23rd ITS World Congress, Melbourne, Australia, 10–14 October 2016

Paper number AP-TP0678

Design of acoustic vehicle count system using DTW

Shigemi Ishida[1*], Song Liu[1], Kohei Mimura[1], Shigeaki Tagashira[2], Akira Fukuda[1]

1. ISEE, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, JAPAN
+81 92 802 3644, {ishida,ryu,mimura,fukuda}@f.ait.kyushu-u.ac.jp
2. Faculty of Informatics, Kansai University
2-1-1 Ryozenji-cho, Takatsuki-shi, OSAKA 569-1095, JAPAN
+81 72 690 2430, shige@res.kutc.kansai-u.ac.jp

Abstract

Vehicle counting is one of the fundamental tasks in the ITS (intelligent transportation system). Although automatic vehicle counters have been proposed to retrieve realtime traffic data, current automatic vehicle counters suffer from high deployment costs, resulting in limited number of deployments. In this paper, we present a vehicle counter using sidewalk microphones. Our vehicle counter relies on two sidewalk microphones and counts vehicles using a sound map, which is a time-difference map of vehicle sound on the two microphones. We developed a vehicle count algorithm using a sound map based on DTW (dynamic time warping). Experimental evaluations reveal that our vehicle count system successfully counted vehicles with a precision of 0.92.

Keywords:

Vehicle count, acoustic sensing, sound map.

1. Introduction

Increasing attention has been focused on the ITS (intelligent transportation system) due to the change of road transportation strategy. The main purpose of the ITS is to improve the safety, efficiency, dependability, and cost effectiveness of transportation systems. In the past decade, products such as car navigation systems have brought the ITS to our daily lives.

Vehicle counting is one of the fundamental tasks in the ITS. 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 data, automatic vehicle count systems have been deployed. However, the deployment of the automatic vehicle count system is 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; small vehicles such as motorbikes tend to be missed because of the small coverage of vehicle sensors. Although camera-based vehicle counters that are capable of motorbike counting are proposed, restrictions on camera location and angle make difficulties on practical deployment.

Therefore, this paper proposes a vehicle counter coming with low deployment cost. Our vehicle counter only relies on two microphones at a sidewalk. Because sound waves are diffracted over obstacles, we can deploy microphones in a low height configuration, which drastically reduces roadwork costs in terms of road closure as well as safety installation. Our vehicle counter can detect all types of vehicles as long as the vehicles generate sound.

There are several studies reporting a vehicle monitoring system using acoustic sensors [1–4]. 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.

We also have developed an automatic vehicle counter using a sound map [5]. In our previous study, we implemented a vehicle detector as a state machine analyzing a sound map. The state machine, however, lacks a sound map model of vehicle passage. Lacking of the sound map model makes performance optimization difficult. The state machine might also fall into an incorrect state, which critically degrades detection performance.

We therefore develop a new vehicle counting algorithm based on DTW (dynamic time warping) depending on a sound map of vehicle passage. By conducting experiments in our university, we demonstrate that our vehicle counter accurately counted vehicles with a precision of 0.92.

The remainder of this paper is organized as follows. Section 2 briefly looks through related works on vehicle counting. Section 3 describes our vehicle counter design and we conduct experiments in Section 4. In Section 5, we discuss limitations of our vehicle counter. Section 6 concludes the paper.

2. Related works

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

Loop coils and photoelectric tubes are categorized into the intrusive vehicle counters. These vehicle counters are required to be installed under the road surface. The installation and maintenance therefore require roadwork closing a target road section, which suffers from high costs. Loop coils and photoelectric tubes also have difficulties in motorbike detection due to their small coverage.

The non-intrusive vehicle counter is based on sensors such as laser, infrared, ultrasound, radar, or camera. 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 [6,7]. 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 [1–3]. 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 [4]. 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 [5]. However, the state machine detects vehicles based on sound delay changes without a model of vehicle passage. Vehicle passage draws a specific curve on a sound map. Lacking of the sound map model provides no assurance of detection performance. The state machine also have a chance to fall into an incorrect state, resulting in performance degradation.

3. System design

A. Overview

Figure 1 depicts an overview of our acoustic vehicle count system. Our vehicle count system consists of three components: a sound retriever, sound mapper, and vehicle counter. The sound retriever consists of two microphones followed by a low-pass filter (LPF). We install two microphones at a sidewalk of a road and record vehicle sound. The LPF removes high frequency environmental noise and the sound mapper calculates cross-correlation function between sounds on the two microphones to draw a sound map. Finally, we apply a template matching utilizing DTW (dynamic time warping) to the sound map and detect vehicles passing in front of the microphones.

In the following subsections give design details of the each component.

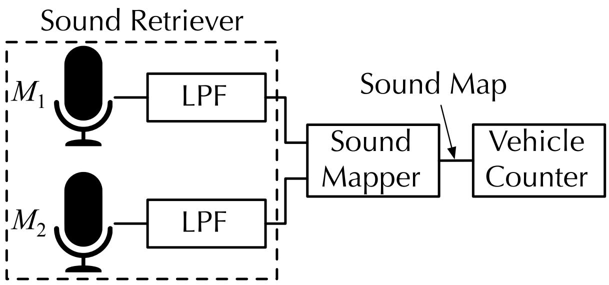

Figure 1 - Overview of acoustic vehicle count system

B. Sound retriever

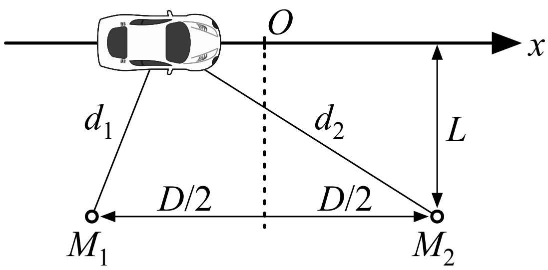

Figure 2 – Microphone setup

Figure 2 depicts a microphone setup. Two microphones M_1 and M_2 are installed at a sidewalk of a road parallel to the road. Distance D between the microphones and distance L between the road and microphones affect vehicle count performance. Sound arrival time difference on the two microphones is maximum when a vehicle is far away, i.e., at $x = \pm\infty$. The maximum sound arrival time difference $\Delta t_{\max}$ is calculated to be

$$\pm\Delta t_{max} = \pm\frac{D}{c}, \tag{1}$$

where c is the speed of sound in air. As D increases, a curve on a sound map increases or decreases more quickly. We therefore expect count accuracy improvement as D increases. Increasing D, however, degrades count accuracy because of increase in coverage, which tends to be affected by environmental noise. Distance L is in a similar situation. Distance D and L might be determined by preliminary experiments as well as physical restrictions.

To minimize the effect of environmental noise, our vehicle count system applies a low-pass filter (LPF) to vehicle sound signals. Majority of frequency component of sound signals generated by vehicle tires is less than 2.0 kHz [8]. The cut-off frequency of the LPF is therefore set to 2.5 kHz including a margin. Because the tire sound is generated by all types of vehicles, our vehicle count system detects all types of vehicles including normal cars, buses, trucks, and motorbikes.

C. Sound mapper

A sound map is time series data keeping track of sound arrival time difference between two microphones. As shown in Figure 2, we install two microphones M_1 and M_2 separated by D by a road at the distance of L. Sound signals generated by a vehicle reach the two microphones with different traveling distance. Let x be the location of a vehicle. The sound traveling distance d_1 and d_2 are calculated to be

$$d_1 = \sqrt{\left(x + \frac{D}{2}\right)^2 + L^2}, \tag{2}$$

$$d_2 = \sqrt{\left(x - \frac{D}{2}\right)^2 + L^2}. \tag{3}$$

We derive the time difference Δt of sound arrival between the two microphones using the speed c of sound in air:

$$\begin{aligned} \Delta t &= \frac{d_1 - d_2}{c} \\ &= \frac{1}{c}\left\{\sqrt{\left(x + \frac{D}{2}\right)^2 + L^2} - \sqrt{\left(x - \frac{D}{2}\right)^2 + L^2}\right\}. \end{aligned} \tag{4}$$

Using Equation (4), we can locate a vehicle from sound delay. Sound delay can be derived using a cross-correlation function $R(t)$. Let $s_1(t)$ and $s_2(t)$ be sound signals on the two microphones. The cross-correlation function $R(t)$ is defined as

$$R(t) = \int s_1(\tau)\, s_2(\tau + t) d\tau. \tag{5}$$

We assume that the two microphones receive same sound signals with time shifted by Δt, i.e., $s_1(t) = s_2(t + \Delta t)$. Because $R(t)$ is maximum at $t = \Delta t$, we can estimate the sound delay Δt by finding the maximum point of $R(t)$.

In our vehicle count system, the generalized cross-correlation (GCC) function [9], which is commonly used in acoustic source localization, is used to estimate sound delay. The count system divides sound signal sequences into small chunks and applies the GCC to the each chunk to estimate sound delay. The sound map is derived by plotting the sound delay of the each chunk.

Figure 3 shows a typical sound map. As a vehicle passes in front of the microphones, sound delay increases or decreases drawing an S-curve; direction of an S-curve depends on direction of the vehicle.

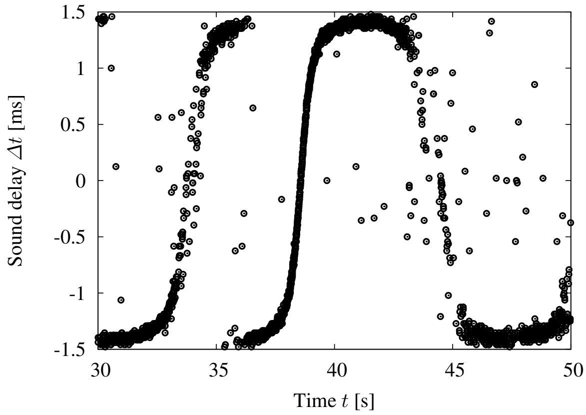

Figure 3 - Example of sound map

D. Vehicle counter

Vehicle passing drawn on a sound map is detected using DTW-based template matching. DTW is an algorithm to calculate the similarity of two sequences with different lengths. The length of an S-curve on a sound map is affected by a vehicle speed. DTW-based template matching is therefore robust to vehicle speed difference.

Prior to the template matching, we prepare templates. We manually create the templates from an S-curve on a sound map derived in a real environment. Because direction of an S-curve depends on a vehicle direction, we prepared the templates for each direction.

Figure 4 shows an overview of our vehicle counting using DTW. We first divide sound map sequences into fixed-length chunks. The first chunk (A) is then compared with a template using DTW. When the chunk (A) unmatches to the template, i.e., DTW distance is greater than a threshold, we combine the chunk (A) with the next chunk (B) and compare the combined chunk with the template. When the combined chunk (A)+(B) matches to the template, the following chunk (C) is compared with the template.

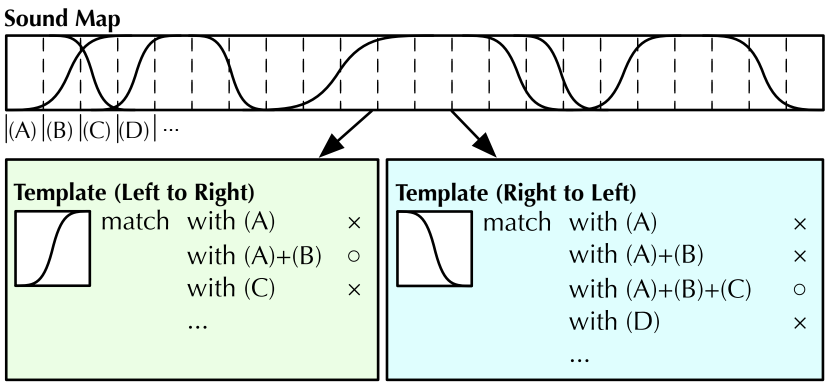

Figure 4 - Overview of vehicle count using DTW

Although DTW-based template matching successfully detects vehicles with different speed, simple shape of the template induces many false detections. As shown in Figure 5, an S-curve on a sound map is asymptotic to $\Delta t = \pm\Delta t_{max}$. The template partially matches successive vehicle passing.

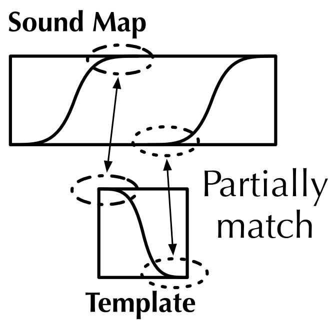

Figure 5 - False positive detection on template matching using DTW

To reduce false detections, we apply a simple filtering method using L2 norm. L2 norm over a sound map sequence Δt, is defined as

$$\|\Delta t\|_{L2} = \sqrt{\sum(\Delta t[k])^2}. \tag{6}$$

L2 norm over a sound map where vehicle is detected is calculated using Equation (6) and compared with a threshold. Detections whose L2 norm is greater than the threshold is determined as false detections.

A sound map sequence of a false detection includes an abrupt change of sound delay Δt from $-\Delta t_{max}$ to Δt_{max}, resulting in high L2 norm. The length of the sound map sequence is different for each vehicle. We normalize L2 norm with the length of the sequence.

4. Experimental evaluation

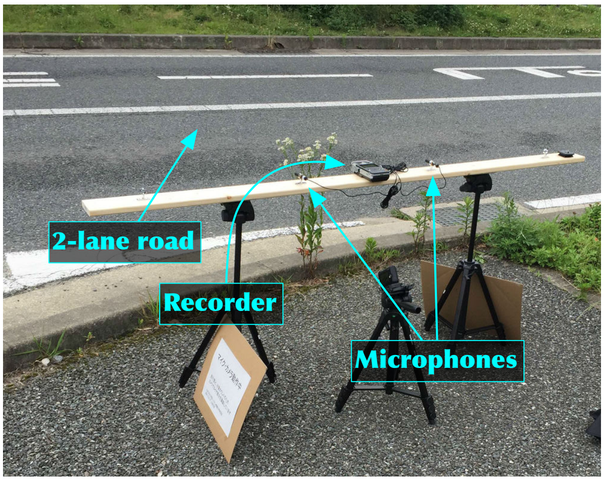

Figure 6 - Experiment setup

As an initial evaluation, we conducted experiments in our university evaluating the basic performance of our vehicle count system. Figure 6 shows an experiment setup. A target road has two lanes, one lane in 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 30 minutes using a 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.

Comparing the results derived by our vehicle count system with the video, we evaluated the numbers of true positives (TPs), false negatives (FNs), false positives (FPs), and true negatives (TNs). TP, FN, FP, and TN are defined as the case that a vehicle detected when a vehicle passing, no vehicle detected when a vehicle passing, a vehicle detected when no vehicle passing, and no vehicle detected when no vehicle passing, respectively. TNs are counted as the number of inter-vehicle intervals. Using the numbers of TPs, FNs, and FPs, we also evaluated an accuracy, precision, recall, and F-measure.

Table 1 - Experiment results

	TP	FN	FP	TN
Left to right	32	7	4	–
Right to left	63	14	4	–
Total	95	21	8	109

Table 1 summarizes the number of TPs, FNs, FPs, and TNs. Note that TN is only defined for total result that is the sum of both vehicle directions. Accuracy, precision, recall, and F-measure were calculated to be 0.88, 0.92, 0.82, and 0.87, respectively. Our vehicle count system successfully counted vehicles with a precision of 0.92.

5. Discussion

A. Template for detection

Big limitation in our vehicle counting scheme is that we need matching templates for both vehicle directions prior to vehicle detection. As described in Section 3.C, we create templates from a sound map derived experimentally. We first record sound signals at a target location and draw a sound map. Template sequences are manually extracted in a part of the sound map.

We can also use ideal templates, which are calculated using Equation (4). Equation (4) instantly gives an ideal sound map sequences using distance D between two microphones and distance L between road and the microphones. Although length of the templates theoretically has no effect on template matching using DTW, a short template abruptly varies as a function of time and suffers from high quantization error. We calculate the minimum length of the templates from speed limit of the target road. The minimum length should be set to a time duration such that a vehicle at the speed limit passes in front of microphones.

B. Multiple vehicle detection

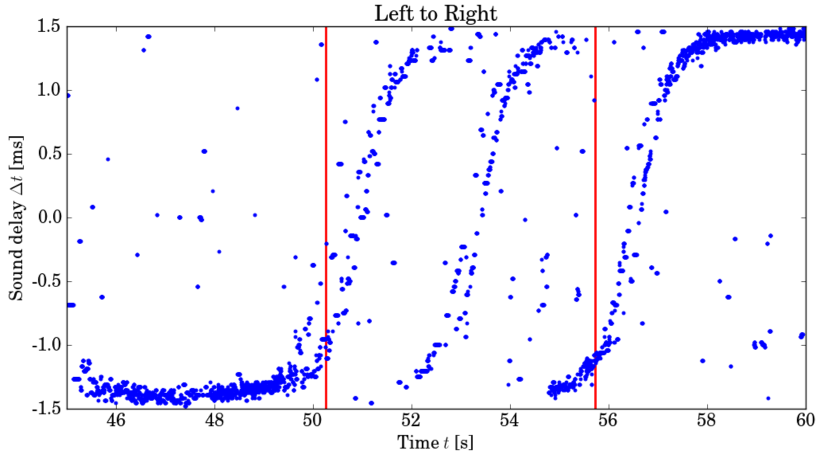

Figure 7 - Multiple vehicle detection example

The vehicle counter tends to detect one vehicle when multiple vehicles simultaneously or successively pass in front of microphones. Figure 7 shows an example of vehicle detection on a sound map when multiple vehicles were successively passing. A red vertical line indicates that a vehicle is detected. Although there are three S-curves drawn, two of the S-curves were detected by our vehicle counter. The vehicle counter performs template matching using DTW on divided chunks of a sound map. DTW-based template matching gives similarity between a chunk and a template sequence, telling us not the number of vehicles but the existence of vehicles. Our vehicle counter is applicable to roads with one lane in each direction.

Multiple vehicles simultaneously passing in different directions are detected because we use different templates for vehicles in different directions. However, detection accuracy might be degraded because of microphone dynamic range limitation. Sound signals from multiple vehicles overlap each other resulting in loud sound signal. If the amplitude of the overlapped sound signal is bigger than a microphone dynamic range, sound signals are distorted, which degrades detection performance.

6. Conclusion

In this paper, we presented an acoustic vehicle count system based on DTW (dynamic time warping) based template matching. Our vehicle counter relies on two microphones installed at a sidewalk and draws a sound map. We then apply template matching using DTW to count vehicles. We conducted experimental evaluations and demonstrated that our vehicle counter accurately counted vehicles with a precision of 0.92.

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. J. Forren, and D. Jaarsma (1997), Traffic monitoring by tire noise, in Proceedings IEEE Conf. Intelligent Transportation Systems (ITSC). pp.177–182.
2. S. Chen, Z.Sun, and B. Bridge (1997). Automatic traffic monitoring by intelligent sound detection, in Proceedings *IEEE Conf. Intelligent Transportation Systems (ITSC)*, pp.171–176.
3. S. Chen, Z. Sun, and B. Bridge (2001). Traffic monitoring using digital sound field mapping, *IEEE Transactions on Vehicular Technology*, vol.50, no.6, pp.1582–1589.
4. B. Barbagli, G. Manes, R. Facchini, et al. (2012). Acoustic sensor network for vehicle traffic monitoring, in Proceedings *1st International Conference on Advances in Vehicular Systems, Technologies and Applications*, pp.1–6.
5. S. Ishida, K. Mimura, S. Liu, et al. (2016). Design of simple vehicle counter using sidewalk microphones, in Proceedings *ITS European Congress*, pp.1–10.
6. N. Buch , M Cracknell, J. Orwell, et al. (2009). Vehicle localisation and classification in urban CCTV streams, in Proceedings *ITS World Congress*, pp.1–8.
7. A. Nurhadiyatna, B. Harjono, A. Wibisono, et al. (2012). ITS information source: Vehicle speed measurement using camera as sensor, in Proceedings *Int'l Conf. Advanced Computer Science and Information Systems (ICACSIS)*, pp.179–184.
8. H. Wu, M. Siegel, and P. Khosla (1998). Vehicle sound signature recognition by frequency vector principal component analysis, in Proceedings *IEEE Instrumentation and Measurement Technology Conf. (IMTC)*, vol.1, pp.429–434.
9. C. Knapp, and G. Carter (1976). The generalized correlation method for estimation of time delay, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.29, no.4, pp.320–327.