. ANN Wikipedia. `https://en.wikipedia.org/wiki/Artificial_neural_network` (2018), "[Online; accessed Nov-2018]"
2. AudioSet by Google. `https://research.google.com/audioset/dataset/emergency_vehicle.html` (2018), "[Online; accessed Nov-2018]"
3. Beritelli, F., Casale, S., Russo, A., Serrano, S.: An automatic emergency signal recognition system for the hearing impaired. In: Digital Signal Processing Workshop, 12th-Signal Processing Education Workshop, 4th. pp. 179–182. IEEE (2006)
4. Bernstein, E.R., Brammer, A.J., Yu, G.: Augmented warning sound detection for hearing protectors. The Journal of the Acoustical Society of America **135**(1), EL29–EL34 (2014)

5. Bhoi, S.K., Khilar, P.M.: Vehicular communication: a survey. IET Networks **3**(3), 204–217 (2013)
6. Bhoi, S.K., Khilar, P.M.: Sir: a secure and intelligent routing protocol for vehicular ad hoc network. IET Networks **4**(3), 185–194 (2014)
7. Grover, J., Prajapati, N.K., Laxmi, V., Gaur, M.S.: Machine learning approach for multiple misbehavior detection in vanet. In: International Conference on Advances in Computing and Communications. pp. 644–653. Springer (2011)
8. Liaw, J.J., Wang, W.S., Chu, H.C., Huang, M.S., Lu, C.P.: Recognition of the ambulance siren sound in taiwan by the longest common subsequence. In: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on. pp. 3825–3828. IEEE (2013)
9. Meucci, F., Pierucci, L., Del Re, E., Lastrucci, L., Desii, P.: A real-time siren detector to improve safety of guide in traffic environment. In: Signal Processing Conference, 2008 16th European. pp. 1–5. IEEE (2008)
10. Meucci, F., Pierucci, L., Del Re, E., Lastrucci, L., Desii, P.: A real-time siren detector to improve safety of guide in traffic environment. In: Signal Processing Conference, 2008 16th European. pp. 1–5. IEEE (2008)
11. Mielke, M., Schäfer, A., Brück, R.: Integrated circuit for detection of acoustic emergency signals in road traffic. In: Mixed Design of Integrated Circuits and Systems (MIXDES), 2010 Proceedings of the 17th International Conference. pp. 562–565. IEEE (2010)
12. Nellore, K., Hancke, G.P.: Traffic management for emergency vehicle priority based on visual sensing. Sensors **16**(11), 1892 (2016)
13. Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C., et al.: Data-driven intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems **12**(4), 1624–1639 (2011)