Initial Design of Two-Stage Acoustic Vehicle Detection System for High Traffic Roads

Masato Uchino*, Billy Dawton*, Yuki Hori*, Shigemi Ishida*,
Shigeaki Tagashira†, Yutaka Arakawa*, and Akira Fukuda*
*Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
Email: {uchino, bdawton, hori}@f.ait.kyushu-u.ac.jp, {ishida, arakawa, fukuda}@ait.kyushu-u.ac.jp
†Faculty of Informatics, Kansai University, Japan
Email: shige@res.kutc.kansai-u.ac.jp

***Abstract*—As the adoption of Intelligent Transport Systems (ITS) grows worldwide, so does the need for lost-cost, fast-deployment vehicle detection systems. SAVeD is a low-cost acoustic detection system developed by the authors which works by fitting a curve indicating vehicle passage to a sound map depicting the difference in arrival time of a passing vehicle's sound at two microphones installed on the roadside. This paper expands on the SAVeD method by proposing a Two-Stage Acoustic Vehicle Detection System for use in high-traffic environments, where multiple simultaneously and successively passing vehicles cause interference in the detection process. To solve this problem, the sound map fitting process is divided into two stages: the detection range is narrowed based on information estimated during the Pre-Fitting stage, and neighborhood point extraction is performed during the Post-Fitting stage to improve vehicle detection accuracy. Initial evaluation performed on a four-lane, two-direction road showed a vehicle detection F-measure of 0.63, a 12-point increase over SAVeD.**

***Index Terms*—Intelligent Transport Systems (ITS), vehicle detection, acoustic sensing, fitting.**

I. Introduction

With the increasing development of road traffic in recent years, the importance of Intelligent Transport Systems (ITS) is greater than ever. The main purpose of ITS is to provide functionalities that improve the safety, efficiency and convenience of road users. Notable examples include car navigation, automatic driving and various driving support systems, all of which need to take into account traffic congestion.

In particular, vehicle detection is indispensable for monitoring and controlling road traffic. Existing vehicle detectors, however, need to be installed under or above a road, which results in high installation and maintenance costs. In an effort to perform low-cost, real-time vehicle detection the authors have been developing an acoustic vehicle detection method [1–3]. This system uses two microphones installed by the roadside to draw a *sound map*, the map of the time difference of arrival (TDOA) of the sounds emitted by vehicles. As the vehicles pass in front of the microphone pair, S-curves are drawn on the sound map enabling the detection of vehicles through the use of curve fitting techniques. System evaluations performed in a variety of traffic and weather conditions on two-way two-lane roads showed high performance when detecting vehicles.

However, in heavy traffic situations the system's detection performance drops significantly. The high number of vehicles passing simultaneously and successively in front of the microphones means that the S-curves of multiple vehicles are mixed within the detection range, causing interference and ultimately stopping the system from detecting passing vehicles.

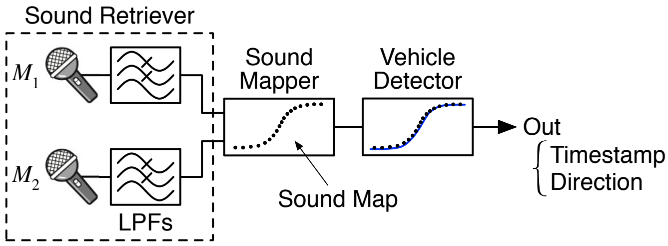


Fig. 1. Overview of SAVeD [1]

In order to tackle the heavy traffic problem, we present a Two-Stage Acoustic Vehicle Detection System which separates the S-curve fitting process into two stages: *Pre-Fitting* and *Post-Fitting*. A vehicle's speed and passing time are estimated during the Pre-Fitting process, and based on this information, the second-stage of Post-Fitting is performed only on the specified time range. After Post-Fitting, in order to reduce the influence of other vehicles, we extract only the sound map points close to the fitting curve.

Initial evaluation of our proposed vehicle detection method shows an F-measure of 0.63, which is an 12-point improvement over our existing method.

The rest of the paper is organized as follows: Section II is a review of our existing acoustic vehicle detector setup and the heavy traffic problem, Section III describes the design of the proposed Two-Stage Acoustic Vehicle Detection System, and the system's performance is evaluated in Section IV. Section V briefly examines existing vehicle detection systems, including an acoustic vehicle detector presented in our previous paper. Finally, Section VI concludes the paper.

II. SAVeD: Acoustic Vehicle Detector

A. Overview of SAVeD

Figure 1 shows an overview of SAVeD, the Sequential Acoustic Vehicle Detector presented in our previous work [1]. SAVeD consists of three parts: the sound retriever, the sound mapper, and the vehicle detector.

The sound retriever is made up of two microphones and two corresponding low pass filters (LPFs). The microphones

doi: 10.1109/PerComWorkshops48775.2020.9156248

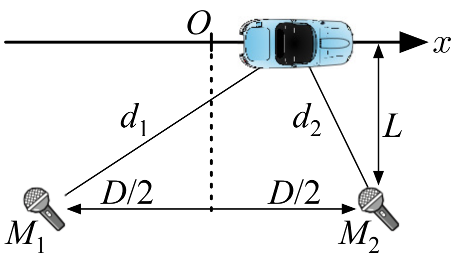


Fig. 2. Microphone setup [1]

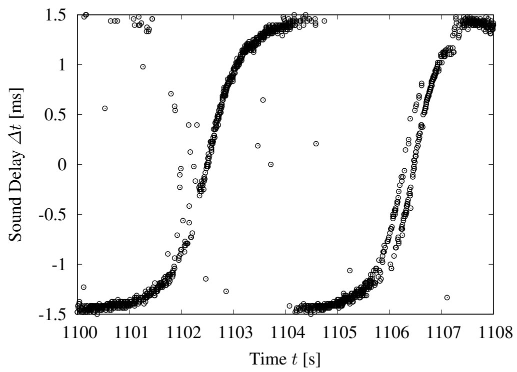


Fig. 3. Typical example of sound map

are set up at the side of the road to record the sounds emitted by passing vehicles, with the LPFs working to reduce the influence of unwanted high-frequency noise. The sound emitted by vehicle tires is sub-2 kHz [4], so we set the LPF's cut-off frequency to 2.5 kHz accordingly.

The sound mapper draws a *sound map*, which is a map of the time difference of arrival (TDOA) of the sound signals captured by the two microphones. When a vehicle passes in front of the microphones at a constant speed v, an S-curve is drawn on the sound map. In the microphone setup shown on Fig. 2, the sound delay Δt can be calculated from the Eq. (1) using sound travelling distances d_1 and d_2:

$$\begin{aligned}\Delta t =& \frac{d_1 - d_2}{c} \\ =& \frac{1}{c}\left\{\sqrt{\left[v(t-t_0)+\frac{D}{2}\right]^2 + L^2}\right. \\ &\left. -\sqrt{\left[v(t-t_0)-\frac{D}{2}\right]^2 + L^2}\right\}, \end{aligned} \tag{1}$$

where t_0 denotes the initial vehicle passing time and c the speed of sound in air. From Eq. (1), we can see that if v is of constant speed, the S-curve appears linearly on the sound map.

Figure 3 shows a typical sound map, with two vehicles moving from left to right. The direction of the S-curve depends on which direction a passing vehicle is travelling in, i.e., the sign of v.

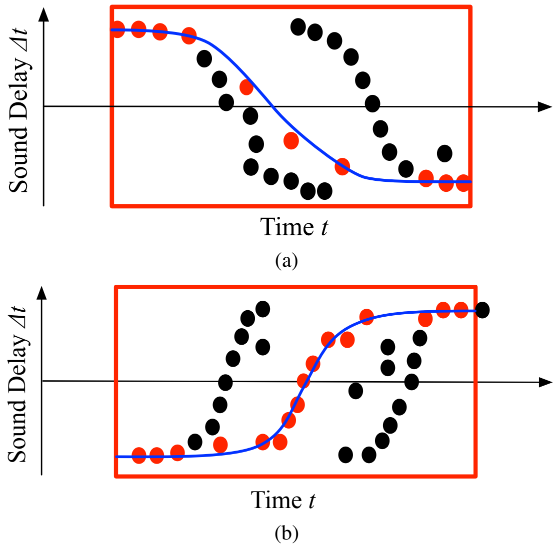


Fig. 4. Incorrect detections due to (a) the un-erased problem, (b) the over-erased problem

The vehicle detector detects the S-curve on the sound map using the Random Sample Consensus (RANSAC) algorithm [5]. We estimate the vehicle speed v and the passing vehicle time t_0 by fitting Eq. (1) to high-likelihood sound map points. The detector sequentially detects vehicles by dividing the sound map into set-width windows and successively fitting each window.

The nature of the RANSAC process means that the detector is still attempting to fit curves to points on the sound map even in the absence of passing vehicles. To prevent this, we apply a threshold for the sum of distances between the fitted S-curve and the sound map points. Once a vehicle is detected, the sound map points corresponding to that vehicle are removed in order to reduce their influence on subsequent vehicles.

B. Heavy Traffic Problem

On heavy traffic roads, there is a marked increase in the number of vehicles passing simultaneously and successively in front of the microphones. Our vehicle detector is therefore faced with two problems when removing sound map points: the *un-erased* problem and the *over-erased* problem, both of which have a detrimental effect on detection accuracy.

1) Un-erased Problem: In this situation, sound map points corresponding to a detected vehicle are not completely erased, which causes them to interfere in subsequent vehicle detections. Figure 4(a) shows an example un-erased situation. The black points are the sound map points, the red frame is the detection window, the blue line is the fitted S-curve, and the red points are the sound map points to be erased. When multiple vehicles pass successively on multiple lanes, the S-curve is wrongly fitted across the S-curve of multiple vehicles as shown in Fig. 4(a). This error during the fitting process has the effect of decreasing detection accuracy.

2) Over-erased Problem: In this situation, sound map points corresponding to vehicles other than the currently passing vehicles are incorrectly erased leading to detection

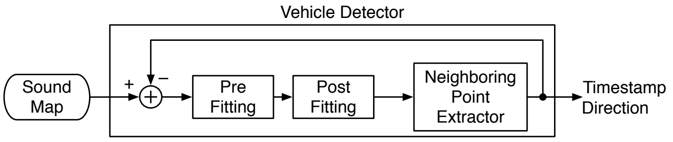


Fig. 5. Overview of Two-Stage Acoustic Vehicle Detection System

errors. Figure 4(b) shows an example over-erased situation. In this case, the fitted S-curve is estimated from sound map points belonging to the vehicles before and after the passing vehicle in addition to those of the vehicle itself. As a result, the sound map points corresponding to the front and rear vehicles are removed, again negatively affecting subsequent vehicle detection accuracy.

III. Two-Stage Acoustic Vehicle Detection System

A. Overview

In order to improve detection performance on heavy traffic roads, we put forward our Two-Stage Acoustic Vehicle Detection System. Figure 5 shows an overview of the proposed system, consisting of three parts: a Pre-Fitting block, Post-Fitting block, and Neighboring Point Extractor.

The Pre-Fitting block detects a vehicle by performing S-curve fitting using the set-width window and RANSAC as shown in Section II-A. When a vehicle is detected, the vehicle speed and passing time are estimated and sent to the Post-Fitting block. The Post-Fitting block resets the detection window based on the vehicle speed and the passing time, and the S-curve fitting is performed again using only sound map points in this window. The Neighboring Point Extractor extracts sound map points close to the passing time among the S-curve in the Post-Fitting, and checks if a vehicle is passing. When the vehicle is detected, we remove sound map points corresponding to the detected vehicles and repeat the process.

The following subsections detail the operation of the Post-Fitting block and the Neighboring Point Extractor.

B. Post-Fitting Block

Figure 6 shows an overview of the Post-Fitting block. The Post-Fitting block dynamically sets the RANSAC window size based on the estimated speed v and the time t_0 obtained from the Pre-Fitting block and repeats the S-curve fitting process using the sound map points in the dynamic window. Using the redefined S-curve temporarily estimated in the Pre-Fitting block results in a more accurate S-curve fitting. If the results of Post-Fitting and Pre-Fitting are significantly different, it is regarded as a detection failure, and we continue the detection process in the following window.

C. Neighboring Point Extractor

The Neighboring Point Extractor detects vehicles using the sound map points close to the estimated passing time. The closer the distance between a passing vehicle and the microphones, the louder the received vehicle sounds are. When a vehicle is passing just in front of the microphones, more

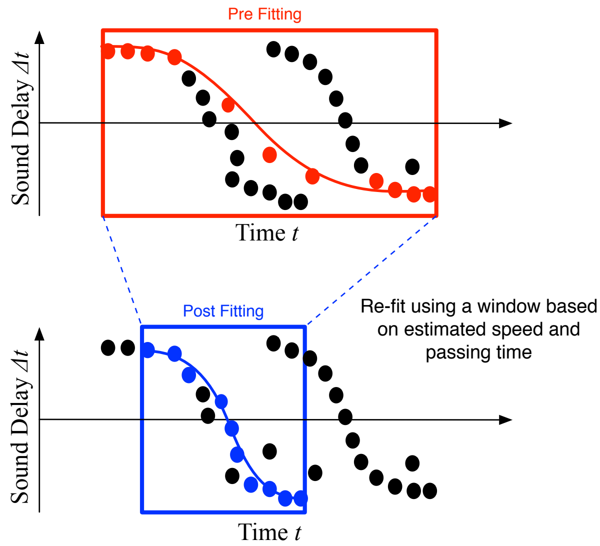


Fig. 6. Overview of the Post-Fitting process

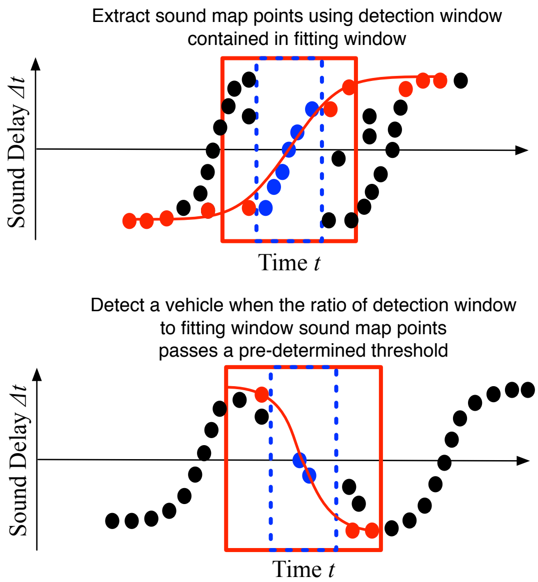
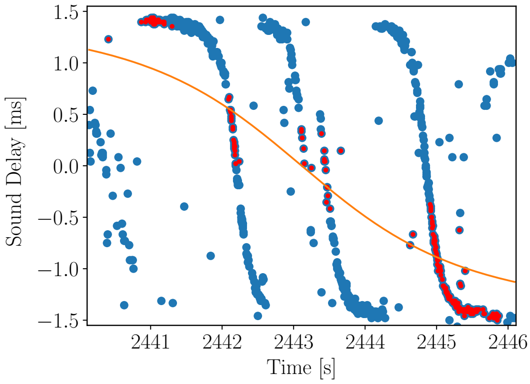


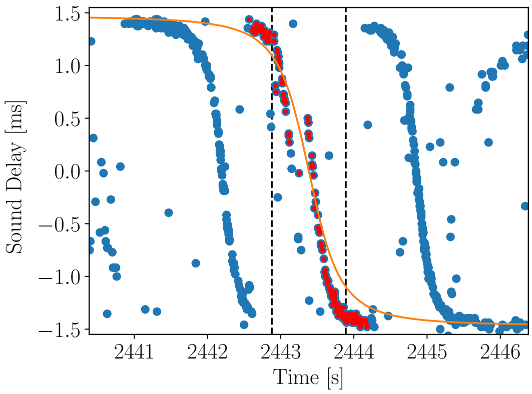
Fig. 7. Overview of the Neighboring Point Extractor. The red square and blue dotted square represent a fitting window and a detection window, respectively.

points are drawn on the sound map. We therefore extract the sound map points near the estimated passing time from the estimated S-curve points.

Figure 7 shows an overview of the Neighboring Point Extractor. The total number of sound map points corresponding to each vehicle is reduced when multiple vehicles pass in front of the microphones. We therefore use the relative number of sound map points in our window instead. As shown in Fig. 7, we set a red detection window and a blue fitting window. A fitting window is the window defined in the Post-Fitting block. A detection window is defined in the Neighboring Point Extractor to extract sound map points near the estimated passing time. The Neighboring Point Extractor dynamically sets the detection window size, which is smaller than the fitting



(a)



(b)

Fig. 8. Example sound maps for sequentially passing vehicles: (a) un-erased and over-erased problems present in the SAVeD, (b) effect of proposed Two-Stage Acoustic Vehicle Detection System. The blue points are the sound map points, the orange line is the S-curve estimated by fitting, and the red points are the sound map points that are close to the S-curve and erased.

window size, based on the estimated speed v. Our system detects a vehicle when the ratio of the points corresponding to the S-curve in the detection window relative to those in the fitting window is greater than a pre-determined threshold. In this paper, we experimentally set 0.4 as the threshold. When a vehicle is detected, we remove the points corresponding to the S-curve of the detected vehicle as described in Section II-A.

Figure 8 shows example sound maps of sequentially passing vehicles. The blue points are the sound map points, the orange line is the S-curve estimated by fitting, and the red points are the sound map points that are close to the S-curve and erased. As shown in Fig. 8(a), SAVeD estimated a "fake" S-curve between the real S-curves of three sequentially passing vehicles, resulting in an un-erased situation. However, when used in the same situation, the Two-Stage Acoustic Vehicle Detection System correctly estimated the appropriate S-curve and only removed points corresponding to the estimated vehicle. The dotted frame in Fig. 8(b) is the detection window from the Neighboring Point Extractor and was set correctly depending on the estimated speed in the Post-Fitting block. This also enabled us to confirm that our proposed system is able to correctly detect the front and rear vehicles.

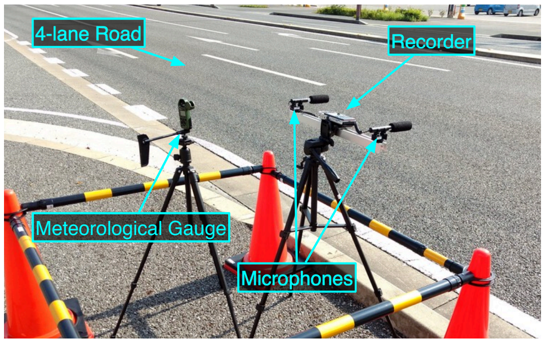


Fig. 9. Experiment setup

IV. Evaluation

We conducted a series of experiments to evaluate the initial performance of our proposed Two-Stage Acoustic Vehicle Detection System.

A. Experimental Setup

Figure 9 shows our experimental setup. The target road has four lanes, two lanes in each direction. Two microphones are installed approximately 2 meters away from the road's side line. As in our previous paper [1], we set the two microphones 50 centimeters apart and 1 meter from the ground. We recorded the sound of passing vehicles for approximately 60 minutes using a Sony PCM-D100 sound recorder with AZDEN SGM-990 microphones. The sound was recorded at a sampling rate of 48 kHz and a word length of 16 bits. Video footage shot from a position overlooking the road was used as ground truth data. The weather conditions during the experiment were measured with a Nielsen-Kellerman Kestrel 5500 anemometer. The average wind speed during the experiment was 1.26 m/s. We gauged our system's performance using the F-measure, a commonly used metric in classifier evaluations. We set a passing time error margin θ_t, which is the difference between the estimated passing time and the actual passing time. When our system detects a vehicle within θ_t, it is counted as a true positive (TP). When our system detects a vehicle outside of θ_t, it is counted as a false positive (FP). In the event that the system fails to detect a vehicle that is passing within θ_t, it is defined as a false negative (FN). We counted the number of true positives (TPs), false positives (FPs), and false negatives (FNs) from which we calculated precision, recall, and F-measure defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{F}_{\text{measure}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

Precision is the ratio of TP passing vehicles to detected passing vehicles, and provides information on how "correct" the classifier's detections are. Recall is the ratio of TP passing vehicles to the actual passing vehicles, which corresponding to the classifier's vehicle detection rate. F-measure is the

TABLE I
VEHICLE DETECTION PERFORMANCE RESULTS

(a) Two-Stage Acoustic Vehicle Detection System

	Left to Right	Right to Left	Total
TP	142	183	325
FN	156	128	284
FP	27	75	102
Precision	0.84	0.71	0.76
Recall	0.48	0.59	0.53
F-measure	0.61	0.64	0.63

(b) SAVeD

	Left to Right	Right to Left	Total
TP	133	132	265
FN	165	179	344
FP	81	80	161
Precision	0.62	0.62	0.62
Recall	0.45	0.42	0.44
F-measure	0.52	0.50	0.51

harmonic mean of precision and recall, which provides a comprehensive evaluation of the classifier.

B. Detection Performance

Table I shows the system performance, i.e., the number of TP, FN, and FP detections as well as the calculated precision, recall, and F-measure for $\theta_t = 1.5s$.

- The precision of the proposed Two-Stage Acoustic Vehicle Detection System is 0.76, 14 points higher than SAVeD. By resetting the detection window in the Post-Fitting block and detecting vehicles using only sound map points in the neighborhood of the estimated vehicle passing time, the amount of FP detections was reduced.
- The recall of 0.53 is a 9-point improvement compared to previous work. This is due to the system's ability to make use of all the sound map points corresponding to a given vehicle.
- The F-measure of 0.63 is 12 points higher than SAVeD. The relatively low F-measures exhibited by both the Two-Stage Acoustic Vehicle Detection System and SAVeD are due to their low recall values compared to their precision values. As the number of simultaneously passing vehicles increases, the number of sound map points corresponding to a single vehicle decreases because only one point is drawn on the sound map at each time step. As a result, the S-curve becomes sparse, increasing the probability of a FN detection.
- The Two-Stage Acoustic Vehicle Detection System decreased the number of left-to-right FP detections by 1/3 and increased the precision by 22 points. On the other hand, the decrease in right-to-left FP detections was not as significant. This is due to the direction in which cars travel with respect to the side of the road: a vehicle passing from right to left passes closer to the microphones. This has a direct effect on the width of the S-curve, the smaller value of L in Eq. (1) shrinking the S-curve accordingly. Therefore, the estimated speeds of these vehicles obtained during the Pre-Fitting process differ considerably from the actual speed, drastically reducing the effectiveness of the Post-Fitting process.

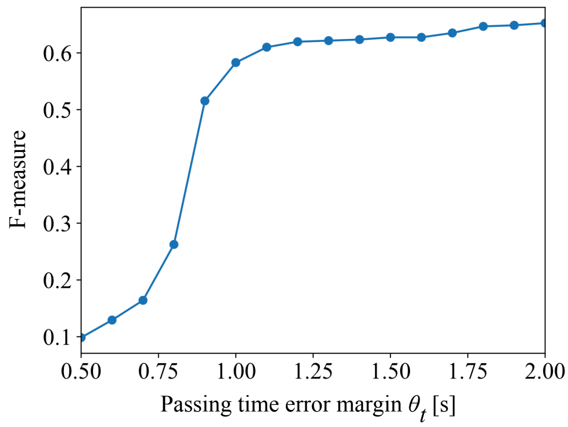


Fig. 10. Proposed system F-measure as a function of passing time error margin θ_t

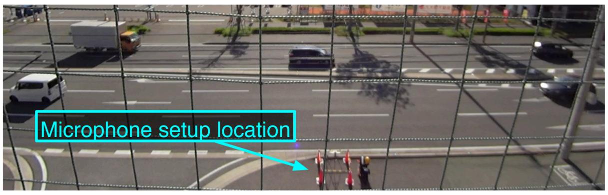


Fig. 11. System monitoring viewpoint

The above results show that our Two-Stage Acoustic Vehicle Detection System is an improvement over the SAVeD in terms of vehicle detection performance.

C. Passing Time Error

The detection performance of our system depends strongly on the estimated passing time used in the Post-Fitting block. As described in Section IV-A, the smaller the passing time error margin $\theta_t s$, the higher the number of FP detections. It is therefore important to examine the influence of the passing time error margin on detection performance.

Figure 10 shows the F-measures for θ_t between 0.5 s and 2.0 s. The F-measure starts to increase at about $\theta_t = 0.6$ s and stabilizes around $\theta_t = 1.2$ s. The optimum value of θ_t corresponds to the point at which the F-measure just begins to stabilize, as a too large value of θ_t will cause the system to incorrectly detect passing vehicles.

In our system, we find that discrepancies between estimated and actual passing times are usually caused by the latter rather than the former. The monitoring environment is shown in Fig. 11. The microphones are surrounded by red cones and located in the center of the image. Due to filming location restrictions, we were unable to place our video camera directly above the microphones, we therefore visually labeled the actual passing time from offset video footage, leading to detection time errors in ground truth data. In addition, due to equipment restrictions the video and audio feeds were not automatically synchronized. We believe that the high detection time error present in the system is mainly due to these two factors.

V. Related Work

To the best of our knowledge, this paper is the first attempt to tackle vehicle detection on multi-lane roads in the field of acoustic vehicle sensing. In this section, we briefly examine currently existing non-acoustic and acoustic vehicle detectors.

A. Current Vehicle Detectors

Current vehicle detectors can be broadly classified into two types: intrusive and non-intrusive.

Intrusive vehicle detectors are a type of detector where the actual sensor is embedded into the road. Loop coils and photoelectric tubes are examples of such systems. Intrusive detectors have a long lifespan and do not need to be regularly maintained; however, they necessitate substantial roadworks to be implemented which results in high installation costs and long installation times. Additionally, due to their small sensing range, intrusive detectors struggle to detect motorbikes.

Non-intrusive vehicle detectors are a type of detector installed either above or next to a road. Examples include laser, infrared, radar, and video camera-based systems. Non-intrusive detectors have a wider range of detection per individual sensor; however, they can incur significant installation costs when deployed above a road, and their performance is affected by weather conditions. To reduce installation and maintenance costs, camera-based vehicle detector using already installed CCTV cameras have been proposed [6, 7]. CCTV, however, is only available in limited areas, especially in city areas with heavy traffic, and its primary purpose is security surveillance, not vehicle detection. This makes CCTV particularly vulnerable to weather conditions, as the installation location and camera angles are optimized for surveillance purposes, not continuous monitoring of traffic from positions above or next to the road.

B. Acoustic Vehicle Detectors

Acoustic detection systems have the advantage of low installation and maintenance costs. By using two microphones to capture acoustic signals generated from vehicle tires, it is possible to detect vehicles of any type (car, bus, motorbikes, etc) over multiple lanes. Several studies have examined the idea of acoustic vehicle detection [8–11].

In our previous work [1], vehicle detection is performed using a sound map. A sound map is a map of time difference of arrival (TDOA) of sound signals on a pair of microphones, with the sound map points themselves corresponding to peaks of a generalized cross-correlation (GCC) function between the two sound signals. The GCC is commonly used in a field of acoustic source localization [12]. This detector detects vehicles by fitting a model to a sound map using a Random Sample Consensus (RANSAC) algorithm [5]. We experimentally demonstrated that this system successfully detected vehicles as well as their direction of travel with an F-measure of 0.83.

We have presented acoustic detectors optimized for adverse weather conditions and for low-power consumption [2, 3, 13]. However, the detection performance of these systems is still greatly reduced in environments with a high number of passing vehicles, such as multi-lane roads.

VI. Conclusion

In this paper, we presented a Two-Stage Acoustic Vehicle Detection System for use on heavy traffic roads. A pair of microphones installed on the side of the road detects passing vehicles by fitting an S-curve to points on a sound map. The main idea put forward is the two-stage detection process, which enables the system to distinguish and detect individual vehicles from a group of simultaneously or successively passing vehicles. Initial evaluation performed on a four-lane road shows that our Two-Stage Acoustic Vehicle Detection System is able to detect vehicles with an F-measure of 0.63, 12 points higher than the SAVeD presented in previous work.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Numbers JP15H05708, JP17H01741, and JP18K18041 as well as the Cooperative Research Project of RIEC, Tohoku University.

References

[1] S. Ishida, J. Kajimura, M. Uchino *et al.*, “SAVeD: Acoustic vehicle detector with speed estimation capable of sequential vehicle detection,” in *Proc. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 906–912.

[2] M. Uchino, S. Ishida, K. Kubo *et al.*, “Initial design of acoustic vehicle detector with wind noise suppressor,” in *Proc. Int. Workshop on Pervasive Computing for Vehicular Systems (PerVehicle)*, Mar. 2019, pp. 814–819.

[3] S. Ishida, M. Uchino, C. Li *et al.*, “Design of acoustic vehicle detector with steady-noise suppression,” in *Proc. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2019.

[4] H. Wu, M. Siegel, and P. Khosla, “Vehicle sound signature recognition by frequency vector principal component analysis,” in *Proc. IEEE Instrumentation and Measurement Technology Conf. (IMTC)*, vol. 1, May 1998, pp. 429–434.

[5] M. A. Fischler and R. C. Bolles, “Random sample censensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[6] N. Buch, M. Cracknell, J. Orwell *et al.*, “Vehicle localisation and classification in urban CCTV streams,” in *Proc. ITS World Congress*, Sep. 2009, pp. 1–8.

[7] A. Nurhadiyatna, B. Hardjono, A. Wibisono *et al.*, “ITS information source: Vehicle speed measurement using camera as sensor,” in *Proc. Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS)*, Dec. 2012, pp. 179–184.

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

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

[10] 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.

[11] B. Barbagli, G. Manes, R. Facchini *et al.*, “Acoustic sensor network for vehicle traffic monitoring,” in *Proc. IEEE Int. Conf. on Advances in Vehicular Systems (VEHICULAR)*, Jun. 2012, pp. 1–6.

[12] 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.

[13] K. Kubo, C. Li, S. Ishida *et al.*, “Design of ultra low power vehicle detector utilizing discrete wavelet transform,” in *Proc. ITS AP Forum*, May 2018, pp. 1052–1063.