International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 14, Issue 7 (July Ver. II 2018), PP.63-69

A study on determination of appropriate aggregation interval to provide driving environment information using vehicle sensing data

Intaek Jung[1]
[1]*Korea Institute Of Civil Engineering And Building Technology, Department Of Future Technology And Convergence Research, 283, Goyang-Daero, Ilsanseo-Gu, Goyang-Si, Gyeonggi-Do, Republic Of Korea.*
Corresponding Author:Intaek Jung1

ABSTRACT: A driving environment analysis platform based on vehicle sensors is being developed to provide driving environment information. To accomplish this, the vehicle sensing data collected from the individual vehicle sensors should be aggregated at regular time intervals. However, the aggregation interval of existing systems is mostly an interval of 5 min, which cannot satisfy users who demand more detailed road weather and traffic information. Therefore, in this study, the appropriate aggregation interval of individual vehicle sensing data is determined to provide road driving environment information. The mean square error was used to determine the aggregation interval, and the aggregation interval was calculated as 30s.

Keywords*:*Vehicle sensor, Aggregation interval, Driving environment information, Mean square error, Platform

---

Date of Submission: 27-07-2018 Date of acceptance: 11-08-2018

I. INTRODUCTION

Owing to rapid industrial development and economic growth, the number of cars has increased significantly. Since October 2014, the number of registered cars in Korea has exceeded 20 million, and is the 15th-highest in the world and the fourth in Asia after Japan, China, and India. This increase in the number of cars is an indicator of today's economic growth; however, it has caused many side effects such as traffic accidents, traffic congestion costs, and environmental pollution. In particular, the number of road traffic accidents per year has increased by more than six times from 37,243 in 1970 to 232,035 in 2015. The number of traffic accident deaths per 10,000 vehicles was 2.2 in 2013 in 32 of the 34 OECD countries [1]. The annual traffic congestion cost has continuously increased, from 4.6 trillion won in 1991 to 33.4 trillion won in 2015 [2].

In addition, the road driving environment can change instantaneously, regardless of time or place, owing to various unexpected situations and events such as bad weather conditions, road icing, traffic congestion, and traffic accidents. If drivers do not recognize the sudden change in road conditions in advance, a large traffic accident is more likely to occur. To cope with safety problems and traffic congestion caused by rapidly changing driving environments, various real-time information collection systems such as an intelligent traffic system (ITS) and road weather information system (RWIS) are in operation. In the case of existing systems, information on specific road points or specific road sections installed with fixed sensors such as weather observation equipment, loop sensors, image sensors, and the like is collected. However, there is a spatial constraint problem where information cannot be collected in places where fixed sensors are not installed. To solve such a space coverage problem, many fixed sensors must be installed, which causes excessive budget requirements. In the case of road traffic information, a traffic information collection system based on mobile sensors using probe cars such as taxis and buses is in operation. However, this is also disadvantageous in that it is not scalable nationwide because the information is collected only in sections where fixed side roadside equipment (RSE) is installed. Currently, an analysis platform is being developed to generate and provide driving environment information using vehicle sensing data collected from individual vehicle sensors, or mobile sensors, to compensate for the disadvantages of fixed sensors [3].

Vehicle sensing data collected on this platform can be viewed as continuous data on the road generated by GPS coordinate units every 1 – 10 s, rather than as discrete data. To provide driving environment information such as road surface temperature, precipitation, and traffic density using individual vehicle sensing data, collected data should be aggregated at regular time intervals. However, most of the aggregation intervals of existing systems are aggregated at intervals of 5 min, and they are not able to provide road weather information and traffic information in shorter time intervals. Therefore, this study determines the appropriate aggregation interval for individual vehicle sensing data to provide driving environment information of various types.

II. CURRENT STATUS AND LITERATURE REVIEW

A. Current status

Table 1 shows the current status of major systems for collecting and providing weather information and traffic information among the current road environment information. In case of weather information, various systems such as the Korea Meteorological Administration (KMA), SK Planet, the RWIS of Korea Expressway Corporation (EX), and Regional Construction and Management Office (RCMO) are in operation, and it has been shown that the whole provides information over a period of 5 to 12 h. In the case of traffic information, there are the National Transport Information Center (NTIC) of the Ministry of Land, Infrastructure, and Transport, the Urban Traffic Information System (UTIS) of the Korean National Police Agency, Roadplus of EX, and the Seoul Transport Operation and Information Service (TOPIS) of the Seoul Metropolitan Government, and the whole provides information 5 min to 1 day.

Table 1: Review status of major weather and traffic information systems

		Collection method	Provision cycle	Spatial unit
Weather information	KMA	ground observation station	1 h, 3 h	observation point, administrative district
	SK Planet (weather planet)	meteorological station based on fixed sensors	5 min, 1h	
	EX (RWIS)	roadside weather observation equipment	15 min, 30 min, 1h	specific road survey points, specific road sections
	RCMO (RWIS)			
Traffic information	NTIC	loop detector, image detector, radar, AVI (Automated Vehicle Identification), DSRC (Dedicated Short Range Communication), GPS	5 min, 1h, 1day	road sections
	UTIS	OBU (On Board Unit), RSE	5 min	
	ROADPLUS	VDS (Vehicle Detection System), TCS (Electronic Toll Collection System), DSRC	5 min, 15min, 1h, 1day	
	TOPIS	loop detector, image detector, radar detector, geomagnetic sensor, DSRC, card cab	5 mins	

As mentioned above, most weather and traffic information systems are using fixed sensors. Although the mobile sensors of some probe vehicles are utilized, these are also a fixed type collection system based on OBU and RSE. Systems utilizing mobile sensors are negligible, and there is no system that utilizes vehicle sensing data collected from individual vehicle sensors. Most of the information providing period is the same as the aggregation time interval of the collected data, and the information is renewed at intervals of 5 min or more as a whole. This corresponds to a macroscopic aggregation interval, and a method for determining the aggregation interval of a more microscopic unit is needed.

B. Literature review

Previous studies related to the determination of the appropriate aggregation interval of collected data have been conducted since the introduction of ITS in Korea in the late 1990s. As shown in Table 2, data collected from various detectors such as loop detectors, AVI, and DSRC were used to determine the appropriate aggregation interval. The applied methods were Cross Validated Mean Square Error (CVMSE), F-test, Forecasting Mean Square Error (FMSE), and the appropriate sample number. The appropriate aggregation interval was calculated as being from 1 min to 60 min.

Table 2: Summary of related literature review

	Methodology	Optimal aggregation interval	Related studies
Loop detector	CVMSE, F-test	1min, 60 min	Gajewski et al.(2000)
	CVMSE	3−5min,	Yoo et al.(2004)
	FMSE	10−20min	
AVI	MSE	3−5min, 5−15 min	Park et al.(2001)
		10−20min	Choi et al.(2012)
	MSE, F-test	5 min	Lim et al.(2004), Lim et al.(2005)
DSRC	MSE, Sample size	5 min	Shim et al.(2011)

Gajewski et al. (2000) collected ITS traffic monitoring data at 20 or 30 s intervals at most traffic information centers, but the time interval for aggregating data varied from 1 min to 5 min, 15 min, and 60 min. Two statistical methods such as the CVMSE method and F-test method have been proposed to determine the optimum time interval of ITS traffic monitoring data. When the variation of traffic volume was low, the optimal aggregation interval was determined to be more than 60 min in both methods, and in the opposite case, the optimal aggregation interval was determined to be less than 1 min in both methods. Therefore, the optimal time aggregation interval determined using the two methods is reasonably derived [4]. Yoo et al. (2004) developed a model that determines the appropriate aggregation time interval for link and path travel time estimation and prediction when collecting travel time data using a loop detector. The CVMSE method was used as the optimal aggregation interval determination model for link and path travel time estimation. The FMSE method was applied to determine the optimal aggregation interval for link and path travel time prediction. The proposed methods were applied to the loop detector data of the Gyeongbu Expressway. The application results were analyzed as 3–5 min for the appropriate aggregation interval for link and path travel time estimation, and 10–20 min for the optimal aggregation interval for link and path travel time prediction [5]. Park et al. (2009) proposed a method of determining the optimal aggregation time interval using MSE for the estimation or prediction of the travel time for an expressway section and route. As a result, the optimal aggregation time interval for the estimation and prediction of the travel time of the road section was determined to be 3 to 5 min, and the optimal aggregation time interval for the travel time estimation and prediction of the traffic route was determined to be 5 to 15 min. The optimal aggregation time interval for predicting the travel time of the road section was larger than the optimal aggregation time interval for estimating the travel time of the road section. The optimal aggregation time interval for the estimation and prediction of the travel time was analyzed to be highly influenced by the correlation between the road segments forming the traffic route [6]. Choi (2012) considered the traffic characteristics affecting the aggregation interval when estimating interrupted flow speed, and suggested the optimal aggregation interval by using a point estimation technique using MSE. As a result, the optimal aggregation interval for each type was analyzed as 10 to 20 min. That is, traffic volume has a major influence on the determination of the appropriate aggregation interval, and a new model is needed, especially when the traffic volume is less than 5,000 vehicles/day. Signal-related variables such as the installation density of signal lights and the green/cycle ratio have no significant influence on the determination of the aggregation interval when the operation management section is longer than 5 km. When determining the aggregation interval, it is necessary to consider the traffic volume of the relevant section. In the road design phase, the aggregation interval should be determined considering the annual average daily traffic (AADT) [7]. Lim et al. (2004, 2005) applied MSE and the F-test as a statistical estimation method to calculate the appropriate time interval for the AVI data collection for an interrupted flow section of a local highway. As a result, it was concluded that MSE is more sensitive to the aggregation interval than the F-test and more accurate determination of the appropriate aggregation interval is possible, the authors concluded that the current 5 min interval is appropriate [8, 9]. Shim et al. (2011) analyzed the satisfaction of the MSE and the appropriate sample number according to the time aggregation interval using highway high-pass DSRC data. As a result, the shorter the time aggregation interval is, the less the MSE decreases but it does not satisfy the appropriate sample number. Therefore, they judged that 5 min is appropriate for the aggregation time interval to satisfy the appropriate sample number [10].

The above previous studies have used only data collected from fixed sensors such as loop detectors, AVI, and DSRC, and there is no evaluation of sensing data collected from individual mobile sensors such as vehicle sensors. The optimal aggregation interval calculated in the previous studies is commonly preferred as a 5 min aggregation interval. However, to provide microscopic environment information, it is necessary to find the optimal value in the aggregation unit that is shorter than 5 min.

III. METHODOLOGY

A. **Development overview**

This study proposes a methodology to determine the appropriate aggregation interval of vehicle sensing data collected from individual vehicle sensors to provide microscopic driving environment information such as wind, road icing, traffic congestion, etc. The microscopic information refers to a aggregation interval that is more subdivided than the time interval for updating information of the current road weather and road traffic information systems. As shown in Fig. 1, the vehicle sensing data refers to the data (road surface temperature, precipitation, traffic density, etc.) generated by the information generation module from data collected every 1 to 10 s measured from an individual vehicle.

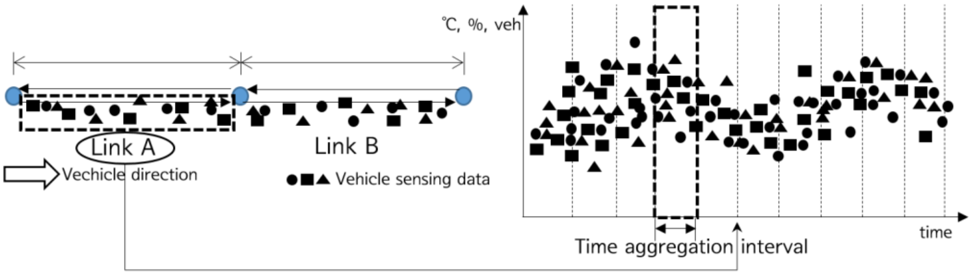

Fig.1: Conceptual diagram for determining aggregation time interval

B. **Collection form of vehicle sensing data**

The vehicle sensing data is data observed from an individual vehicle, and a driving environment analysis platform based on vehicle sensors for collecting the vehicle sensing data is under development. As shown in Fig. 2, the vehicle sensors consist of radar, camera, temperature sensor, GPS, etc., and the same vehicle sensors as in a conventional vehicle were installed on the experimental vehicle and the vehicle sensing data at the site was collected.

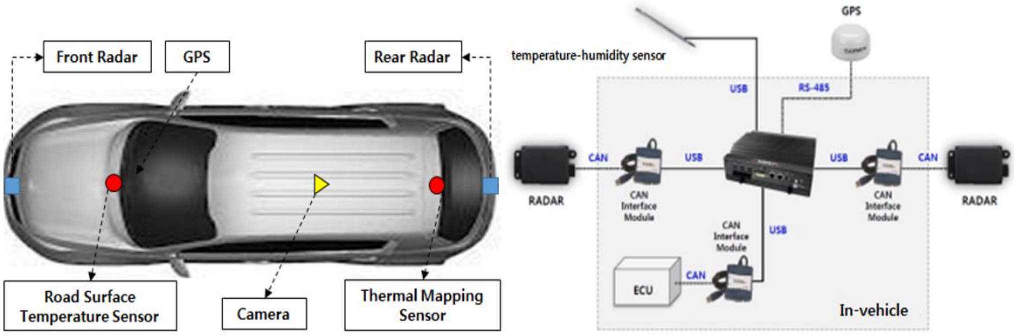

Fig. 2: Concept of vehicle sensor installation

The vehicle sensing data of this study is the road surface temperature, precipitation, and traffic density generated by the information generation module of the platform using data collected from individual vehicle sensors. Table 3 lists the collected vehicle data and the data obtained by processing the data collected in 1 to 10 s.

Table 3: Collection form of vehicle sensing data

No	Parameter	Value type	Unit	Value
1	GPSTime	Time	(HMS)	8:24:29.000000
2	Date	Data	(MDY)	9/17/2015
3	Latitude	Float8	(+/- D M S)	37 39 38.41317
4	Longitude	Float8	(+/- D M S)	126 43 00.58528
5	Road surface temperature	Float4	(℃)	18.9
6	Precipitation	Float4	(mm/h)	0.0
7	Traffic density	Float4	(veh/km/lane)	78.0

C. **Determining the appropriate aggregation interval**

In this study, the MSE method, which is generally applied in previous studies, is applied to determine the appropriate time interval of individual vehicle sensing data. This method determines the aggregation interval at which the error of the collected data becomes the minimum. The method of determining the appropriate aggregation interval in this study is performed through the following four steps as follows.

Step 1: Setting the range for determining the appropriate aggregation interval

Aggregation interval (h) = 10 s, 20 s, 30 s, 40 s, 50 s, 1 min, 2 min, 3 min, 4 min, 5 min

Step 2:MSE calculation by aggregation interval(h)

$$\text{MSE} = \text{E}[\frac{\sum_{i=1}^{p_h}(d_{h,i}-\bar{d}_h)^2}{P_h} + (\bar{d}_h - \mu_h)^2], \hat{\mu}_h = \bar{d}_h = \frac{\sum_{i=1}^{p_h} d_{h,i}}{p_h - 1}$$

where, p_h: number of vehicles during time interval h

$d_{h,i}$: vehicle sensing data of the i-th vehicle during the time interval h

μ_h: average vehicle sensing data of the population during time interval h

$\bar{d}_h$: average vehicle sensing data during time interval h
$\hat{\mu}_h$: parameter estimate over time interval h

Step 3:Convert MSE to maximum time interval (H = 1 hour)
Convert MSE to 1 h based on a time interval of h<H

$$\widehat{MSE}_H = \frac{\sum_{h=1}^{H}(p_h \times MSE_h)}{p_H}$$

where,$\widehat{MSE}_H$: conversion value of MSE based on time interval h
p_H: number of vehicles during time interval h

Step 4:Determine the appropriate aggregation interval (h_t^o)
Determine h_t^o so that $\widehat{MSE}_H$ is minimized according to time interval h
min $= \widehat{MSE}_{H,h}$, h $= h_t^o$

IV. CASE STUDY AND RESULTS

Currently, a driving environment analysis platform based on vehicle sensors is being developed to monitor continuous and local road weather and traffic data (road surface temperature, precipitation, traffic density). Vehicle sensing data such as atmospheric temperature, atmospheric humidity, and road surface temperature were collected using an experimental vehicle equipped with the same vehicle sensors as the test vehicle used in the test road sections (Jayuro, Jangwol IC–Isanpo IC). It is necessary to use the vehicle sensing data collected from a large number of vehicles passing through a given road section. However, at present, the vehicle sensing data collected from two experimental vehicles are used according to the development conditions. Because the vehicle sensing data is collected using only two test vehicles, it is not possible to observe the data continuously over the same interval. Therefore, from October 24th to October 28th, 2016, the time intervals were divided so as not to overlap each other day by day, and then continuous data of more than 3 h per day were generated by combining the collected data by time of day. The results of the construction of the generated vehicle sensing data are shown in Table 4.

Table 4: Result of data construction using vehicle sensing data

Vehicle ID	Time	Latitude	Longitude	Road surface temperature(℃)	Air temperature(℃)	Air humidity(%)
1	2016/10/25 14:00:01	37.660670	126.716829	27.92	53.99	28.20
1	2016/10/25 14:00:02	37.660670	126.716829	27.92	53.89	28.00
1	2016/10/25 14:00:03	37.660670	126.716829	27.91	54.51	28.20
1	2016/10/25 14:00:05	37.660670	126.716829	27.91	53.80	28.70
1	2016/10/25 14:00:06	37.660670	126.716829	27.90	53.86	28.70
1	2016/10/25 14:00:08	37.660670	126.716829	27.89	53.81	28.70
1	2016/10/25 14:00:09	37.660670	126.716829	27.89	53.93	28.00
1	2016/10/25 14:00:10	37.660670	126.716829	27.89	53.96	28.00
1	2016/10/25 14:00:11	37.660670	126.716830	27.88	54.02	28.20
1	2016/10/25 14:00:12	37.660670	126.716830	27.88	53.98	28.10
1	2016/10/25 14:00:13	37.660670	126.716830	27.87	53.87	28.10
1	2016/10/25 14:00:14	37.660670	126.716830	27.86	54.20	28.60
1	2016/10/25 14:00:15	37.660670	126.716830	27.86	53.84	28.80
1	2016/10/25 14:00:16	37.660670	126.716830	27.86	53.97	29.10
1	2016/10/25 14:00:17	37.660670	126.716830	27.85	54.86	29.20
⋮	⋮	⋮	⋮	⋮	⋮	⋮

The method of determining the optimal aggregation interval proposed in this study was applied using the vehicle sensing data constructed above. As shown in Table 5, the average number of samples per hour was more than 13.4 counts/ 5 min, which was originally shown by Shim et al. (2013) [11].

Table 5: Average sample size by aggregation interval per hour(unit: counts)

	Aggregation interval									
	10 s	20 s.	30 s	40 s	50 s	1min.	2min.	3min.	4min.	5min.
14:00 to 15:00	89	178	267	355	444	533	1,066	1,599	2,133	2,666
15:00 to 16:00	120	240	360	480	599	719	1,439	2,158	2,877	3,597
16:00 to 17:00	74	147	221	295	368	442	884	1,326	1,768	2,210
Average	94	188	283	377	470	565	1,130	1,694	2,259	2,824

As shown in Table 6, MSE is minimized at the aggregation interval of 30 s as a result of calculating the MSE according to the aggregation interval set in the following (h_t^o. In the previous studies, the 5-min aggregation interval was suggested as the optimal aggregation interval, which is different from the result of this study. It is considered that this reflects the difference between types of collected data such as atmospheric temperature, atmospheric humidity, road surface temperature, and vehicle sensing data. Also, unlike the previous studies, in which the MSE decreases with increasing aggregation interval, MSE increases with increasing aggregation interval in this study. This means that the longer the aggregation interval is, the greater the variability of vehicle sensing data.

Table 6: MSE by aggregation interval per hour(unit: ℃, %)

		Aggregation interval									
		10 s	20 s.	30 s	40 s	50 s	1min.	2min.	3min.	4min.	5min.
Road surface temperature	14:00 to 15:00	12.9	12.9	12.9	12.9	13.0	13.0	13.1	13.1	13.1	13.2
	15:00 to 16:00	10.2	10.1	9.9	10.0	10.1	10.1	10.3	10.4	10.5	10.4
	16:00 to 17:00	13.0	13.1	13.2	13.4	13.5	13.6	13.7	13.9	14.0	14.1
	Average	12.0	12.0	12.0	12.1	12.2	12.2	12.4	12.5	12.5	12.6
Air temperature	14:00 to 15:00	10.9	11.0	10.9	11.0	11.0	11.0	11.2	11.3	11.3	11.3
	15:00 to 16:00	8.5	8.4	8.4	8.4	8.4	8.4	8.5	8.6	8.6	8.6
	16:00 to 17:00	8.8	8.8	8.8	8.8	8.9	8.9	9.1	9.1	9.2	9.4
	Average	9.4	9.4	9.4	9.4	9.5	9.5	9.6	9.7	9.7	9.7
Air humidity	14:00 to 15:00	46.0	45.7	45.6	45.8	45.7	45.8	46.0	46.7	46.5	46.9
	15:00 to 16:00	61.4	61.3	61.3	61.7	61.8	61.8	62.5	62.8	63.3	64.1
	16:00 to 17:00	45.3	45.5	45.6	46.5	47.0	47.5	49.3	51.5	51.0	52.5
	Average	50.9	50.8	50.8	51.3	51.5	51.7	52.6	53.7	53.6	54.5

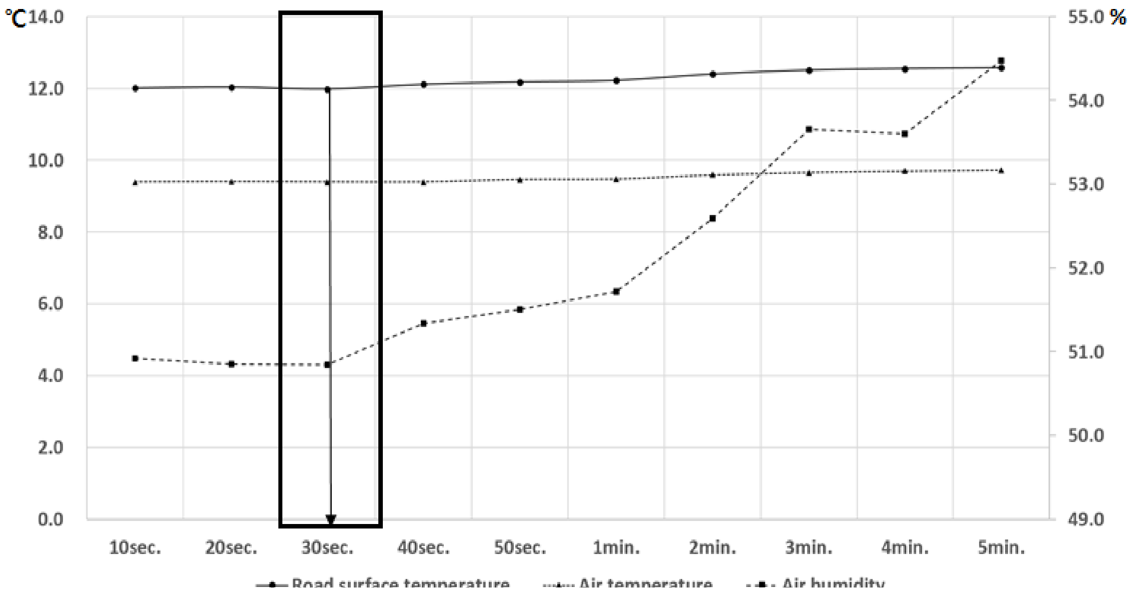

Figure 3: Average MSE by aggregation interval

V. CONCLUSIONS

In this study, the optimal time interval of individual vehicle sensing data using MSE was evaluated to provide microscopic road environment information. As a result, the appropriate aggregation interval was calculated to be 30 s. In the future, it is expected that more detailed information of the driving environment (road icing, bad weather, traffic congestion, etc.) will be provided than by the data aggregation unit provided by existing systems. Currently, a driving environment analysis platform for collecting various road environment data is being developed, and vehicle sensors capable of measuring only some data (atmospheric temperature, atmospheric humidity, road surface temperature) among all vehicle data have been installed in two experimental vehicles. Using the two test vehicles, vehicle sensing data were collected for a specific time zone and test road section. In the future, it should be possible to perform a comprehensive analysis using vehicle sensing data collected from a large number of vehicles at various intervals and time zones. This study represents a preliminary study for determining the proper aggregation interval. In the future, related research such as an estimation method for the population of the vehicle sensing data and testing of the identity between the estimated population and the sample group should be continued. In addition, it is necessary to compare the method with methodologies other than the method of determining the appropriate aggregation interval generally applied in the previous studies.

ACKNOWLEDGMENT

This research was supported by a grant from a fusion research project (Development of Driving Environment Observation, Prediction and Safety Technology based on Automotive Sensors) funded by the Korea Institute of Civil Engineering and Building Technology.

REFERENCES

[1]. The TAAS (Traffic Accident Analysis System) website. [Online]. Available: http://taas.koroad.or.kr/

[2]. The KTDB (Korea Transport DataBase) website. [Online]. Available: https://www.ktdb.go.kr/

[3]. I. Jung, K. Chong, "Development of Information Technology Infrastructures through Construction of Big Data Platform for Road Driving Environment Analysis", Journal of the Korea Academia-Industrial cooperation Society, vol. 19, no. 3, pp. 669-678, 2018.

[4]. Byron J. Gajewski, Shawn M. Turner, William L. Eisele and Clifford H. Spiegelman, "Intelligent Transportation System Data Archiving: Statistical Techniques for Determining Optimal Aggregation Widths for Inductive Loop Detector Speed Data", Transportation Research Record: Journal of the Transportation Research Board, vol. 1719, pp. 85-93, 2000.

[5]. S. Yoo, J. No, D. Park, "Investigating Optimal Aggregation Interval Size of Loop Detector Data for Travel Time", Journal of Korean Society of Transportation, vol. 22, no. 6, pp. 109-118, 2004.

[6]. D. Park, S. You, J. Rho, H. Cho, K. Lee, "Investigating optimal aggregation interval sizes of loop detector data for freeway travel-time estimation and prediction", vol. 36, no. 4, pp. 580-591, 2009.

[7]. D. Choi, "Determine Optimal Aggregation Interval for Travel Time Estimation on Rural Interrupted Traffic Flow Considering Traffic Flow Characteristics", University of Science and Technology, Master thesis, 2012.

[8]. H. Lim, S. Lee, H. Lee, "A Study on the Optimal Aggregation Interval for Travel Time Estimation on the Rural Arterial Interrupted Traffic Flow", The Journal of The Korea Institute of Intelligent Transport Systems, vol. 3, no. 2, pp. 129-140, 2004.

[9]. H. Lim, S. Lee, H. Lee, "A Comparative Study on the Statistical Methodology to Determine the Optimal Aggregation Interval for Travel Time Estimation of the Interrupted Traffic Flow", Journal of Korean Society of Transportation, vol. 23, no. 3, pp. 109-123, 2005.

[10]. S. Shim, "An Expressway Path Travel Time Estimation Model Using DRSC Travel Data", Ajou University, Doctoral thesis, 2011.

[11]. S. Shim, K. Choi, S. Lee, S. Namkoong, "An Expressway Path Travel Time Estimation Using Hi-pass DSRC Off-line Travel Data", Journal of Korean Society of Transportation, Vol. 31, No. 3, pp.45-54, 2013.

Intaek Jung, "A study on determination of appropriate aggregation interval to provide driving environment information using vehicle sensing data" International Journal Of Engineering Research And Development , vol. 14, no. 07, 2018, pp. 63-69