SAVeD: Acoustic Vehicle Detector with Speed Estimation capable of Sequential Vehicle Detection

Shigemi Ishida[1], Jumpei Kajimura[1], Masato Uchino[1], Shigeaki Tagashira[2], and Akira Fukuda[1]

***Abstract*— In the ITS (intelligent transportation system), vehicle detection is one of the core technologies. We are developing an acoustic vehicle detector that detects vehicles using a sound map, which is a map of sound arrival time difference on two microphones. We developed vehicle detection algorithms based on state machine and DTW (dynamic time warping) to detect S-curves on a sound map drawn by passing vehicles. However, the detection algorithms often fail to detect simultaneous and sequential passing vehicles.**

This paper presents *SAVeD*, a sequential acoustic vehicle detector. The SAVeD fits an S-curve model to sound map points using a RANSAC (random sample consensus) robust estimation method to detect each vehicle. The SAVeD then removes sound map points corresponding to the detected vehicle and continues vehicle detection process for the following vehicles. Experimental evaluations demonstrated that the SAVeD improves detection accuracy by more than 10 points compared to the state-machine based algorithm.

I. Introduction

The past decade has seen the rapid development of ITS (intelligent transportation system). The main purpose of the ITS is to improve the safety, efficiency, dependability, and cost effectiveness of transportation systems. Many cars come with car navigation, cruise control, and anti-collision braking systems, which implies that the ITS is becoming prevalent in our lives today.

In the ITS, vehicle sensing is one of the core technologies. In Japan, the deployment of vehicle sensors is limited to high traffic roads and freeways because of high deployment and maintenance costs. Although vehicle sensors have installed on many roads in some countries, these sensors are becoming old and are to be replaced in the coming decades. Some literature reported low-cost vehicle detector based on CCTVs [1, 2] and probe-car data [3–9]. These technologies are applicable to high traffic roads.

We are developing an acoustic vehicle detector coming with low deployment and maintenance costs as another choice of low cost vehicle sensing [10, 11]. We use two microphones to capture acoustic signals generated from vehicle tires and draw a *sound map*, which is a map of time difference of vehicle sound on the two microphones, to detect vehicles. Our previous studies reported vehicle detection algorithms based on state-machine [10] or DTW (dynamic time warping) [11].

The acoustic vehicle detector, however, suffers from low detection performance when multiple vehicles are simultaneously or sequentially passing in front of microphones. Multiple vehicles draw multiple curves on a sound map. The multiple curves interfere each other, degrading detection performance.

This paper therefore presents *SAVeD*, a sequential acoustic vehicle detector with speed estimation capability. The SAVeD detects vehicles one by one while removing sound map points corresponding to the detected vehicles, which minimizes the interference between vehicles in a vehicle detection process. Sound map points and vehicles are associated by fitting a vehicle passing model to sound map points using a RANSAC (random sample consensus) robust estimation method [12]. Vehicle speed is derived from the fitting result. We conducted experiments to demonstrate that the SAVeD improved detection performance compared to the state-machine based vehicle detection.

Specifically, our key contributions are threefold:

- We present the design of SAVeD, a sequential acoustic vehicle detector with speed estimation. To the best of our knowledge, this is a first attempt to explicitly detect multiple vehicles on a sound map, which is a map of time difference of vehicle sound on two microphones.
- We present a vehicle speed estimation method using a sound map. Model fitting on sound map intuitively gives vehicle speed. We compensate the estimated vehicle speed for lane-to-lane difference of physical dimensions.
- We show detection performance and speed error of the SAVeD by experimental evaluations.

The remainder of this paper is structured as follows. Section II briefly looks through related works. Section III reviews our acoustic vehicle detector and design challenges for detection of simultaneous and sequential passing vehicles. Section IV describes the design of SAVeD, and experimental evaluations are conducted in Section V. Finally, Section VI concludes the paper.

II. Related Works

To the best of our knowledge, detection of simultaneous and sequential passing vehicles is a first attempt in the field of sound-map based acoustic vehicle sensing. This

This work was supported in part by JSPS KAKENHI Grant Numbers JP15H05708, JP17K19983, and JP17H01741 as well as the Cooperative Research Project of the Research Institute of Electrical Communication, Tohoku University.

[1]S. Ishida, J Kajimura, M Uchino, and A. Fukuda are with the Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan (email: {ishida, kajimura, uchino, fukuda}@f.ait.kyushu-u.ac.jp)

[2]S. Tagashira is with the Faculty of Informatics, Kansai University, Osaka 569-1095, Japan (email: shige@res.kutc.kansai-u.ac.jp)

This is an accepted version of the paper.

doi: 10.1109/ITSC.2018.8569727

section briefly reviews related works on vehicle detection technologies.

Current vehicle detectors are grouped into two types: intrusive or non-intrusive.

Loop coils and photoelectric tubes are categorized into the intrusive vehicle detectors. These vehicle detectors need to be installed under the road surface, which results in high installation and maintenance costs due to roadwork closing a target road section. Loop coils and photoelectric tubes also suffer from a motorbike detection problem; motorbikes are highly undetected because of small sensing coverage of the detectors.

The non-intrusive detectors are based on laser, infrared, ultrasound, radar, or camera. The non-intrusive vehicle detectors are installed above or by a road for better performance. Deployment above a road requires high installation and maintenance costs in terms of roadwork. Although installation of roadside non-intrusive vehicle detectors requires no roadwork, the roadside detectors are capable of single lane vehicle detection. Most of non-intrusive detectors are based on laser, infrared, or ultrasound, which have small coverage suffering from the motorbike detection problem.

To reduce installation and maintenance costs, camera-based vehicle detectors using CCTVs installed in the environment have been proposed [1, 2]. CCTVs, however, are available in limited areas, especially in city areas. Camera location and angle are designed for security surveillance not for vehicle sensing, resulting in low detection accuracy especially in bad weather conditions.

On the contrary, acoustic approach is a promising candidate for low cost vehicle sensing. Previous studies have reported sound-map based vehicle detectors [13–16]. We have also reported state-machine based and DTW (dynamic time warping) based algorithms that detect vehicles on a sound map [10, 11].

However, the sound-map based approaches exhibit low detection performance when multiple vehicles are simultaneously or sequentially passing in front of microphones. The sound-map based approaches define no model of vehicle passing drawn on a sound map. Sound map points and passing vehicles are not associated in the detection process, which implicitly induces interference between multiple vehicles.

Several studies have reported acoustic vehicle detectors relying on loudness on microphones instead of a sound map [17, 18]. The loudness based approaches require microphones at both side of a road and might suffer from low accuracy because of environmental noise including pedestrian voice.

III. Acoustic Vehicle Detector

Figure 1 depicts an overview of an acoustic vehicle detector. The acoustic vehicle detector consists of three blocks: a sound retriever, sound mapper, and vehicle detector.

A sound retriever consists of two microphones and LPFs (low-pass filters). Two microphones are installed by a road to collect acoustic signals generated from vehicle tires. The LPFs remove high frequency environmental noise.

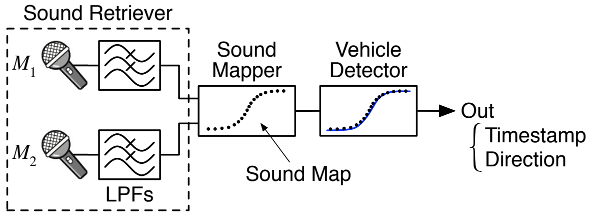

Fig. 1. Overview of acoustic vehicle detector

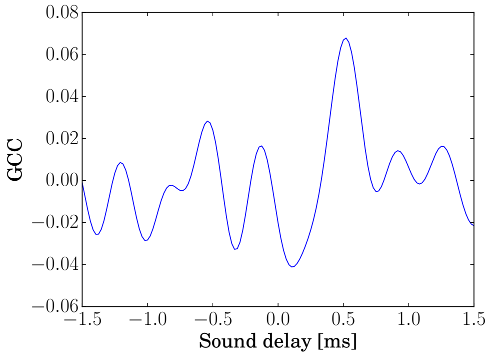

Fig. 2. Example of GCC (generalized cross-correlation) result

A sound mapper draws a sound map, which is a map of time difference of sound arrival on the two microphones. The time difference of sound arrival, i.e., sound delay, is estimated from a cross correlation function. The cross correlation function $R(t)$ over two continuous functions $s_1(t)$ and $s_2(t)$ is generally defined as:

$$R(t) = \int s_1(\tau)\, s_2(\tau + t)\, d\tau. \quad (1)$$

We substitute $s_1(t)$ and $s_2(t)$ for sound signals received on the two microphones. When the two microphones receive the same acoustic signals with sound delay Δt, i.e., $s_1(t) = s(t + \Delta t)$, $R(t)$ becomes maximum at $t = \Delta t$. We can estimate the sound delay Δt by finding a peak of a cross correlation function $R(t)$.

We use the GCC (generalized cross-correlation) function [19], which is commonly used in the field of acoustic source localization, to estimate the sound delay Δt. Figure 2 shows an example of the GCC result. We can see a peak at sound delay $\Delta t = 0.52$ milliseconds. The peak indicates that microphones received sound signals with delay of $\Delta t = 0.52$ milliseconds. Drawing the estimated delay Δt as a function of time t gives a sound map.

A passing vehicle draws an S-shaped curve on a sound map. As shown in Fig. 3, we install two microphones M_1 and M_2 separated by D in parallel to a road at L away from the road center. Sound signals generated by a vehicle travel in air and reach microphones M_1 and M_2 with traveling distances d_1 and d_2, respectively. Let x be the location of a vehicle.

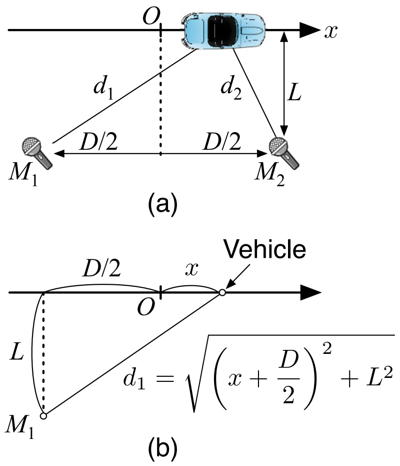

Fig. 3. Microphone setup. (a) Definition of distances D, L, d_1, and d_2. (b) d_1 and d_2 is calculated by the Pythagorean theorem.

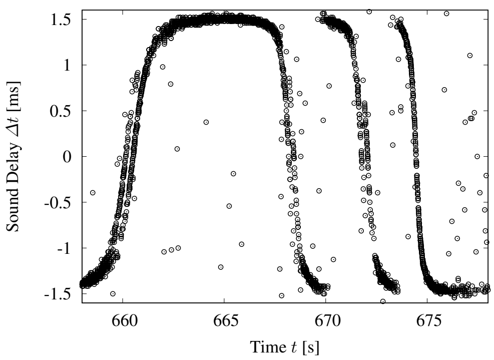

Fig. 4. Typical example of sound map. Passing vehicle draws an S-curve on sound map. In this example, four vehicles were passing: one vehicle from left to right and three vehicles from right to left.

The sound traveling distance d_1 and d_2 are calculated by the Pythagorean theorem as:

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

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

We derive sound delay Δt between microphones M_1 and M_2 using the speed c of sound in air:

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

Equation (4) indicates that an S-curve appears on a sound map when x is increasing or decreasing linearly; a passing vehicle in a constant speed draws an S-curve.

Figure 4 shows a typical sound map. Direction of S-curves indicates direction of passing vehicles. Figure 4 indicates that four vehicles were passing: one from left to right and three from right to left.

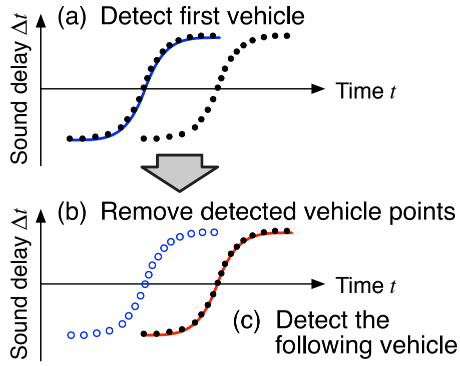

Fig. 5. Key idea of SAVeD. (a) SAVeD detects a preceding vehicle (b) and removes sound map points corresponding to the detected vehicle. (c) SAVeD continues detection process for the following vehicles.

A vehicle detector analyzes a sound map to find S-curves to detect vehicles. We have reported state-machine based [10] and DTW (dynamic time warping) based vehicle detection algorithms [11]. These algorithms take no considerations on simultaneous and sequential passing vehicles. S-curves sometimes mix up, which degrades detection performance because the curves interfere each other in the detection algorithms.

Two challenges come up with multiple vehicle detection.

1) *How to split S-curves of different vehicles?*
Multiple S-curves drawn by multiple vehicles interfere each other, resulting in low detection accuracy. We need to process S-curves one by one to avoid such interference.
2) *How to minimize the effect of sound map noise?*
To draw a sound map, we estimate sound delay by finding a peak on a cross correlation function. Sound signals from multiple vehicles interfere, which makes cross correlation weaker. The weak cross correlation tends to be affected by environmental noise. The sound map therefore becomes noisy when multiple vehicles are in front of microphones.

The following section describes SAVeD, a sequential acoustic vehicle detector that addresses the above two challenges.

IV. Design of SAVeD

A. Key Idea

The key idea of SAVeD is that sound map points corresponding to detected vehicles are removed prior to detection of the following vehicles. Figure 5 depicts the key idea of SAVeD. The SAVeD detects vehicles in three steps. (a) The SAVeD first detects vehicles from the left side of a sound map (b) and next removes sound map points corresponding to the detected vehicles. (c) The SAVeD continues to detect the following vehicles. Multiple S-curves drawn by multiple vehicles partially overlap on a sound map. The sequential

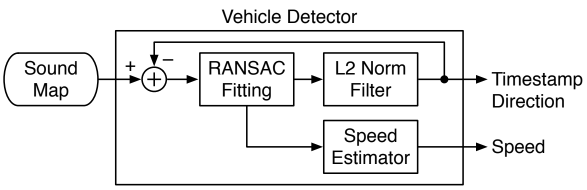

Fig. 6. Overview of vehicle detector block in SAVeD

detection steps successfully detect vehicles one by one with small influence of the following vehicles.

The sound map points are associated with a vehicle by fitting an S-curve model to sound map points. We use a RANSAC (random sample consensus) robust estimation method for model fitting to adapt to a noisy sound map.

B. Design Overview

The SAVeD sequentially detects vehicles with negative feedback of detected vehicle information in a vehicle detector block. Figure 6 depicts an overview of vehicle detector block in the SAVeD. The SAVeD vehicle detector block consists of RANSAC fitting, L2 norm filter, and speed estimator modules.

A RANSAC fitting module fits an S-curve model Eq. (4) to sound map points to detect vehicles. The RANSAC fitting incorrectly detects vehicles when two vehicles are sequentially passing. We apply a filter based on L2 norm to reduce such false positive detections. The SAVeD system has a negative feedback loop to remove sound map points corresponding to the detected vehicles. The speed estimator module estimates vehicle speed based on a fitted S-curve.

The following subsections describe the each module in more detail.

C. RANSAC Fitting Module

The RANSAC fitting module fits an S-curve model Eq. (4) to sound map points. Equation (4) formulates sound delay Δt by the location x of a vehicle. We first rewrite Eq. (4) to formulate sound delay Δt by time t. Assume that a vehicle is passing right in front of microphones at $t = 0$ at a constant speed of v. Equation (4) is now rewritten as

$$\Delta t = \frac{1}{c}\left\{\sqrt{\left(vt + \frac{D}{2}\right)^2 + L^2} - \sqrt{\left(vt - \frac{D}{2}\right)^2 + L^2}\right\}. \tag{5}$$

The RANSAC fitting module estimates vehicle speed v in Eq. (5) to fit the model to sound map points.

Figure 7 depicts an overview of a fitting process using RANSAC. The RANSAC fitting process consists of four steps.

1) The RANSAC fitting module randomly samples a sound map point. Multiple points might be sampled in this step. The number of samples is usually set to minimum number required to estimate unknown parameter in a model formula. Remind that vehicle speed v is a real number, we can estimate v with one point from Eq. (5).
2) The RANSAC fitting module estimates v. Let $\hat{v}$ be the estimated vehicle speed. An S-curve is drawn on a sound map with the estimated speed $\hat{v}$,
3) and the sum of distances between the S-curve and each sound map point is calculated. To reduce computational complexity, we use distance in a vertical Δt axis instead of the shortest distance between the S-curve and sound map points.
4) The RANSAC fitting module repeats steps 1)–3) to find a curve whose distance sum is minimum, completing the fitting process.

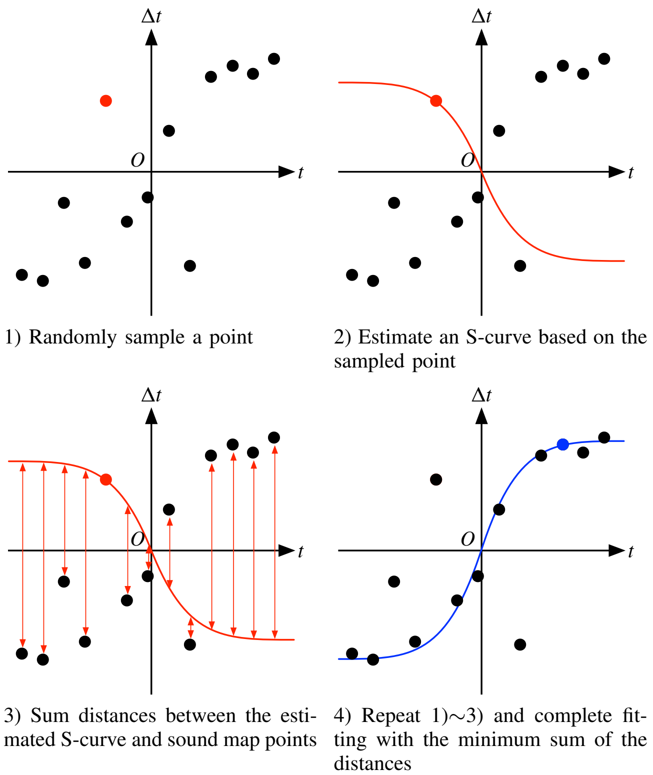

Fig. 7. Overview of RANSAC fitting

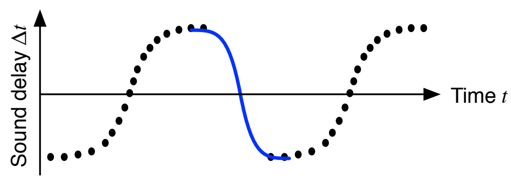

Fig. 8. False positive detection by sequentially passing vehicles

The RANSAC fitting module finally applies a filter based on the distance sum to detect vehicles. The RANSAC fitting always gives an estimated S-curve that best fits to sound map points even if no vehicle is passing. We apply a threshold to the sum of distances after step 3) to check if a vehicle is passing.

D. L2 Norm Filter Module

The RANSAC fitting module incorrectly detects an extra vehicle when two vehicles are sequentially passing, as shown in Fig. 8. The falsely estimated S-curve partially matches to two actual S-curves at $\Delta t \simeq \pm D/c$. The partial match makes it difficult to avoid the false positive detections in the RANSAC fitting.

To reduce such false positive detections, the SAVeD employs a filtering process based on L2 norm. L2 norm is calculated over the sound map corresponding to the vehicle passing derived by the RANSAC fitting. Detections with L2 norm above a threshold are filtered out as they are false positive detections. Let $\mathbf{D}$ be a set of sound map points where vehicle is detected in the RANSAC fitting. L2 norm $\|\mathbf{D}\|$ is defined as

$$\|\mathbf{D}\| = \sqrt{\sum_{i \in \mathbf{D}} i^2}. \tag{6}$$

As shown in Fig. 8, the sound map corresponding to false positive detections includes an abrupt change of sound delay Δt between $\pm D/c$, which results in high L2 norm. The number of samples $\|\mathbf{D}\|$ might be different for each vehicle. We normalize the L2 norm with the size $|\mathbf{D}|$ prior to thresholding.

E. Speed Estimator Module

The vehicle speed is estimated in the RANSAC fitting module, as described in Section IV-C. The estimated vehicle speed $\hat{v}$, however, depends on a lane where the vehicle is running because the sound map curve model, i.e., Eq. (5), includes distance L between microphones and a road.

The speed estimator module therefore compensates the estimated speed $\hat{v}$ for lane-to-lane difference of distance L. We first formulate vehicle speed from a sound map model. Differential of Eq. (5) gives the slope m at $t = 0$ as

$$\begin{aligned} m &= \left.\frac{d}{dt}\Delta t\right|_{t=0} \\ &= \frac{v}{c}\left\{\left(\frac{L}{D}\right)^2 + \frac{1}{4}\right\}^{-\frac{1}{2}}. \end{aligned} \tag{7}$$

Rewriting Eq. (7), we derive vehicle speed v as

$$v = mc\sqrt{\left(\frac{L}{D}\right)^2 + \frac{1}{4}}. \tag{8}$$

We then compensates a vehicle speed by using the slope m of the S-curve estimated in a RANSAC fitting module. Let L_{mod} be the microphone-road distance used in the RANSAC fitting and L_{act} be the actual distance between microphones and a vehicle running lane. The compensated vehicle speed $\tilde{v}$ is calculated as

$$\begin{aligned} \tilde{v} &= mc\sqrt{\left(\frac{L_{\mathrm{act}}}{D}\right)^2 + \frac{1}{4}} \\ &= \hat{v}\sqrt{\frac{4{L_{\mathrm{act}}}^2 + D^2}{4{L_{\mathrm{mod}}}^2 + D^2}}. \end{aligned} \tag{9}$$

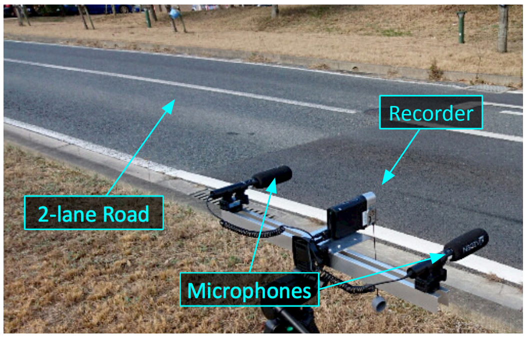

Fig. 9. Experiment setup

TABLE I
NUMBERS OF SEQUENTIAL AND SIMULTANEOUS PASSING VEHICLES AMONG ALL PASSING VEHICLES

Direction	All	Sequential	Simultaneous
Left to Right	124	54	15
Right to Left	54	8	18
Total	178	62	33

In the following section, we conduct our experimental evaluations on two-lane road, one lane in each direction. We change the distance L_{act} in Eq. (9) based on an S-curve direction. It is our future work to adapt the speed estimation method to roads with multiple lanes in each direction.

V. Evaluation

To demonstrate the effectiveness of the SAVeD, we experimentally evaluated detection accuracy and vehicle speed error.

A. Experiment Setup

Figure 9 shows an experiment setup. A target road has two of 3-meter width lanes; one lane in each direction. We installed two microphones at a sidewalk of the road at 1.5 meters away from a road edge at a height of one meter. The vehicle sound was recorded for approximately 20 minutes at 48-kHz sampling with 16-bit code length using a video recorder. The experiment was conducted at rush hour as many vehicles are simultaneously or sequentially passing in front of the microphones. We used AZDEN SGM-990 microphones connected to a SONY HDR-MV1 video recorder. Referring to [10], we set distance between the microphones to 50 centimeters. We also recorded video monitoring the road as ground truth data.

During the experiment, 178 vehicles passed. Table I shows the numbers of sequential and simultaneous passing vehicles in each direction. Vehicles with another vehicle passing within 2 seconds were defined as sequential or simultaneous passing vehicles. Approximately half of passing vehicles were sequential or simultaneous passing vehicles. The passing vehicles include not only normal cars, but also buses, trucks, and motorbikes. Table II summarizes the number of vehicles for each vehicle type in each direction.

TABLE II
NUMBER OF PASSING VEHICLES FOR EACH VEHICLE TYPE

Direction	All	Bike	Bus	Truck	Normal
Left to Right	124	77	15	0	32
Right to Left	54	17	10	2	25
Total	178	94	25	2	57

B. Vehicle Detection Performance

We evaluated vehicle detection performance by F-measure, which is commonly used in classifier evaluations. F-measure is defined as:

$$\mathrm{F}_{\mathrm{measure}} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \tag{10}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{11}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{12}$$

where TP, FN, and FP are the numbers of true positives, false negatives, and false positives, respectively. True positives, false negatives, and false positives are defined as the cases that a vehicle is detected when a vehicle is passing, that no vehicle is detected when a vehicle is passing, and that a vehicle is detected when no vehicle is passing, respectively.

To demonstrate relative detection performance, we compared the performance of the following three methods.

1) *SAVeD:*
The SAVeD method is a proposed method presented in Section IV. Sound map points corresponding to detected vehicles are removed prior to continuing a vehicle detection process.
2) *Non-removal:*
The non-removal is a vehicle detection without removing sound map points. Other parts, including RANSAC fitting and L2 norm filtering, is the same as the SAVeD.
3) *State-machine:*
The state-machine method is a vehicle detection method presented in our previous work [10]. A state-machine based algorithm analyzes a sound map to detect vehicles.

Table III shows detection performance, i.e., the numbers of TPs, FNs, and FPs as well as the calculated precision, recall, and F-measure of the three methods. Table III indicates the following.

- The SAVeD method showed the most largest F-measure of 0.83 among the three methods. Compared to the state-machine method, the SAVeD method improved an F-measure by more than 10 points.
- Recall of the SAVeD and non-removal methods is greater than that of the state-machine method. The SAVeD and non-removal methods utilizes RANSAC fitting presented in Section IV-C, which increased the number of TPs for simultaneous and sequential vehicles.
- The state-machine method exhibited the most largest precision of 1.00. The state-machine method detects vehicles only and only when almost complete S-curves appeared on a sound map, which resulted in high precision with the large number of FNs.
- Precision of the SAVeD method was greater than that of the non-removal method. The SAVeD detects vehicles one by one while removing sound map points corresponding to the detected vehicles. The non-removal method incorrectly detected a vehicle as multiple vehicles because an S-curve was detected multiple times, which resulted in large number of FPs.

The above results confirm that the SAVeD method successfully detected vehicles with an F-measure greater than the state-machine method.

C. Vehicle Speed Error

As an initial evaluation of speed estimation method of the SAVeD, we evaluated vehicle speed error for several passing vehicles. The vehicle speed error is defined as relative difference between actual and estimated vehicle speeds. Let v_i and $\widetilde{v_i}$ be actual and estimated speeds of vehicle i, respectively. The vehicle speed error ε_i is defined as

$$\varepsilon_i = \left| \frac{\widetilde{v_i} - v_i}{v_i} \right|. \tag{13}$$

The actual vehicle speed is manually estimated from video images monitoring a target road.

Table IV shows vehicle speed error calculated over 12 randomly selected vehicles. Maximum and mean vehicle speed error was 30.5 % and 16.8 %, respectively. High error mainly occurred when motorbikes passed, which was mainly caused by the error of distance L. Width of each lane on the target road is 3 meters. We assumed that the vehicle is passing at the center of a dedicated lane in our sound map model. Motorbikes freely choose their running position on the lane, which resulted in the high error.

Note that ground truth of vehicle speed is manually estimated from a video images of a target road because we don't have an equipment to measure vehicle speed. Further investigation is required to strengthen our contribution.

Although vehicle speed error was quite high, we believe that speed estimation in SAVeD is still useful to recognize road traffic condition on each lane. The SAVeD has advantages over radar-based speed estimator in non line-of-sight deployment and in multiple lane monitoring. Radar-based speed estimator needs to be installed at line-of-sight to a road. In contrast, SAVeD allows non line-of-sight deployment, which eases restrictions on deployment. While multiple radar-based speed estimators are required to monitor multiple lanes, SAVeD only requires two microphones installed at a sidewalk to monitor multiple lanes. We emphasize that the SAVeD is designed for vehicle sensing, not for speed measurement. The speed estimation is an optional feature to recognize road traffic condition on each lane.

VI. CONCLUSION

In this paper, we presented SAVeD, a sequential acoustic vehicle detector. The SAVeD relies on two microphones installed at a sidewalk to draw a sound map, which is a map

TABLE III

EXPERIMENT RESULTS OF VEHICLE DETECTION PERFORMANCE

	SAVeD			Non Removal			State Machine		
	Left to Right	Right to Left	Total	Left to Right	Right to Left	Total	Left to Right	Right to Left	Total
TP	103	44	147	103	41	144	62	37	99
FN	21	10	31	21	13	34	62	17	79
FP	18	11	29	70	47	117	0	0	0
Precision	0.85	0.80	0.84	0.60	0.47	0.56	1.00	1.00	1.00
Recall	0.83	0.81	0.83	0.83	0.76	0.81	0.50	0.69	0.56
F-measure	0.84	0.81	0.83	0.69	0.58	0.66	0.67	0.81	0.71

TABLE IV

VEHICLE SPEED ERROR

Vehicle		Actual Speed	Estimated Speed	Error
No. i	Type	v_i [m/s]	$\widetilde{v_i}$ [m/s]	ε_i [%]
1	normal	9.7	8.70	10.3
2	motorbike	15.1	11.84	21.6
3	normal	7.4	6.58	11.1
4	normal	-17.0	-15.15	10.9
5	motorbike	11.9	8.51	28.5
6	motorbike	8.8	6.41	27.1
7	motorbike	-10.9	-9.60	12.0
8	motorbike	-15.2	-11.29	25.7
9	truck	-8.8	-9.00	2.2
10	normal	-11.9	-10.10	15.1
11	normal	-10.9	-11.63	6.7
12	motorbike	11.4	7.92	30.5

of sound arrival time difference on the two microphones. We presented a vehicle detection algorithm based on a RANSAC robust estimation method that analyzes a sound map to detect S-shaped curves drawn by passing vehicles. Experimental evaluations revealed that the SAVeD successfully detected vehicles with an F-measure of 0.83, which was more than 10-point improvement compared to a state-machine based algorithm presented in our previous paper. The SAVeD estimated speed of vehicles on each lane. Although the speed estimation error was high up to 30.5 %, the speed estimation is still useful for traffic monitoring on each lane.

We believe that the SAVeD can be a candidate of low-end substitution of current vehicle detectors that are used in traffic monitoring mainly for car navigation systems. As future works, we plan to conduct experiment at a more wider road with multiple lanes in each direction. We also need to improve speed estimation accuracy prior to practical use.

REFERENCES

[1] 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.

[2] 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.

[3] C. de Fabritiis, R. Ragona, and G. Valenti, "Traffic estimation and prediction based on real time floating car data," in *Proc. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2008, pp. 197–203.

[4] D. B. Work, O.-P. Tossavainen, S. Blandin *et al.*, "An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices," in *Proc. IEEE Decision and Control (CDC)*, Dec. 2008, pp. 5062–5068.

[5] F. Calabrese, M. Colonna, P. Lovisolo *et al.*, "Real-time urban monitoring using cell phones – a case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.

[6] X. Zhou, W. Wang, and L. Yu, "Traffic flow analysis and prediction based on GPS data of floating cars," in *LNEE*, vol. 210, Nov. 2012, pp. 497–508, proc. Int. Conf. on Information Technology and Software Engineering.

[7] G. Guido, V. Galleli, F. Saccomanno *et al.*, "Treating uncertainty in the estimation of speed from smartphone traffic probes," *Transporation Research Part C: Emerging Technologies*, vol. 47, pp. 100–112, Oct. 2014.

[8] T. Seo, T. Kusakabe, and Y. Asakura, "Estimation of flow and density using probe vehicles with spacing measurement equipment," *Transporation Research Part C: Emerging Technologies*, vol. 53, pp. 134–150, Apr. 2015.

[9] T. Seo and T. Kusakabe, "Probe vehicle-based traffic state estimation method with spacing information and conservation law," *Transporation Research Part C: Emerging Technologies*, vol. 59, pp. 391–403, Oct. 2015.

[10] S. Ishida, K. Mimura, S. Liu *et al.*, "Design of simple vehicle counter using sidewalk microphones," in *Proc. ITS EU Congress*. EU-TP0042, Jun. 2016, pp. 1–10.

[11] S. Ishida, S. Liu, K. Mimura *et al.*, "Design of acoustic vehicle count system using DTW," in *Proc. ITS World Congress*. AP-TP0678, Oct. 2016, pp. 1–10.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] A. Toyoda, N. Ono, S. Miyabe *et al.*, "Traffic monitoring with ad-hoc microphone array," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 318–322.

[18] T. Toyoda, N. Ono, S. Miyabe *et al.*, "Vehicle counting and lane estimation with ad-hoc microphone array in real road environments," in *Proc. Int. Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP)*, Mar. 2016, pp. 622–625.

[19] 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.