一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
NS2016-88 (2016-10)

[Encouragement Talk] Enhanced Sound Mapping for Successive Vehicle Detection in Acoustic Vehicle Count System

Song LIU[†], Shigemi ISHIDA[†], Shigeaki TAGASHIRA[††], and Akira FUKUDA[†]

[†] Graduate School / Faculty of Information Science and Electrical Engineering, Kyushu University
Motooka 744, Nishi-ku, Fukuoka-shi, Fukuoka, 819–0395 Japan
[††] Faculty of Informatics, Kansai University
Ryozenji-cho 2-1-1, Takatsuki-shi, Osaka, 569–1095 Japan

Abstract Vehicle counting is one of the fundamental tasks in the intelligent transportation system (ITS). We are developing an acoustic vehicle count system that relies on two microphones at a sidewalk. The system extracts key data reflecting the road traffic conditions from received audio signals using a correlation based algorithm, the data recorded is called *sound map*. However, the acoustic vehicle count system suffers from a sparse sound map problem; a sound map becomes weak when multiple vehicles pass in front of the microphones. This paper therefore presents a new algorithm called *enhanced sound mapping* to address the sparse sound map problem. Comparison between original and enhance sound maps are presented theoretically and experimentally.

Key words Vehicle count, acoustic sensing, enhanced sound mapping.

1. Introduction

Increasing attention has been focused on the intelligent transportation system (ITS) due to the change of road transportation strategy. The main purpose of ITS is to improve the safety, efficiency, dependability, and cost effectiveness of our transportation system. According to a market research report, ITS market is expected to grow at a compound average growth rate (CAGR) of 11.57 % between 2015 and 2020, and reach $33.89 billion by 2020 [1].

In the ITS, vehicle counting is one of the fundamental tasks. In Japan, vehicle counting has been mainly conducted as a road traffic census almost every five years since 1928 by the Ministry of Land, Infrastructure, Transportation and Tourism (MLIT). The traffic census investigates temporal traffic volume, which restricts usage of the traffic data to non-realtime applications.

To retrieve realtime traffic volume, automatic vehicle count systems have been deployed. The deployment of the automatic vehicle count system is, however, limited to high traffic roads because of its high installation and maintenance costs. The automatic vehicle count systems also suffer from a motorbike counting problem. Although camera-based vehicle counters that are capable of motorbike counting are proposed, restrictions on camera location and angle put impractical restrictions on deployment.

We are developing an acoustic vehicle count system that comes with low installation and maintenance costs. Our vehicle count system is a microphone-based system installed at a sidewalk. Because sound waves are diffracted over obstacles, we can deploy microphones in a low height configuration, which drastically reduces roadwork costs closing a target road section. Our vehicle counter detects all types of vehicles as long as the vehicles generate sound.

There are several studies reporting a vehicle monitoring system using acoustic sensors [2] ~ [5]. These studies used a microphone array to draw a sound map, i.e., a map of time difference of vehicle sound on different microphones. The studies manually analyzed the sound map and demonstrated that the sound map can be used for vehicle counting.

In our previous studies, we also have implemented vehicle detectors using a sound map [6] ~ [8]. Although our vehicle counters successfully counted vehicles with an F-measure of up to 0.92, many false-negative detections occurred because of a sparse sound map problem.

We therefore presents an enhanced sound mapping scheme, a modified version of sound mapping presented in [3], to address the sparse sound map problem. The enhanced sound mapping is based on an observation that sound delay mapped as a peak on a cross-correlation function barely changes in a short time. We sum up cross-correlation functions calculated from sound signals at slightly different time to enhance peaks on cross-correlation functions. Initial experimental evaluation reveals that the enhanced sound mapping reduces the number of false negative detections from 11 to 4; recall was improved by approximately 10%.

The remainder of this paper is organized as follows. Section 2 overviews related works on vehicle counting. Section 3 describes our vehicle count system and a sparse sound map problem. We then present the enhanced sound mapping in Section 4, followed by experimental evaluations in Section 5. Finally, Section 6 concludes the paper.

2. Related Works

To the best of our knowledge, sound map enhancement is novel in the field of acoustic vehicle detection. This section briefly looks through related works in terms of vehicle counting.

Current vehicle counters are divided into two types: intrusive and non-intrusive.

Loop coils and photoelectric tubes are categorized into the intrusive vehicle counters. These vehicle counters share the same defect that they are required to be installed under the road surface. The installation and maintenance of these counters is by far the most dominating cost factor in their life cycle due to a roadwork closing a target road section. Loop coils and photoelectric tubes also have difficulties in motorbike detection due to their small coverage.

The Non-intrusive counters including laser, infrared, ultrasound, radar, and video are supposed to overcome the cost problem, yet they have their own problems instead. The non-intrusive vehicle counter needs to be installed above or by a road for better performance. Deployment above a road requires high installation and maintenance costs in terms of roadwork. Roadside non-intrusive vehicle counters are capable of single lane detection and only works on small roads. Most of non-intrusive counters are based on laser, infrared, or ultrasound. These counters have small coverage and face the motorbike detection problem.

To reduce installation and maintenance costs, camera-based vehicle counters using CCTVs installed in the environment have been proposed [9], [10]. CCTVs, however, are only available in limited areas, especially in city areas. Performance of vehicle counters using CCTVs is also affected by weather condition because camera location and angle are not suitable for vehicle counting but for security surveillance.

On the contrary, acoustic approach is a promising candidate for vehicle counting at a low installation and maintenance costs. Using a roadside microphone array, we can locate a sound source, i.e., a vehicle on a road. Acoustic approach is capable of counting vehicles on multiple lanes at a sidewalk because sound waves are diffracted over obstacles.

Several studies have reported on a vehicle monitoring system using acoustic sensors. Forren et al. and Chen et al. proposed traffic monitoring schemes using a microphone array [2] ~ [4]. The monitoring schemes draw a sound map, i.e., a map of time difference of vehicle sound on different microphones and analyze the sound map to monitor vehicles. The monitoring schemes are missing design details of vehicle counting. The monitoring schemes also install a microphone array in a high height configuration at a roadside and monitor vehicles on multiple lanes. The high height configuration fails to reduce installation costs in terms of safety installation.

Barbagli et al. reported an acoustic sensor network for traffic monitoring [5]. The acoustic sensor network installs sensor nodes at road sides. Each sensor node draws a sound map and combines the sound map with an energy detection result to monitor traffic flow distribution. The sensor network requires many sensor nodes at both sides of the road to monitor realtime traffic flow, which results in high deployment and maintenance costs. The paper also lacks an evaluation of accuracy on vehicle counting because the sensor network focuses on monitoring traffic flow with small energy consumption.

We also have developed an automatic vehicle counter as a state machine that keeps track of curves drawn on a sound map [6] and as a pattern detector using template matching based on dynamic time warping (DTW) [7], [8]. Our previous vehicle counters, however, suffers from false-negative detections for successive vehicles resulting in performance degradation.

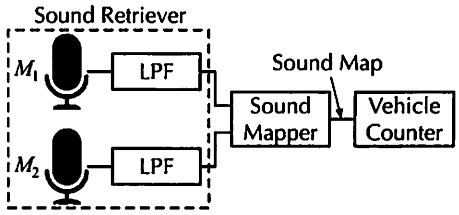

Figure 1 Overview of acoustic vehicle count system

3. Acoustic Vehicle Count System

3.1 System Overview

Figure 1 depicts the overview of our acoustic vehicle count system. The vehicle count system consists of three components as shown in Fig. 1. The sound retriever is where the system collects acoustic information from the road. Our system uses only two microphones and low pass filters (LPFs) have been applied for noise reduction to increase the robustness of signal. After receiving data from the sound retriever, sound mapper calculates time difference of arrival (TDOA) between left and right channel recordings to generate a sound map. Finally, the vehicle counter counts vehicles from the sound map by applying a simple vehicle detect algorithm.

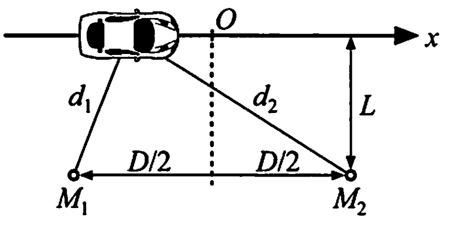

Figure 2 Microphone setup

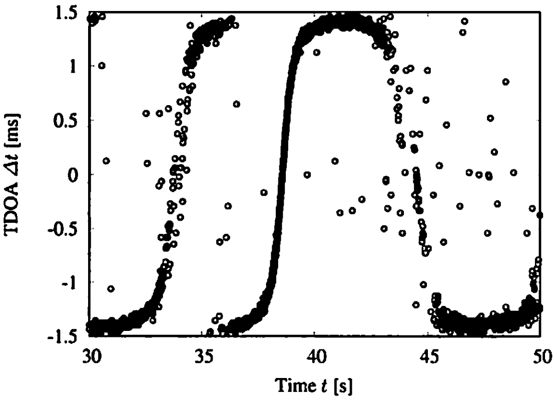

Figure 3 Example of sound map

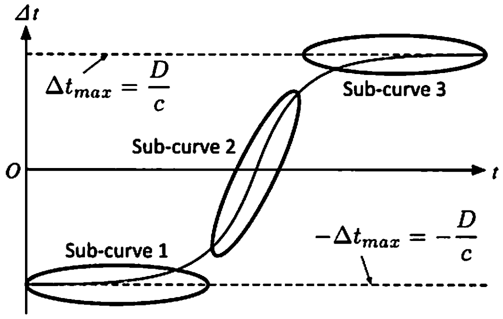
Δt
$\Delta t_{max} = \frac{D}{c}$
Sub-curve 3
Sub-curve 2
O
t
$-\Delta t_{max} = -\frac{D}{c}$
Sub-curve 1

Figure 4 Vehicle passage drawn on sound map

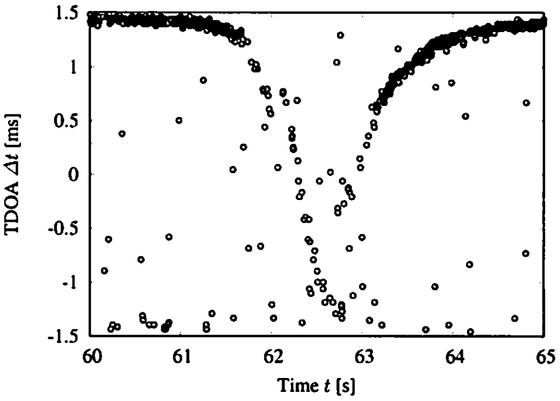

Figure 5 Example of sparse sound map

Figure 2 depicts the microphone setup. Two microphones M_1 and M_2 are installed parallel to a road at a distance of D. L stands for distance between the two microphones. Sound signals generated by the vehicle travel to the microphones with different route d_1 and d_2, therefore form a time difference between signals received by M_1 and M_2. Let x be the location of a car. The time difference Δt of sound arrival on microphones is calculated as $\Delta t = (d_1 - d_2)/c$, where c is the speed of sound in air. We therefore derive

$$\Delta t = \frac{1}{c}\left\{\sqrt{\left(x+\frac{D}{2}\right)^2+L^2}-\sqrt{\left(x-\frac{D}{2}\right)^2+L^2}\right\}. \tag{1}$$

Equation (1) gives us the ability to locate the vehicle if we can calculate Δt from the sound signals. Sound difference can be derived by cross-correlation function defined as $R(t) = s_1(t) * s_2(t)$, where $s_1(t)$ and $s_2(t)$ are the pair of signals received by the microphones and $*$ denotes the convolution operation. The cross-correlation function obtains its maximum at $t = \Delta t$. We use Generalized Cross-Correlation (GCC) function [11] instead of cross-correlation function to increase the robustness against noise.

A typical soundmap, i.e., TDOA Δt as a function of time, is shown in Fig. 3. As a car passes in front of the microphones, Δt rises up or drops down on the sound map and draws an S-curve; direction of the S-curve corresponds to direction of the car.

The count algorithm is a state machine that detects sub-curves on sound map. In the algorithm, an S-curve is divided into three sub-curves as shown in Fig. 4; sub-curves 1, 2, and 3 are observed when a car is approaching, passing in front of, and leaving from the microphones, respectively. The count algorithm starts with a sub-curve 1 detection state and keeps track of TDOA to detect sub-curve 2 followed by sub-curve 3.

To increase the overall accuracy against long wheelbase vehicles, we apply a preprocessing on a sound map prior to applying the count algorithm. Every data point on the sound map is replaced with a weighted rectangle, as they overlap with each other, a more explicit bold curve is formed. Not only does this preprocessing help with the detection of long wheelbase vehicles, but it also reduces noise greatly.

3.2 Sparse Sound Map Problem

In acoustic vehicle count systems using sound map, the number of microphones is usually set to four or more because of the need of a precise and explicit sound map. We only rely on two microphones in our vehicle count system, we derive a sparse sound map. Figure 5 shows an example of a sparse sound map. With these sparse vague tracks of TDOA, the vehicle count algorithm fails to detect vehicles resulting in false negative detections.

The counting algorithm fully depends on the sound map, which is highly related to the accuracy and clarity of sound maps. Sound maps made by the dual microphone system work fine most of the time, but when vehicles coming successively in the same direction or simultaneously in oppo-

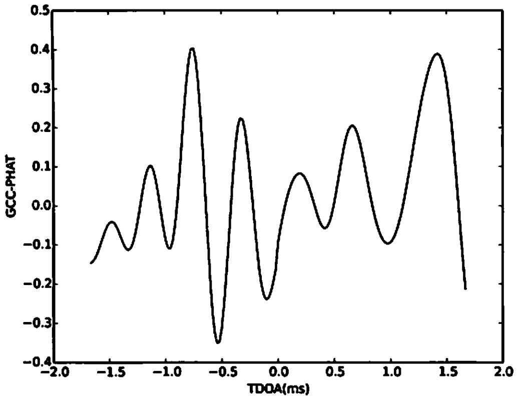

Figure 6 GCC result jeopardized by noize

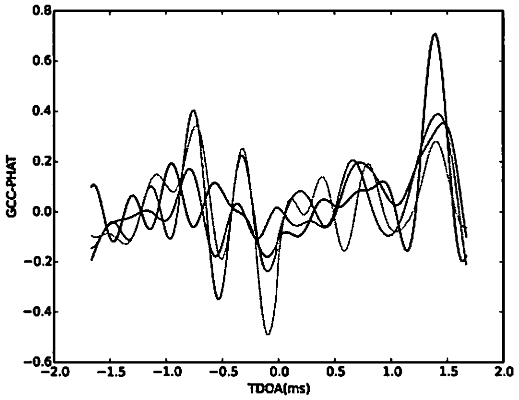

Figure 7 Four adjacent GCC results with short time differences

site direction, the sound interference between cars will make the sound map sparse and implicit.

The sparse sound map originates with peak detection in a GCC result. Figure 6 depicts the example of a GCC result jeopardized by noise. Although this result has the peak at TDOA $\Delta t = 1.4$, noises have generated an even more superior peak at $\Delta t = -0.6$. This will make the sound mapper record a false TDOA and ignore the true result that is very common in the traditional sound mapping. Because we cannot enhance the sound signal or reduce the interference directly, GCC cannot solve the sparse sound map problem.

4. Enhanced Sound Mapping

4.1 Key Idea

To address the sparse sound map problem described in the previous section, we developed an enhanced sound mapping scheme. Our key idea is to sum up successive GCC results derived at slightly different time. TDOA barely changes in a short time duration such that a vehicle moves negligible distance. Figure 7 shows four adjacent GCC results. Time difference between each waveform is 2.9 milliseconds. Waveforms of those results differ from each other. Although each waveform exhibits many peaks that make difficult to accurately detect vehicles, all of them have a peak at TDOA $\Delta t = 1.4$ milliseconds. In a short time duration of $2.9 \times 3 = 8.7$ milliseconds, a vehicle moving at 40 km/h moves 32.2 millimeters, peak position barely moves in GCC results.

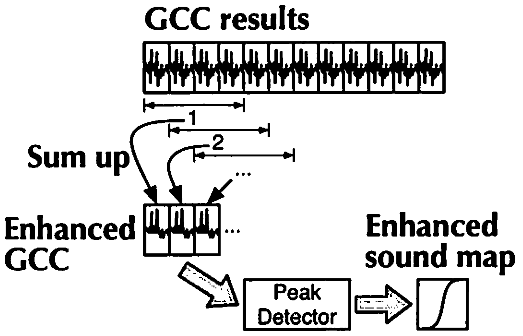

Figure 8 Overview of enhanced sound mapping

4.2 Enhanced Sound Mapping

Figure 8 depicts the overview of enhanced sound mapping. The enhanced sound mapping is quite simple. Multiple adjacent GCC results are added to retrieve enhanced GCC results. The enhanced GCC results are passed to peak detector deriving an enhanced sound map.

Here we analyze the enhanced sound mapping. Given two signals $x_i(n)$ and $x_j(n)$, the GCC with phase transform (GCC-PHAT) is defined as:

$$\hat{G}_{PHAT}(f) = \frac{X_i(f)[X_j(f)]^*}{|X_i(f)[X_j(f)]^*|}, \qquad (2)$$

where $X_i(f)$ and $X_j(f)$ are the Fourier transforms of the two signals and $[\ ]^*$ denotes the complex conjugate. The TDOA on two microphones is therefore estimated as:

$$\Delta t = \arg\max \hat{R}_{PHAT}(t), \qquad (3)$$

where $\hat{R}_{PHAT}(t)$ is the inverse Fourier transform of Eq. (2).

Consider two GCC-PHAT functions f_1 and f_2 with very small time difference. The sound signals that made f_1 are different from the sound signals made made f_2 because a vehicle cannot keep making the same noise. On the other hand, the vehicle is barely moved in such a short time and the location of the vehicle is almost unchanged, which results in the almost same TDOA Δt. Therefore waveforms of f_1 and f_2 is totally different except for the peaks at $t = \Delta t$. If several GCC-PHAT functions of this kind are added together, the peaks are enhanced and the effect of sound interference and of the noise are reduced.

Consider the situation in which there is only one sound source on a road,

$$\begin{cases} s_L(t) = s(t) + n_L(t) \\ s_R(t) = s(t - \Delta t) + n_R(t), \end{cases} \qquad (4)$$

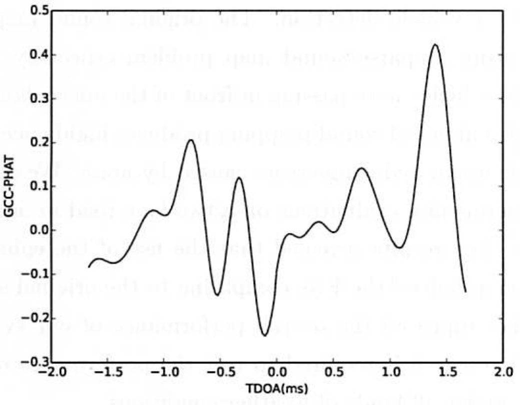

Figure 9 Example of enhanced GCC result

where $s_L(t)$ and $s_R(t)$ denote signals received by the microphones, $s(t)$ denotes the signal generated by the sound source, $n_L(t)$ and $n_R(t)$ denote the noise signals. The GCC-PHAT result of $s_L(t)$ and $s_R(t)$ is

$$\begin{aligned}\hat{G}_{PHAT}(f) &= e^{-2\pi j f \Delta t} + e^{j(\angle S(f) - \angle N_R(f))} \\ &+ e^{j(\angle N_L(f) - \angle S(f) + 2\pi f \Delta t)} + e^{j(\angle N_L(f) - \angle N_R(f))},\end{aligned} \tag{5}$$

where $S(f)$, $N_R(f)$, and $N_L(f)$ are the Fourier transforms of $s(t)$, $n_R(t)$, and $n_L(t)$, respectively. Equation (5) indicates that the time domain signal for GCC-PHAT result ideally consists of one impulse signal and three white noise signals, all of the same power. Due to the physical restriction, the power of the impulse signal usually spreads through time, as a consequence the impulse signal forms a peak in the time domain. Note that when the sum of noises exceed the peak value or jeopardize the peak, sound mapper makes a mistake by either recording a fake TDOA or recording nothing.

Figure 9 shows the example of an enhanced GCC result. We add four adjacent GCC results together to increase the signal to noise ratio (SNR). Preliminary experiments reveal that setting the time difference to 25 % of the length of the sequence is enough to make the last three terms in Eq. (5) change. Meanwhile, the first term barely changes because the vehicle cannot move far enough to make a considerable time shifting in TDOA. Comparing Figs. 6 and 9, we can confirm that the peak at $t = -0.6$ has been weakened sharply but the peak at $t = 1.4$ gets even stronger.

Assuming the noise to be Gaussian, after the summation, power of the noise increased by double while the power of the signal increased by four times. The SNR is therefore increased by double and the error rate of the sound mapper is dramatically reduced. Figure 10 depicts a comparison between the original and enhanced sound maps. When multiple sound sources were recorded by the system at the same time, the sparse original sound map loses so many details and S-curves on it become indistinct. The enhanced sound map remains explicit and detailed during the whole period.

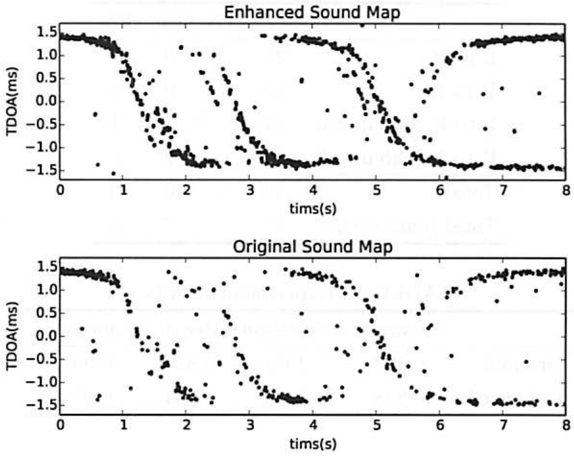

Figure 10 Original and enhanced sound maps

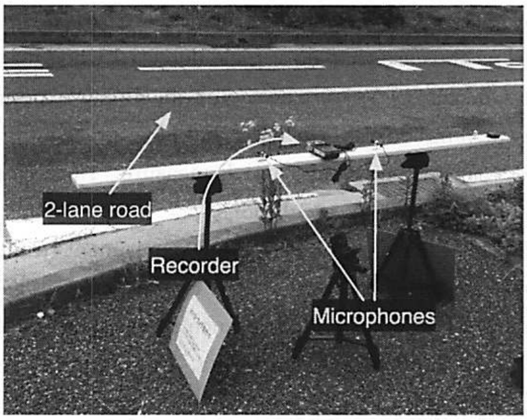

Figure 11 Experimental setup

5. Evaluation

We conducted experiments in our university to evaluate the basic performance of our vehicle count system. Figure 11 shows an experiment setup. The target road has two lanes, one lane for each direction. Two microphones were installed approximately two meters away from the road center. Distance between the two microphones was 50 centimeters, which is determined based on preliminary experiment results. We recorded vehicle sound for approximately 7 minutes using Sony PCM-D100 recorder with OLYMPUS ME30W microphones. The sound was recorded with a sampling frequency of 48 kHz and word length of 16 bits. We also recorded video monitoring the road which was used as ground truth data. Two kinds of sound map, original and enhanced, were generated using the same data and both of them have been evaluated under the same count algorithm.

Comparing the results derived by our vehicle count system with ground truth, we evaluated the number of true positives (TPs), false negatives (FNs), and false positives (FPs). TP, FN, and FP are defined as the case that a vehicle de-

TABLE 1 Experiment results

	TP	TN	FP	FN
L to R	28	–	0	9
R to L	22	–	0	2
L to R (enhanced)	34	–	5	3
R to L (enhanced)	23	–	2	1
Total	50	–	0	11
Total (enhanced)	57	–	7	4

TABLE 2 Experiment results

	Accuracy	Precision	Recall	F-measure
Original	0.820	1.00	0.820	0.901
Enhanced	0.838	0.891	0.934	0.912

tected when a vehicle passing, no vehicle detected when a vehicle passing, and a vehicle detected when no vehicle passing, respectively. We excluded true negatives (TNs), which is defined as the case that no vehicle detected when no vehicle passing, because TNs were not countable in our experiments. Using the numbers of TPs, FNs, and FPs, we also evaluated accuracy, precision, recall, and F-measure. Note that TNs were set to 0 in these calculations.

Tables 1 and 2 summarized our experiment results. For original method, the number of FPs was zero. The original method exhibited high tolerance to environmental noise such as wind and people chattering. However, 11 FNs have been made, which mainly happened when two vehicles were simultaneously coming from the opposite directions or consecutively coming from the same direction.

As for enhanced method, FNs have been reduced to 4, enhanced sound map performed splendidly in the situation where two or more cars coming at the same time. The FPs on the other hand, have increased by 7. This was partially because the road we took experiment on was located next to a parking lot. According to the video we took, more than 4 car activities had been made during the experiment in the parking lot. It turns out that the enhanced sound map recorded those activities and resulted in some FPs in our experiment. These FPs were actually TPs against the background activities.

Even with those FPs, the enhanced method still improved accuracy, recall, and F-measure of the system comparing to the original method. There was no mistake on detection of vehicle direction. Note that the count algorithm is developed specifically for the original sound map, hence there is still room for improvement.

6. Conclusion

In this paper, we propose an enhanced sound mapping scheme for sidewalk vehicle count system based on a microphone array. In our vehicle count system, a sound map, i.e., time difference of sound arrival on two microphones, is used for vehicle detection. The original sound mapping suffers from a sparse sound map problem especially when multiple vehicles were passing in front of the microphone array. The enhanced sound mapping produces highly accurate sound maps by reducing errors caused by noise. We carried out experimental evaluations on a two-lane road in our university. The results revealed that the use of the enhanced sound map halved the FNs comparing to the original sound map and improved the overall performance of our system. Further research is required to test the performance of the system under all kinds of weather conditions.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Numbers 15H05708 and 15K12021 as well as the Cooperative Research Project of the Research Institute of Electrical Communication, Tohoku University.

References

[1] Research and Markets, “Intelligent transportation system market by component, system (ATMS, ATIS, ITS-enable transportation pricing system, APTS, and CVO), application, and geography — analysis & forecast to 2015–2020,” Technical report, Research and Markets, July 2015.

[2] J.F. Forren and D. Jaarsma, “Traffic monitoring by tire noise,” Proc. IEEE Conf. Intelligent Transportation Systems (ITSC), pp.177–182, Nov. 1997.

[3] S. Chen, Z.P. Sun, and B. Bridge, “Automatic traffic monitoring by intelligent sound detection,” Proc. IEEE Conf. Intelligent Transportation Systems (ITSC), pp.171–176, Nov. 1997.

[4] S. Chen, Z. Sun, and B. Bridge, “Traffic monitoring using digital sound field mapping,” IEEE Trans. Veh. Technol., vol.50, no.6, pp.1582–1589, Nov. 2001.

[5] B. Barbagli, G. Manes, R. Facchini, and A. Manes, “Acoustic sensor network for vehicle traffic monitoring,” Proc. IEEE Int. Conf. Advances in Vehicular Systems (VEHICULAR), pp.1–6, June 2012.

[6] S. Ishida, K. Mimura, S. Liu, S. Tagashira, and A. Fukuda, “Design of simple vehicle counter using sidewalk microphones,” Proc. ITS EU Congress, pp.1–10, EU-TP0042, June 2016.

[7] S. Liu, S. Ishida, K. Mimura, S. Tagashira, and A. Fukuda, “Initial evaluation of acoustic vehicle count system utilizing dynamic time warping,” IEICE General Conf., pp.1–2, BS-3-44, March 2016.

[8] S. Ishida, S. Liu, K. Mimura, S. Tagashira, and A. Fukuda, “Design of acoustic vehicle count system using DTW,” Proc. ITS World Congress, Oct. 2016. (will appear).

[9] N. Buch, M. Cracknell, J. Orwell, and S.A. Velastin, “Vehicle localisation and classification in urban CCTV streams,” Proc. ITS World Congress, pp.1–8, Sept. 2009.

[10] A. Nurhadiyatna, B. Hardjono, A. Wibisono, W. Jatmiko, and P. Mursanto, “ITS information source: Vehicle speed measurement using camera as sensor,” Proc. Int. Conf. Advanced Computer Science and Information Systems (ICACSIS), pp.179–184, Dec. 2012.

[11] C.H. Knapp and G.C. Carter, “The generalized correlation method for estimation of time delay,” IEEE Trans. Acoust., Speech, Signal Process., vol.24, no.4, pp.320–327, Aug. 1976.