Outlier Removal Technique for Estimating Link Travel Times Using DSRC Probes

Jinhwan Jang

Highway Research Department, Korea Institute of Civil Engineering and Building Technology
283, Goyangdae-Ro, Illsanseo-Gu, Goyang-Si, Gyeonggi-Do, 10223, South Korea

jhjang@kict.re.kr

Abstract- As electronic toll collection systems based on 5.8 GHz dedicated short-range communications (DSRC) have become popular, real-life travel time collection via DSRC systems is attracting more interest. The DSRC system, composed of an on-board unit and a road-side unit, has the advantage of gathering section travel times directly. However, outliers resulting from vehicles entering or exiting a section, parking activities, and so on have been considered a critical hurdle to be overcome in estimating reliable travel times. In this study, a modified Ferguson test is applied to clean DSRC probe data on a signalized rural highway in Korea. The proposed method was tested with 24-h block real-world data, indicating satisfactory results.

Keywords- travel time estimaiton, DSRC, probe data, outlier

I. INTRODUCTION

When roads are congested, drivers want to find the fastest routes to their destinations, while road managers are eager to search for the most efficient way to handle congestion. In Korea, the annual cost of congestion amounts to $25 billion and the resulting air pollution is of great concern to the public (Baik et al. 2009). To efficiently alleviate the problem, an advanced traveler information system (ATIS) has been steadily deployed across the nation since it was first introduced in early 1990s. As of now, around 30 cities have been equipped with ATIS devices─traffic detectors, closed circuit televisions (CCTV), variable message signs, and so on (Jang 2013). In an ATIS environment, drivers can choose the shortest route to their destinations based on the reliable real-life travel-time information provided.

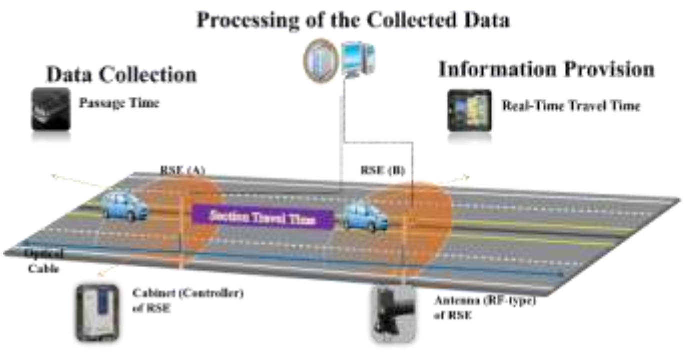

Fig. 1 Schematic of DSRC traffic information system

To this end, it is imperative that the provided travel-time information be reliable. Conventionally, traffic detectors have been installed to estimate travel time information. However, with the popularity of electronic toll collection systems (ETCSs), dedicated short-range communication (DSRC) probes have been considered more efficient to gather link travel time information. As of 2016, approximately 70% of vehicles on freeways are equipped with DSRC on-board units (OBUs). Traffic managers only need to install road-side units (RSUs) to scan unique OBU identifications. As DSRC probes can directly collect section travel times, they have been considered superior to conventional detectors, with which estimation of section travel time is necessary and generally known to be erroneous (Zhang et al. 2013). Fig. 1 shows the schematic of a widely deployed DSRC traffic information system in Korea.

In a DSRC system, however, outliers that deviate markedly from valid observations are frequently observed, as shown in Fig. 2, mainly due to entry and exit maneuvers between corridors, parking activities, and so on. Sometimes, a car is detected in one direction and redetected in the opposite direction sometime later, usually on the way back. In this case, the first observation produces the correct travel time, but the later one generates an abnormally long travel time. In old systems, time synchronization between two consecutive RSUs is not performed properly, causing substantial outliers.

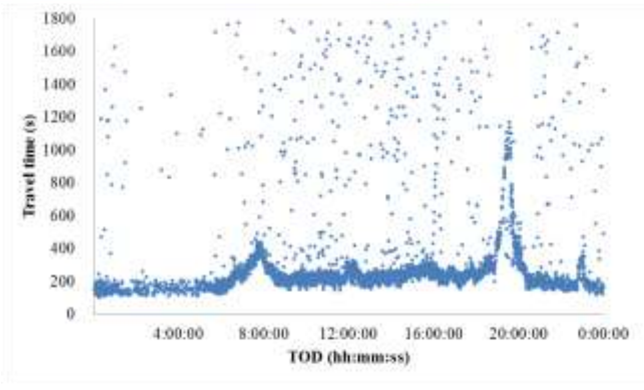

Fig. 2 Raw DSRC probe travel times

To tackle this outlier issue of DSRC probes, a novel method is proposed to censor outliers contained in raw DSRC probe data out. To take DSRC probe outlier characteristics account, the Ferguson test was initially applied, but sometimes it cannot filter out apparent outliers in certain situations where many outliers exist. To resolve the deficiency of the Ferguson test, it was supplemented with a validity window logic. The method was evaluated with real-world DSRC probe data, revealing a satisfactory outcome.

II. LITERATURE REVIEW

Some previous studies to estimate reliable travel times using probe-based systems have been performed. Southwest Research

Institute (SwRI 1998) developed the TransGuide algorithm, which eliminates outliers outside the predefined valid range. Dion et al. (2006) proposed a technique to reliably estimate link travel time. Their algorithm is based on the assumption that probe travel time follows a log-normal distribution. Ma et al. (2010) proposed a median filter approach that uses the median as a measure of location. ITS Korea et al. (2008) developed a method that uses a simple confidence interval concept with an assumption that probe travel time follows a normal distribution. Boxel et al. (2011) developed an innovative method for removing probe data gathered using Bluetooth scanners. They used the confidence interval concept (Kendall et al. 1973) using the standard residual of Greenshield's speed-density model. They used the least median of squares to estimate parameters in the model to prevent it from being compensated with outliers. Clark et al. (2002) suggested a statistical method for cleaning outlying observations included in probe travel time data gathered from license number plates. The traditional confidence interval concept with the median and quartile deviation was used. Jang (2016) developed algorithms for filtering outliers in DSRC probe travel times on signalized rural arterials. His algorithm considered low sample size situations where previous methods were not proven to be effectively operated.

III. Proposed Methodology for Cleaning DSRC Probe Data

A. *Ferguson Statistical Test*

The Ferguson test is one of the statistical tests for identifying outliers. Most statistical tests─Dixon, Tietjen-Moore, t-tests, and so on─detect outliers only once with their statistics. Hence, some outliers could be masked by abnormally aberrant values, which result in a shift in the mean (or level) and variance (or scale). Consequently, the seemingly outlying observations cannot be discerned and cannot be filtered out (ASTM 2008). To address this problem, Ferguson (1969) proposed the sample coefficient of skewness for one-tailed tests. If the computed for an observed value exceeds the values listed in Table 1, the observation farthest from the mean is removed and the iteration process continues until no further observations are determined as aberrant values.

$$\sqrt{b_1} = \sqrt{n}\sum_{i=1}^{n}(x_i-\bar{x})^3/[\sum(x_i-\bar{x})^2]^{3/2}, \text{ one-tailed} \quad (1)$$

where, $\sqrt{b_1}$ = statistic, x_i *and* $\bar{x}$ = observed values and the mean

Table 1 Ferguson Test Statistic (One-Sided)

Statistic	Significance level	Number of samples (n)					
		5	10	15	20	25	50
$\sqrt{b_1}$	1%	1.34	1.31	1.20	1.11	1.06	0.79
	5%	1.05	0.92	0.84	0.79	0.71	0.53

B. *Evaluating the Ferguson Test*

The Ferguson test is applied to the 24-h block DSRC probe data illustrated in Fig. 2. The test data were retrieved from a database of a rural highway management system in Korea. The road section on which this study is based spans around 3 km and includes one signalized intersection and one interchange.

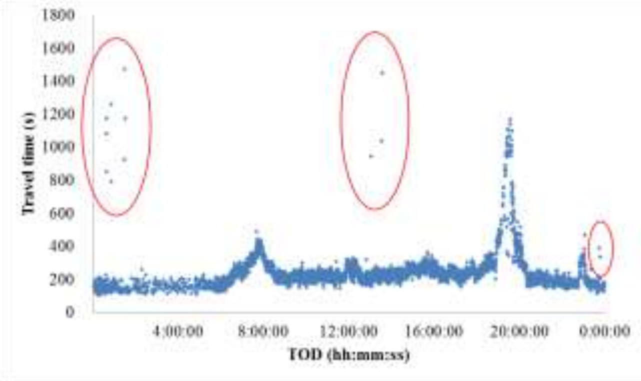

Fig. 3 Probe travel times after application of the Ferguson test

In most cases, it shows good performance. However, the red circles marked in Fig. 3 show apparent outliers remaining after the test was applied. If many outliers are observed in a collection interval, they make the mean deviate markedly from valid observations. The deviated mean, in turn, generates a low value of the statistic ($\sqrt{b_1}$) due to a compensation effect between negative values (from valid observations) and positive ones (from aberrant observations).

C. *Adding Validity Window to Ferguson Test*

To resolve the above-mentioned deficiency revealed by the Ferguson test, it is supplemented with the concept of a validity window. The validity window is determined using the average of the filtered travel times in the immediately preceding collection interval and a window parameter (α). The validity window concept based on previous interval values is generally known to have disadvantages under conditions of sudden travel time fluctuations during morning peak hours or periods of intermittent system failure (Dion et al. 2006; Jang 2016). However, combined with the Ferguson test, the disadvantages can be effectively surmounted.

$$T_{AB}(t) = \frac{\sum_i (t_{B,i}-t_{A,i})}{N(S_{AB}(t))}, \qquad \text{where } i \in S_{AB}(t) \quad (2)$$

$$S_{AB}(t) \equiv \{k \mid t - T_w < t_{B,k} \le t\} \cap \left\{m \,\middle|\, \left|\frac{t_{B,m}-t_{A,m}-T_{AB}(t-T_w)}{T_{AB}(t-T_w)}\right| \le \alpha\right\} \quad (3)$$

Where, n = number of samples in a 5-min block of travel times, $T_{AB}(t)$ = average of valid travel times from A to B at time t, $S_{AB}(t)$ = set of valid travel times from A to B at time t, $t_{A,i\ (\text{or } m)}$ = detection time of vehicle i (or m) at point A, $t_{B,i\ (\text{or } m)}$ = detection time of vehicle i (or m) at point B, T_w = collection (or aggregation) interval, α = parameter

To estimate the parameter (α), a statistical confidence interval concept was applied to travel time data in the same section, which were cleaned manually based on the CCTV images installed between the consecutive RSUs. To make a confidence interval valid, the normality assumption should be verified in advance. To test the normality assumption of the cleaned travel times, the K-S statistic was used as shown in Table 2; as a result, the assumption was proven to be valid and a value corresponding to a 99.9% confidence interval was applied. Fig. 4 shows the histogram of the parameter (α), superimposed by the normal curve.

Table 2 Normality test and applied values for parameters

Para-meter	Number of sample	Estimated parameters	K-S statistic for normality test at significance level of 0.05 (p-value)	Maxi-mum	Applied value
Alpha (α)	415	μ = –0.01, σ = 0.12	0.79	0.29	0.35

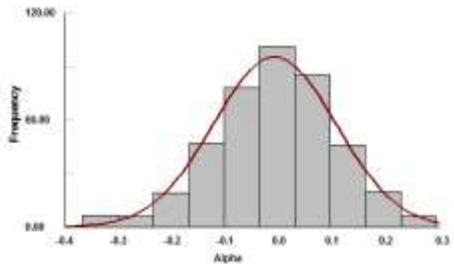

Fig. 4 Histogram of the parameter (α) superimposed by the normal curve

The whole probe data cleaning procedure is represented in Fig. 5. After matching the probe data with a unique coded OBU ID, the minimum sample size required to apply the Ferguson test is checked, and then the iteration process is performed based on the test. After terminating the process, a validity window set by previous interval values checks the individual travel times filtered by the Ferguson test.

Fig. 5 Proposed probe travel time cleaning procedure

Fig. 6 illustrates the result of application of the suggested algorithm and shows that the seemingly apparent outlying observations in Fig. 3 were thoroughly removed. Due to budget constraints, baseline travel times against which the filtered travel times could be quantitatively evaluated could not be obtained. Even though the evaluation is somewhat qualitative, the cleaned probe travel times represented the typical travel time pattern of the section well.

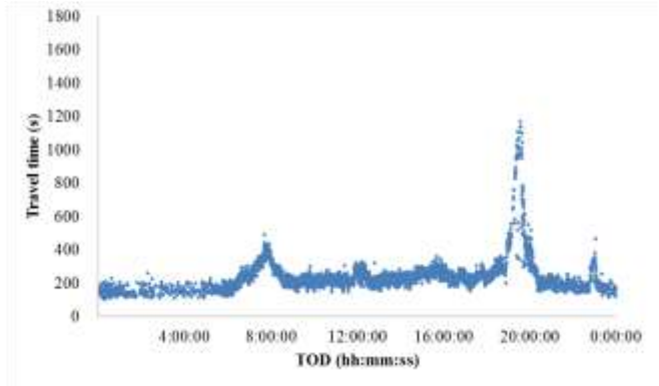

Fig. 6 Probe travel times after application of the proposed algorithm

The filtered data can be used for multiple purposes. The primary application area would be real-time traveler information systems by aggregating them into a certain time interval (e.g. 5 min). Also, travel time monitoring systems for gathering fundamental traffic data for road planning, design, operations, and maintenance could gain benefit by employing the developed method to generate reliable travel time data.

IV. CONCLUSIONS

In this study, an outlier treatment method for DSRC probe travel time on a signalized rural highway was presented. The DSRC system is generally known to gather point-to-point travel times efficiently compared to conventional point detectors, but without delicate treatment of outlying observations, gathered travel times could be useless. In this regard, a cleaning process is essential for obtaining reliable travel time information in DSRC-based ATIS.

The filtering procedure developed in this study uses the Ferguson test, which can remove outliers repeatedly. However, after application of the test, some outliers were still observed. To make up for the weak point of the Ferguson test, a validity window that was predetermined with valid travel times in the immediately preceding collection interval was applied. Consequently, the proposed procedure was qualitatively proved to have a satisfactory outcome. However, only a single set of 24-h block data was used for verification of the algorithm. Hence, more comprehensive evaluations with abundant data need to be conducted to verify the spatiotemporal transferability of the developed method. Also, to maximize the driver's utility, prediction techniques for generating future travel times need to be considered.

ACKNOWLEDGEMENT

This work was supported by a grant from the Korea Agency for Infrastructure Technology Advancement (KAIA) (No. 16TBIP-C111209-01).

REFERENCES

[1] ASTM International (2008). Standard Practice for Dealing with Outlying Observations, Designation: E 178-08.

[2] D. V. Boxel, W. H. Schneider IV, and C. Bakula (2011). "An Innovative Real-Time Methodology for Detecting Travel Time Outliers on Interstate

Highways and Urban Arterials." TRB 2011 Annual Meeting, CD-ROM, Washington D.C.

[3] F. Dion and H. Rakha (2006). "Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data." Transportation Research Part B, Elsevier.

[4] ITS Korea, Hitecom System, and Aju University (2008). Development of Practical Technology for DSRC Traffic Information System, Korea Expressway Corporation (in Korean).

[5] J. Jang (2013). "A Review on Section Data Cleaning Methods." Transportation Technology and Policy, vol. 10, No. 2, Korean Society of Transportation (in Korean).

[6] J. Jang (2016). "An Outlier Filtering Algorithm for Travel time Estimation using Dedicated Short Range Communications Probes on Rural Highway." IET Intelligent Transport Systems, Online ISSN 1751-9578.

[7] M. G. Kendall and A. Stuart (1973). The Advanced Theory of Statistics, Volume 1: Distribution Theory, Edward Arnold, London.

[8] N. Baik, B. Kim, J. Jang, Y. Go, and B. Lee (2001). Technical Development for Advancement of Hazardous Roadway Information Service under Inclement Weather. Korea Institute of Construction Technology (in Korean).

[9] S. D. Clark, S. Grant-Muller, and H. Chen (2002). "Cleaning of Matched License Plate Data." Transportation Research Record, No. 1804.

[10] Southwest Research Institute (1998). Automatic Vehicle Identification Model Deployment Initiative–System Design Document, Texas Department of Transportation.

[11] T. S. Ferguson (1961). "Rules for Rejection of Outliers." Revue Inst. Int. de Stat. (RINSA), Vol. 29, Issue 3.

[12] X. Ma and H. Koutsopoulos (2010). "Estimation of the Automatic Vehicle Identification Based Spatial Travel Time Information Collected in Stockholm." IET Intelligent Transport Systems, Vol. 4, Iss. 4.