sustainability

MDPI

Article

Classification of Speed Change and Unstable Flow Segments Using Geohash-Encoded Vehicle Big Data

Kyu Soo Chong

Korea Institute of Civil Engineering and Building Technology, 283 Goyang-daero, Ilsanseo-gu, Goyang-si 10223, Gyeonggi-do, Republic of Korea; ksc@kict.re.kr; Tel.: +82-31-910-0652

Abstract: Precise and detailed speed information is indispensable for ensuring safe and efficient transportation. This is particularly true within unstable flow (UF) segments, which are especially prone to accidents due to the significant speed variations between vehicles and across lanes, and in the context of evolving transportation systems, where autonomous and non-autonomous vehicles are increasingly mixing. To address the limitations of existing methods in providing such data, this study aims to improve the detail, accuracy, and granularity of road information for micro-segments by leveraging individual vehicle big data. The proposed approach utilizes the geohash algorithm for spatial segmentation and introduces a novel criterion for identifying UF segments based on the relationship between space mean speed (SMS) and time mean speed (TMS). The presented strategy was validated through a comprehensive analysis of DTG (Digital Tachograph) data from freight vehicles on Expressway No. 50 in the Gyeonggi region in the Republic of Korea. As a result, a total of 301 segments were identified, including 178 eastbound and 123 westbound segments. UF segments corresponded to partitions falling beyond the reference standard deviation range. Compared with VDS (Vehicle Detection System) and conzone speeds, the proposed method provided more precise and continuous speed information, surpassing those obtained from conventional link-based approaches.

Keywords: geohash; space mean speed; time mean speed; unstable segment; speed deviation

Citation: Chong, K.S. Classification of Speed Change and Unstable Flow Segments Using Geohash-Encoded Vehicle Big Data. *Sustainability* **2023**, *15*, 14684. https://doi.org/10.3390/su152014684

Academic Editors: James J.Q. Yu, Shuai Wang, Bo Fan, Shiyao Zhang and Christos Markos

Received: 10 August 2023
Revised: 18 September 2023
Accepted: 5 October 2023
Published: 10 October 2023

1. Introduction

By the year 2022, the Republic of Korea witnessed a total of 196,836 traffic accidents, resulting in 2735 fatalities. Alarmingly, on expressways alone, the accident count reached 4860, with 184 lives lost—nearly 2.7 times the fatality rate compared with other roadways. Because of the median barriers featured by expressways, most of those accidents were side or rear-end collisions (1207 and 1698, respectively), whereas only 40 corresponded to head-on collisions, as shown in Table 1. A significant contributor to accidents is unstable flow (UF) caused by various interferences affecting traffic movement, such as interactions at entrance and exit ramps and accidents, which lead to rapid speed reductions as traffic volume escalates. The prevalence of UF across continuous flow zones like highways underscores its potential role in traffic accidents, necessitating the classification of road segments based on their traffic flow characteristics.

Currently, traffic information segments (a unit of the road network that provides speed information) are established on nodes and links derived from the road's central axis. Nodes and links are generated at locations such as road intersections, origin and destination points, traffic control sections, points of structural or road management changes, administrative zone transitions, and entry/exit junctures. However, due to these static links, traffic information segments do not correspond to actual speed alteration segments or indicate speed attributes, i.e., despite internal speed differentials within the same segment, they are presented as a singular data entity.

Table 1. Traffic accident statistics in 2022.

Accident Type	Road Type	Number of Accidents	Number of Deaths	Number of Injured
Head-on collision	National highway	7849	110	12,590
	Provincial road	5376	88	8707
	Metropolitan city road	28,854	112	42,697
	City road	25,989	207	38,946
	County road	3568	48	5587
	Expressway	1333	10	2374
	Unknown	3588	15	5255
Rear-end collision	National highway	4521	66	8403
	Provincial road	2052	27	3736
	Metropolitan city road	11,917	48	19,860
	City road	8820	69	15,075
	County road	694	16	1176
	Expressway	1943	100	4852
	Unknown	1257	3	2091
Etc./unknown	National highway	4729	41	7093
	Provincial road	2171	18	3301
	Metropolitan city road	17,893	75	25,026
	City road	13,276	69	18,890
	County road	739	8	1097
	Expressway	1374	27	2775
	Unknown	3733	14	5191

The rapid advancement of autonomous vehicles and Cooperative Intelligent Transportation Systems (C-ITS) is expected to transform the road/traffic environment into a complex amalgamation of autonomous and non-autonomous vehicles in the near future. This underscores the critical importance of ensuring precision and accuracy in the surrounding vehicle attribute data (such as location and speed) obtained through Vehicle-to-Everything (V2X) communication [1]. Notably, the acquisition of precise micro-segmented speed data is considered a pivotal factor for ensuring safe and comfortable vehicle operation, facilitating autonomous driving, and seamless C-ITS integration in environments where autonomous and non-autonomous vehicles coexist [2]. The provision of precise and reliable information on the speed of proximate vehicles within a relatively large spatial domain is particularly critical to ensuring safety when implementing autonomous driving with high-speed vehicles, given the reduced reaction time available for unforeseen circumstances [3].

The present study utilizes substantial Digital Tachograph (DTG) data—big data containing location information acquired from individual vehicles—to identify and expound upon UF occurrences, such as instances of speed alterations and segments characterized by velocity disparities between lanes. We introduce discerning criteria for segregating segments characterized by uneven speed distribution, employing space mean speed (SMS) to detect both intra- and inter-vehicle speed differences within a single segment. Furthermore, to provide detailed traffic information, we divide UF segments where speed differences occurred by lane.

2. Literature Review

A junction is a location where vehicles merge and sort, prompted by competitive conflicts arising from divergent traffic demands. Primarily established to link the traffic streams of an expressway and a standard route, this segment exhibits a higher frequency of lane changes compared with the main expressway artery, resulting in inherently unstable traffic flow, heightened congestion, and a higher propensity for accidents. The most common type of connections are entrance and exit ramps; the impact area is calculated based on the junction's geometric design, road alignment, and lane count, whereas service levels are determined based on traffic volume [4]. Efforts to mitigate UF have encompassed the integration of Variable Message Signs (VMS) and density metrics within a congestion

prediction model that leverages the cumulative moving average of vehicle speed arrays [5]. A vehicle trajectory model rooted in Vehicle Detection System (VDS) data was used in the construction of a dynamic Origin–Destination (O-D) prediction model to explore congested traffic flow dynamics [6]. Factors that can explain lane interdependencies were identified in the study "Development of Impulse Propagation Model between Lanes through Temporal-Spatial Analysis" [7]. This led to the development of a lane impact propagation model, employing multiple regression analysis to quantify impact based on the analyzed segment and lane. Notably, prior research grappled with the challenge of precisely timing UF instances through the examination of detector data from junctions, such as entrance and exit ramps.

Given the inherent variation in speeds within a traffic stream, an accurate single depiction of the speed of a traffic stream necessitates understanding speed distribution, typically approximated by a normal distribution. The average speed is usually divided into two components: time mean speed (TMS) and SMS. TMS represents the arithmetic mean of speeds exhibited by vehicles traversing a point or short road segment over a defined interval. On the other hand, SMS quantifies the sum of all vehicles passing through a specific segment within a specific period, divided by the distance traveled [8]. Consequently, the speed of vehicles in a specific segment is expressed by the relationship between TMS and SMS, with UF surfacing when individual vehicle speeds diverge from a given value.

Because they are fixed on individual vehicles and provide crucial information about their operation, DTG data emerge as a promising resource for generating finer-grained micro-speed information [9]. The widespread integration of diverse sensors within vehicles has ushered in an era of location-based big data, sourced from individual vehicle sensors. These data assume a pivotal role by not only contributing to road and traffic information but also offering insights into the operational conditions of neighboring vehicles [10]. Particularly, DTG data based on satellite positioning systems differ from VDS data collected through devices such as loop detectors and Closed-Circuit Television (CCTV) due to their ability to furnish copious amounts of data, encompassing vehicle locations and instantaneous speeds. While VDS data bear reliability, their utility is constrained by their focus on point information, thereby limiting their ability to capture the dynamic behaviors of vehicles characterized by frequent accelerations and decelerations [11]. Moreover, the generation of diverse traffic information faces spatial limitations, potentially compromising the reliability of the yielded data depending on the traffic environment [12].

The expeditious processing of DTG big data is an imperative facet in collating and presenting micro-segment speed information of vehicles. In this regard, the velocity in processing location-based big data is the most important aspect in deciding whether big data-driven road and traffic information can be delivered in real time [13]. Notably, the potency and rapidity of spatial analysis exert a commanding influence on the processing of location-based big data. However, the present approach of spatial analysis, entailing the connection of current point data with link data, proves inadequate for real-time data provisioning due to its requirement of map matching technology. This arises from the necessity to precisely map coordinates onto the geographical layout and subsequently match them with the requisite buffer area encompassing surrounding links. Moreover, depending on the distance between data points and reference points, varied matching strategies—ranging from point-to-point and point-to-curve to curve-to-curve approaches—to align individual vehicle big data with map data may lead to considerable errors [14,15]. Therefore, providing accurate speed information based on DTG big data hinges on a comprehensive assessment of accurate and efficient matching procedures between spatial orientation and data representation.

Various methodologies have been proposed for generating accurate vehicle speed data. Han [16] examined methodologies for calculating SMS on expressways, utilizing vehicle trajectory data from radar detectors. Ko [17] discussed the challenges of continuous micro-traffic analysis, contrasting prevalent image-based traffic analysis approaches with a microscopic traffic analysis framework using drone-acquired photographs. These aerial

images facilitated the extraction of essential metrics, including speed, traffic volume, and density. Liu [18] employed GPS data to compute SMS for node links, elucidating the nexus between speed and vehicle collision incidents along a segment of the Shanghai trunk route.

While the aforementioned studies have contributed commendable methodologies for generating accurate speed data pertinent to their respective datasets, certain limitations persist. First, data procured from roadside stations, loop detectors, CCTV, and probe cars operate within a static link framework relative to the road's central axis. This approach, however, may fall short of representing genuine speed variations or capturing speed homogenous segments characterized by distinctive speed attributes. Fixed sensors like radar and loop detectors solely furnish speed information at designated points, neglecting the broader spatiotemporal distribution of speed [19]. Second, speed data emanating from node links cannot adequately reflect micro-segment attributes. Lastly, methodologies reliant on mobile platform videos face difficulties in acquiring continuous, relevant data and maintaining constant observation [17].

Numerous investigations have explored the geographic alignment of vehicle big data based on Global Navigation Satellite Systems (GNSS). Zhao [20] mapped GPS data on a 60 m grid to predict vehicle speeds. Ibarra-Espinosa [21] generated regional model data for traffic flow and speed, mapping vehicle GPS data on a fixed grid. Liu [18] studied spatial patterns of urban traffic congestion through a grid-based analysis of node capacity shortages, bottlenecks, and intersection congestion trends. Nonetheless, because these existing studies rely on a fixed grid size, their accuracy may vary depending on the grid configuration. Moreover, reducing grid size to enhance precision might inadvertently delay data processing.

The central aim of this study was to create dynamic homogeneous speed zones from DTG data and to use them to facilitate the generation of speed information pertinent to micro-segments. These homogeneous speed zones were determined by comparing TMS and SMS. Typically, drivers are presented with an average speed for a designated segment, which is further partitioned into TMS and SMS. TMS and SMS represent, respectively, the arithmetic mean of vehicle speeds traversing a point or small road segment within a specified timeframe, and the quotient of speed by the distance covered by all vehicles transiting a road segment during the same period [8]. Usually, TMS is lower than SMS and both would be equal if all vehicles within a segment maintained an identical pace. Thus, the similarity between TMS and SMS serves as a metric for assessing the uniformity of vehicle speeds within a given segment. The present study evaluates TMS-SMS congruence based on the correlation between the two speeds. Past research has drawn upon empirical analyses to elucidate the relationship between TMS and SMS in the context of vehicles on expressways or national highways [22]. Garber [23] established an empirical linear linkage between TMS and SMS within expressway settings. Leong [24] also determined a linear relationship between these two speeds on a Malaysian national highway.

3. Methodology

3.1. Determination of Unstable Flow

SMS and TMS are equal if all vehicles in the same space have the same speed. However, divergent individual vehicle speeds introduce discernible differentials between these two means, as illustrated in Figure 1. The relationship between these mean speeds is mathematically expressed in Equation (1). Employing the ratio of the square of TMS speed's standard deviation to TMS itself offers an effective measure to quantify the dissimilarity between these mean speeds. This quantification is particularly informative when speed variations are uniformly spread across a designated space. In such instances, the extent of speed variability among vehicles may fluctuate, as dictated by the variance denoted as σ'. Notably, the range of Speed Fluctuation (SF) per speed resides within the confines defined by the function of the two mean speeds, as described by Equation (2). This depiction is graphically elucidated in Figure 2.

$$SMS = TMS - \sigma^2_{TMS}/TMS \quad (1)$$

$$SMS = TMS - \sigma'^2/TMS \quad (2)$$

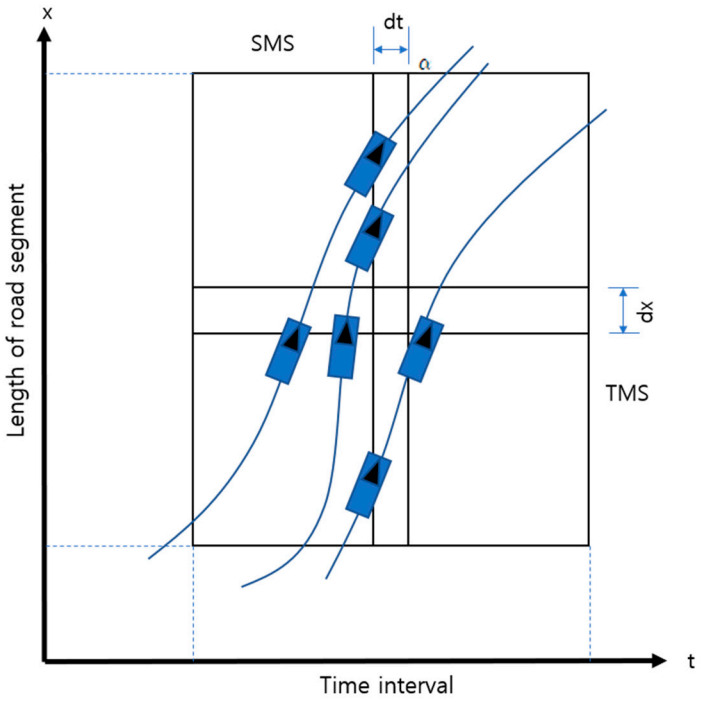


Figure 1. Schematic of TMS and SMS concepts.

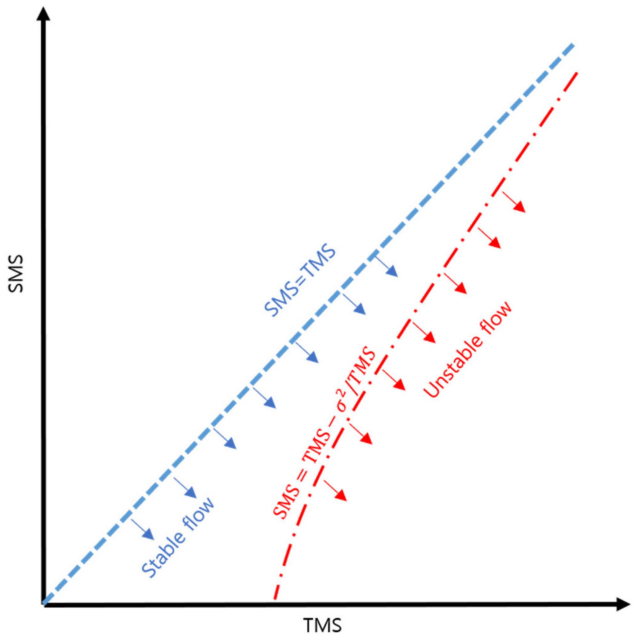


Figure 2. Relationship between SMS, TMS, and unstable flow concepts.

3.2. Establishment of Unstable Flow Boundaries

UF manifests as a significant speed discrepancy among vehicles transiting a brief segment. This divergence in speed precipitates abrupt decelerations, in turn affecting trailing vehicles. In this context, the standard deviation, as outlined in Equation (3), serves as the pivotal demarcation for speed deviations. The crux of this boundary definition lies in identifying segments characterized by akin actual driving speeds. To expound upon

this, Figure 3 offers a visualization of DTG data pertinent to Expressway no. 50, specifically focusing on the right-handed upbound data stream. The calculation of the speed's standard variation was executed through link data analysis across eight distinct straight or curved segments, each devoid of entrance and exit ramp junctions—a configuration summarized in Table 2. The resulting standard deviation was derived from a sample of six segments tallied at 9.45 kph, thus serving as the benchmark standard deviation value for this investigation using Figure 4.

$$\mathrm{SMS} = \mathrm{TMS} - 89/\mathrm{TMS} \quad (3)$$

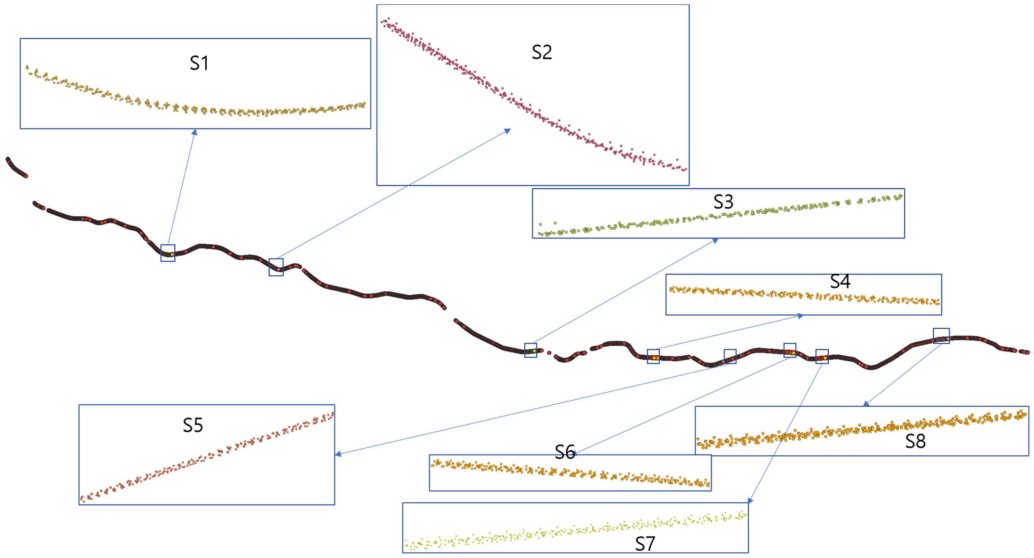


Figure 3. Identifying segments characterized by akin actual driving speeds.

Table 2. A configuration summarized.

Statistics	S1	S2	S3	S4	S5	S6	S7	S8	Average
Number	280	342	202	195	219	275	299	121	241
Sum	23,507	27,657	16,885	17,631	19,003	25,488	28,132	11,176	21,185
Average	83.95	80.87	83.59	90.42	86.77	92.68	94.09	92.36	88.09
Median	81	81	83	91	87	87	92	93	86.88
Standard deviation (population)	7.60	8.29	8.61	11.41	9.71	9.34	9.37	11.25	9.45
Standard deviation (samples)	7.61	8.30	8.63	11.44	9.73	9.35	9.38	11.29	9.47
Minimum	73	66	63	71	73	78	80	77	72.63
Maximum	108	103	109	108	108	109	109	109	107.88
Range	35	37	46	37	35	31	29	32	35.25
Minority	106	99	69	83	93	89	84	97	90.00
Majority	81	86	77	81	87	87	87	109	86.88
Variety	27	36	32	27	31	26	29	17	28.13

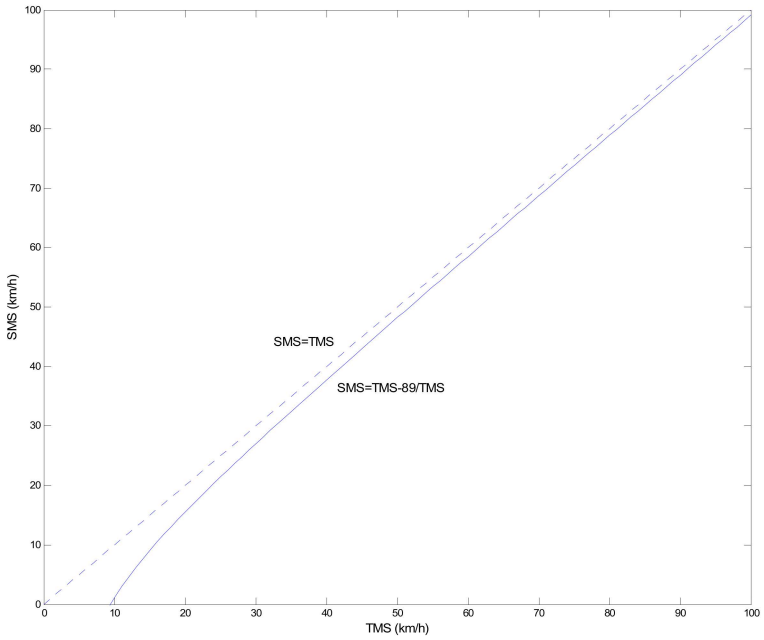


Figure 4. Establishment of unstable flow boundaries.

3.3. Dynamic Segmentation Method

In this study, the process of spatial segmentation at the study location was executed using the geohash algorithm. The geohash algorithm, a grid-based and string encoding technique, generates a hierarchical grid map by iteratively quartering (2×2) the target region. Figure 5 illustrates the fundamental components of the geohash algorithm. The black numerals within the figure depict the code assigned to the initial partitioned area, while the blue, green, and red letters signify grid code numbers corresponding to the second, third, and fourth subdivisions, respectively. Employing the geohash algorithm, the study partitioned the geographic expanse of the Republic of Korea, encompassing longitudes 126° to 130° and latitudes 34° to 38°.

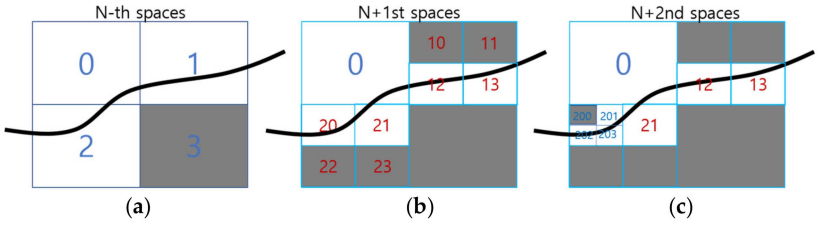


Figure 5. Geohash-based space segmentation method. (**a**) N-th (**b**) N+1st (**c**) N+2nd spaces.

Dynamic segmentation, a pivotal facet of this study, orchestrates the hierarchy of segmentation predicated on UF determination. To elaborate, the partitioning process hinges on the interplay between TMS and SMS, dictating the UF categorization for each grid. In essence, grids identified as SF halt further partitioning, while non-SF grids undergo continued subdivision. This dynamic segmentation methodology is comprehensively depicted in Figure 5. The initial step, depicted in Figure 5a, entails the division of the grid along both latitude and longitude axes, with sequential codes 0, 1, 2, and 3 allocated. Grid 3, devoid of road passages, is excluded from this partitioning. Similarly, Figure 5b displays a parallel division, with space 1 from Figure 5a now labeled as 10, 11, 12, and 13. To maintain granularity while averting excessive coverage or exceedingly diminutive road lengths, the study instituted a minimum of 5 partitions and a maximum of 12. This selection rationale balances the need to avoid overly extensive regions (fewer than 5 partitions) and the requirement for road segments of adequate length (more than 12 partitions) to furnish

micro-segment speed data for expressways. As earlier outlined, the partitioning process follows the pre-determined UF categories. Notably, space 0 qualifies as SF, while spaces 1 and 2 are identified as UF, as illustrated in Figure 6a. The ensuing space segmentation follows the pattern outlined in Figure 6b. Upon determining space no. 20 as UF, partitioning proceeds accordingly. When the partition count reaches 12, space 203 again meets the criteria for UF classification, as evident in Figure 6c, culminating in the conclusion of further partitioning.

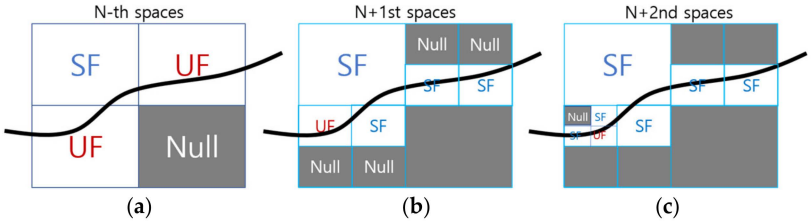


Figure 6. Space segmentation according to the determination of UF. (**a**) N-th (**b**) N+1st (**c**) N+2nd spaces.

3.4. Research Data

The focal point of this study rests upon expressways, and for the purpose of calculating micro-segment speeds, DTG data sourced from freight vehicles and expressway node link data were employed. The study's aim revolves around partitioning lengthy road segments based on variations in vehicle speeds and subsequently generating speed data for each delineated segment. Consequently, emphasis was placed on expressways characterized by extended segment lengths and higher speed limits. This selection rationale stemmed from the prevalence of continuous entrance and exit ramp junctions, accident occurrences, and sporadic instances of congestion—features more pronounced on expressways as compared with those on urban roads. Urban roads, typified by relatively shorter segment lengths, lower speed limits, and fluctuations in traffic flow due to intermittent patterns, were thus excluded from the study's scope.

To effectively capture prototypical vehicle driving behaviors and amass comprehensive position and speed data, DTG data from freight vehicles were harnessed. While DTG data can be culled from taxis, buses, and trucks, this study chose to focus solely on freight vehicles. The rationale behind this exclusion was rooted in the potential divergence in driving patterns exhibited by taxis and buses, largely stemming from their prevalence within urban domains. Notably, taxis and city buses often engage in frequent stops, dictated by factors like passenger embarkation, disembarkation, and service halts. Moreover, the presence of bus-exclusive lanes can introduce distinctive attributes to intercity buses' driving behaviors. In contrast, freight vehicles exhibit unique characteristics, predominantly traversing at lower speeds and frequently occupying the right lane. Aligning with the objective of securing congruence in driving patterns and data availability, this study concentrated on freight vehicle DTG data. A comprehensive overview of the obtained data's scope is enumerated in Table 3.

Data extraction was undertaken based on expressway routes, utilizing the standard node link framework established by the National Transport Information Center [25]. The experimental setup encompassed 951 nodes and 2133 links situated on Expressway no. 50 within the Gyeonggi region. The attributes encompassed in this extraction process are itemized in Table 3.

Table 4 further delineates a comparison between national node links for expressways and Gyeonggi-do regional node links applicable to urban freeways.

Table 3. DTG data definition.

Time	1–30 April 2018	**Region of Interest (Road)**	Expressway No. 50 in Gyeonggi-do
Data Size	Approx. 3.13 TB (Approx. 22.35 billion files, intervals of 1 s)	**Number of Vehicles**	2775
Data Fields	**Field Name**	**Attribute**	**Note**
	T_KEY	Trip key	Encoding
	CARNUM	Car number	Encoding
	SPEED	Vehicle speed	km/h
	WGS84_X	Vehicle location X	Latitude coordinate (WGS84)
	WGS84_Y	Vehicle location Y	Longitude coordinate (WGS84)
	AZIM	GIS azimuth	Degree
	TIME	Time	Date-hour-minute-second (18042809074000)

Table 4. Node-link data definition.

	Expressway		**Data Fields**		
	Node	**Link**	**Field Name**	**Attribute**	**Value**
Nationwide	4501	9835	LINK_ID	Link ID	Digit (2440063100)
			LANES	Number of lanes	Digit (5)
			ROAD_NO	Road number	Digit (1)
Gyeonggi-do	951	2133	ROAD_NAME	Road name	Text (Youngdong)
			CONNECT	Link	Digit (000/101)
			MAX_SPD	Maximum speed limit	Digit (110 km/h)
			LENGTH	Link length	Digit (344.7723755 km)

4. Results and Discussion

4.1. Analysis Results

The DTG data sourced from freight vehicles along the Interstate 50 corridor facilitated the acquisition of comprehensive individual vehicle information traversing the expressway. Extracting data from peak traffic periods demonstrated noteworthy fluctuations in speed homogeneity across distinct timeframes. Notably, Tuesday registered the lengthiest average trip duration of 339.5 min, coupled with the shortest average journey distance of 165.1 km. Consequently, data from Tuesday at 9:30 a.m.—falling within the peak hour range of 8:00 a.m. to 10:00 a.m., a period characterized by heightened truck activity—was selected for further analysis. Raw data fields from truck DTGs underwent preprocessing, with the encrypted trip key and chassis number serving as criteria for bulk elimination of erroneous entries. Paramount data for this study encompassed vehicle speed, vehicle position, and GIS azimuth angle fields, with the 'TIME' field employed to facilitate time-based filtering.

The outcome of the 12-step grid partitioning of the Gyeonggi region along Expressway no. 50, as executed using the geohash algorithm with DTG data, is visually presented in Figure 7. If it is located above the standard deviation relationship curve between SMS and TMS, it is classified as SF, and if it is located below, it is classified as UF and the space is separated repeatedly.

Figure 8 delineates the resultant 12-step grid division for the eastbound Expressway no. 50 in Figure 8a, road line configurations and overlapping segmentation in Figure 8b, and the extraction of the unstable traffic component within the final 12 segments in Figure 8c. A parallel analysis conducted on Interstate 50's westbound trajectory is showcased in Figure 9. Table 5 comprehensively encapsulates the statistics derived from this analysis. Here, the part where the matched road alignment was cut off was excluded due to insufficient data for analysis.

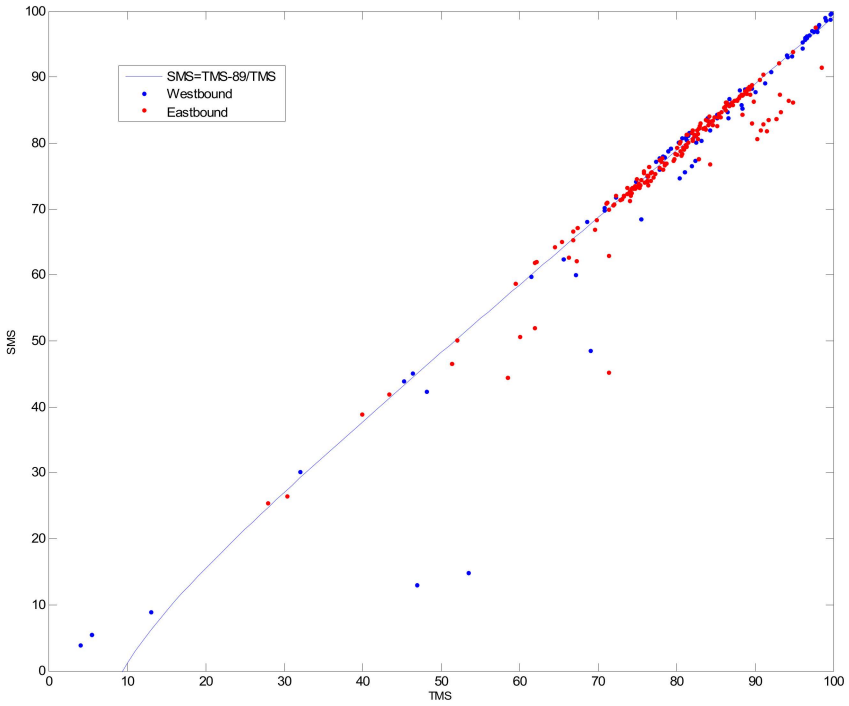
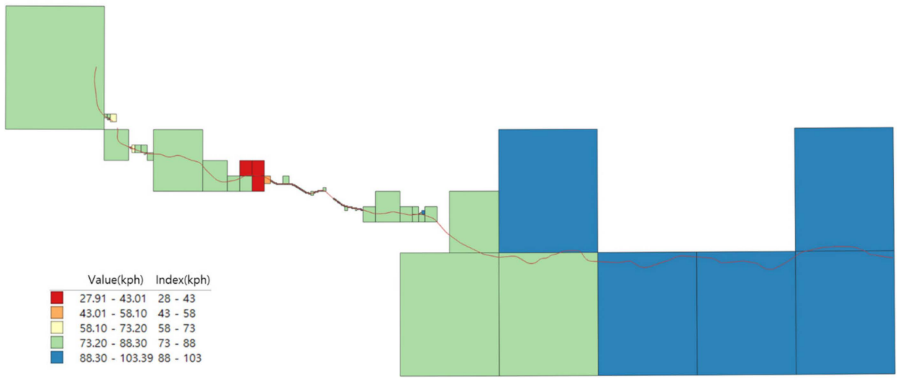
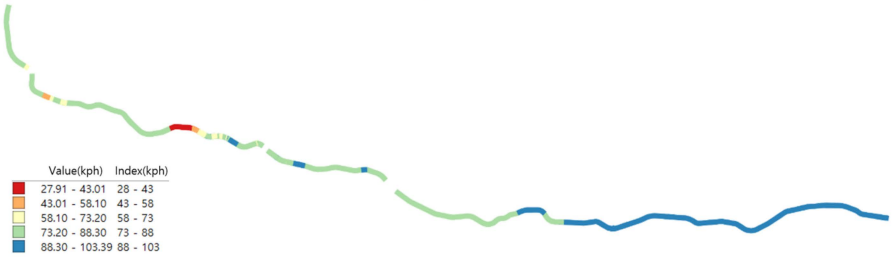


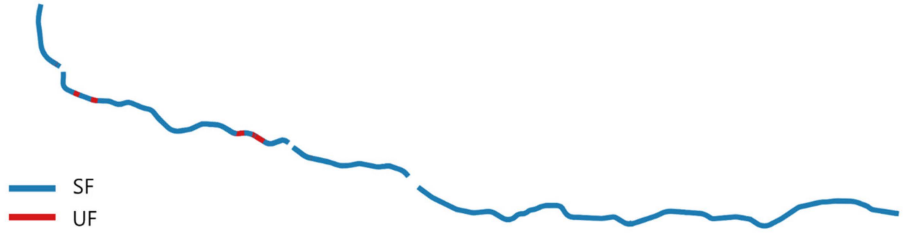
Figure 7. Comparison of TMS, SMS, and UF boundaries in the12-step grid split result.



(**a**) 12 step grid split result

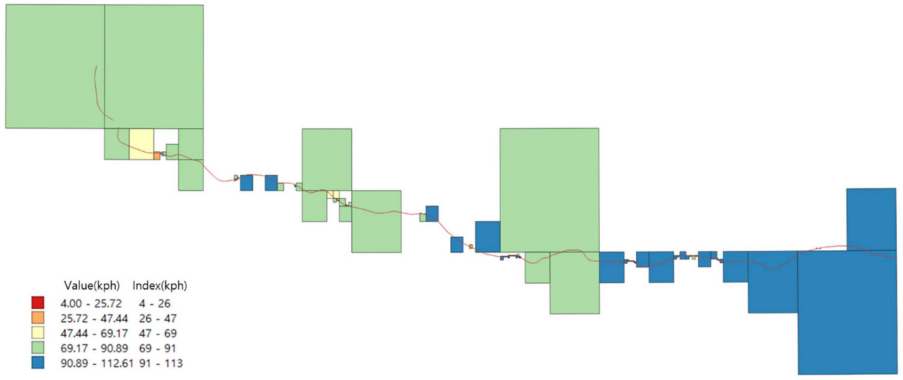


(**b**) Extracted overlaps of 12 step split grid and road

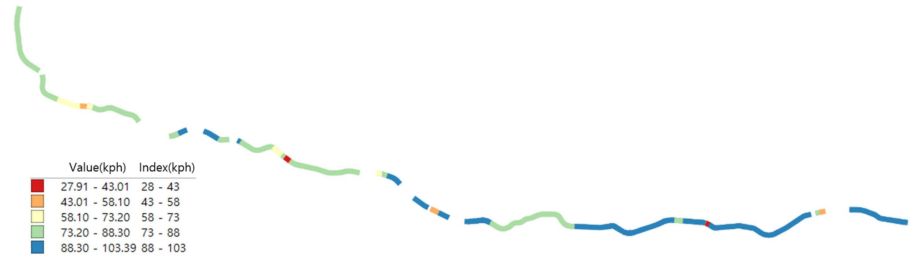


(**c**) UF segments extracted from the 12 step split grids

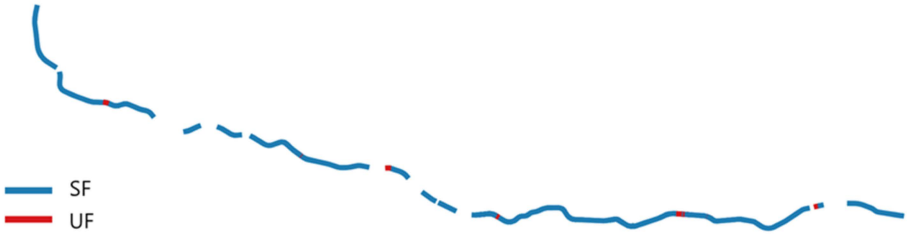
Figure 8. Eastbound UF results for Expressway No. 50.



(**a**) 12 step grid split result



(**b**) Extracted overlaps of 12 step split grid and road



(**c**) UF segments extracted from the 12 step split grids

Figure 9. Westbound UF results for Expressway No. 50.

Table 5. UF results for Expressway No. 50.

Grid Level	Eastbound				Westbound			
	ID Instance	Average TMS	Average SMS	Grid Count	ID Instance	Average TMS	Average SMS	Grid Count
5	20301	89.29	88.19	8	20301	85.65	84.92	4
6	212033	83.78	82.22	2	213033	89.54	88.67	5
7	2120231	80.44	78.78	3	2120333	88.88	87.83	10
8	21202303	62.46	60.65	8	21231011	93.92	93.44	8
9	212032032	74.85	73.77	9	212311101	80.10	79.12	9
10	2031022200	77.25	76.52	8	2031210231	83.56	82.66	11
11	20312013210	75.60	74.31	5	21230001013	89.85	89.31	13
12	203131231212	80.46	77.97	135	212310103130	88.70	86.19	63

The segmentation results encompassing 301 segments, stemming from the 12-fold division of 188 grids, identified outliers beyond the standard deviation threshold. Termed UF segments, these anomalies are featured within the westbound direction of Expressway no. 50, as depicted in Table 6. Within this context, 13 out of the total 63 segments emerged as UF segments, distinct in their characteristics.

Table 6. Comparison of UF in the 12-step split grids for Expressway no. 50.

Grid ID	TMS	SMS	Flow	Grid ID	TMS	SMS	Flow	Grid ID	TMS	SMS	Flow
212300011203	97.91	96.91	SF	212311110300	82.37	77.41	UF	212300000311	111.12	111.11	SF
212300011202	100.63	100.58	SF	212032031223	48.19	42.32	UF	212300000310	110.42	110.39	SF
212300011201	78.33	77.94	SF	203121023003	81.14	80.84	SF	213200013001	96.40	95.99	SF
212300011200	103.86	103.84	SF	203121023002	67.12	60.04	UF	213200013000	5.50	5.45	SF
212300001201	112.61	112.53	SF	203130321211	4.00	3.87	SF	212310102113	98.22	97.98	SF
212300001200	111.06	111.04	SF	203130321210	78.92	78.86	SF	212310102112	101.53	101.50	SF
212300010123	99.64	98.79	SF	212022113233	81.69	81.56	SF	212300001123	108.50	108.48	SF
212300010122	102.44	102.12	SF	212022113232	80.31	80.17	SF	212300001122	109.44	109.44	SF
212300010023	104.96	104.80	SF	212022103131	46.93	13.00	UF	212300000301	107.94	107.94	SF
212300010022	108.47	108.47	SF	212022103130	13.11	8.92	SF	212300000300	105.38	105.36	SF
212310103102	69.00	48.57	UF	213200011222	53.50	14.89	UF	212310103121	100.24	100.14	SF
212310103003	96.67	96.29	SF	212022113223	79.31	79.15	SF	212310103120	101.06	101.00	SF
212310103002	97.29	97.01	SF	212022113222	77.86	77.69	SF	212300001133	107.96	107.88	SF
212300011212	88.40	85.29	UF	212022113220	81.50	81.22	SF	212300001132	108.90	108.85	SF
212311100311	101.88	101.81	SF	213032233303	45.33	43.88	SF	212300001033	110.05	110.03	SF
212311100310	102.60	102.48	SF	213032233202	75.50	68.51	UF	212300001032	110.78	110.75	SF
212311110311	80.36	74.69	UF	212311111301	86.60	83.80	UF	212300010033	102.92	102.70	SF
212311110310	81.90	76.58	UF	212311111300	83.12	80.40	UF	212300010032	104.68	104.50	SF
212310103013	96.00	95.30	SF	212311111310	96.00	94.34	SF	212310103133	100.81	100.64	SF
212310103012	96.93	96.40	SF	212300010311	104.15	104.12	SF	212310103132	99.67	99.49	SF
212311110301	81.03	75.68	UF	212300010310	104.85	104.82	SF	212310103130	103.25	103.25	SF

4.2. Interpretation of Results

The velocity insights concerning Korean expressway segments are conventionally established through the concept of 'conzones,' wherein VDS facilitate the determination of instantaneous and average speeds within these delineated areas. A 'conzone' constitutes an expressway segment characterized by a consistent volume of passing vehicles, commonly associated with nodes such as interchanges, junctions, and toll gates. The particulars of this mechanism are exhibited in Table 7, which illustrates real-time data captured at 09:00 on 10 April 2018, encompassing various detectors and zones. VDS installations, spaced at irregular intervals spanning from 330 m to 2.96 km, engage in vehicle speed detection. Subsequently, the velocity data gleaned from two or more VDSs are amalgamated into an averaged conzone speed, imparting uniform interval-based information. Nevertheless, this approach falls short in accounting for instances such as accidents within a segment or the influence of subdivisions, thereby obscuring insights into UF occurrences.

Table 7. Speed information guidance criteria for westbound segment of Expressway No. 50.

Count	VDS_ID	Distance (km)	VDS Zone Length (m)	Conzone ID	Instantaneous Speed (km/h)
1	0500VDE00100	1.3	1240	0500CZE010	
2	0500VDE00200	2.3	1050	0500CZE010	
3	0500VDE00300	3.4	1300	0500CZE010	
4	0500VDE00400	4.9	850	0500CZE010	
5	0500VDE00500	6.05	2960	0500CZE020	
6	0500VDE00700	8.7	1340	0500CZE025	
7	0500VDE00800	9.9	740	0500CZE025	
8	0500VDE00900	11.3	2500	0500CZE030	66
9	0500VDE01000	12.6	660	0500CZE040	106
10	0500VDE01100	13.8	1300	0500CZE040	126
⋮	⋮	⋮	⋮	⋮	
95	0500VDE09400	109.9	1350	0500CZE200	89
96	0500VDE09500	111.2	1950	0500CZE200	
Minimum			330		49
Maximum			2960		126

The spatial domain under scrutiny encompasses the two-way traffic of Expressway no. 50 within Gyeonggi Province, characterized by a cumulative one-way expanse of 102.14 km. To derive segment speeds, the study employed DTG-based individual vehicle statistics. Segments featuring uniform speeds bereft of substantial inter-vehicle discrepancies underwent further partitioning into a geohash-derived grid, establishing a sense of uniformity, termed SF. In contrast, segments characterized by significant speed disparities among vehicles were systematically subdivided into smaller units designated as UF, until each grid reached a minimum size of 100 m. Notably, instances manifesting pronounced disparities in vehicle speeds, as illustrated in Table 8, were classified as UF segments.

Table 8. Average speed and traffic flow on westbound segment of Expressway No. 50.

Count	Length (m)	CHGRID	Average Speed (km/h)	Flow
1	599.52	212311111300	83.12	UF
2	599.52	212311111301	86.60	UF
3	599.60	21320000	107.62	SF
4	472.33	212301	85.19	SF
⋮	⋮	⋮	⋮	⋮
8	4900.36	212310103133	100.81	SF
9	468.00	212301	85.19	SF
10	4900.36	212310103130	103.25	SF
⋮	⋮	⋮	⋮	⋮
27	39.16	21231011	99.78	SF
⋮	⋮	⋮	⋮	⋮
320	599.52	2123111112	82.50	SF
321	599.52	21320000	107.62	SF
Minimum	39.16		4.00	
Maximum	5450.80		112.61	

With UF segments thus identified, the intricate road network map was subsequently scrutinized to pinpoint influential variables. Figure 10 elucidates the outcomes of this analytical endeavor; the red line in the picture is the UF segments analysis result. This line matches the center line of the road, and the shape of the road matches the background of the map, highlighting distinct scenarios:

(1) UF emerges as a segment characterized by a mixture of vehicles undergoing speed reduction before the expressway toll plaza, intermingled with those maintaining high velocities, yielding an average speed of 51 km per hour.
(2) The UF category is typified by an average speed range of 66–67 kph, wherein vehicles of varying velocities merge from the right.
(3) The first segment maintains an average speed range of 83–93 kph, juxtaposed with a second segment registering a speed of 60 kph. The former segment attains Level of Service A, yet it faces the influence of lane-changing vehicles decelerating into the right-hand turn lane. Concurrently, vehicles operating at standard speeds along the mainline further affect this segment. In contrast, the second segment reflects Level of Service C, contending with the deceleration of vehicles converging towards the toll plaza ahead, alongside a blend of both sluggish and rapid-moving vehicles.
(4) The final classification designates the speed range of 91–99 kph as UF. This range is subject to fluctuations attributable to frontal classifications, culminating in a mixture of slow and swift vehicles traversing the lanes.

Hence, UF characterizes short road segments spanning less than 100 m, wherein pronounced fluctuations in vehicle speeds ensue, potentially culminating in side or rear-end collisions. As exemplified in Figure 10, instance 1 highlights scenarios where segment driving speeds remain relatively low, yet substantial disparities emerge between front and rear vehicles due to hindrances preceding the lead vehicle or the downstream impact

of diminished speeds. Additionally, example 4 underscores that even within a relatively seamless segment marked by elevated overall speeds, noteworthy discrepancies in speed distribution between left and right lanes may arise. These differentials can be attributed to abrupt lane shifts, deceleration events related to entering and exiting vehicles, or the influence of junctions and classification zones. Such nuances elude comprehension through conventional average segment speed analyses, underscoring the criticality of delving deeper into speed variations to provide insights into hazardous segment attributes. This distinctive facet distinguishes the findings of this study from the existing body of research, spotlighting the paramount importance of driving safety considerations.

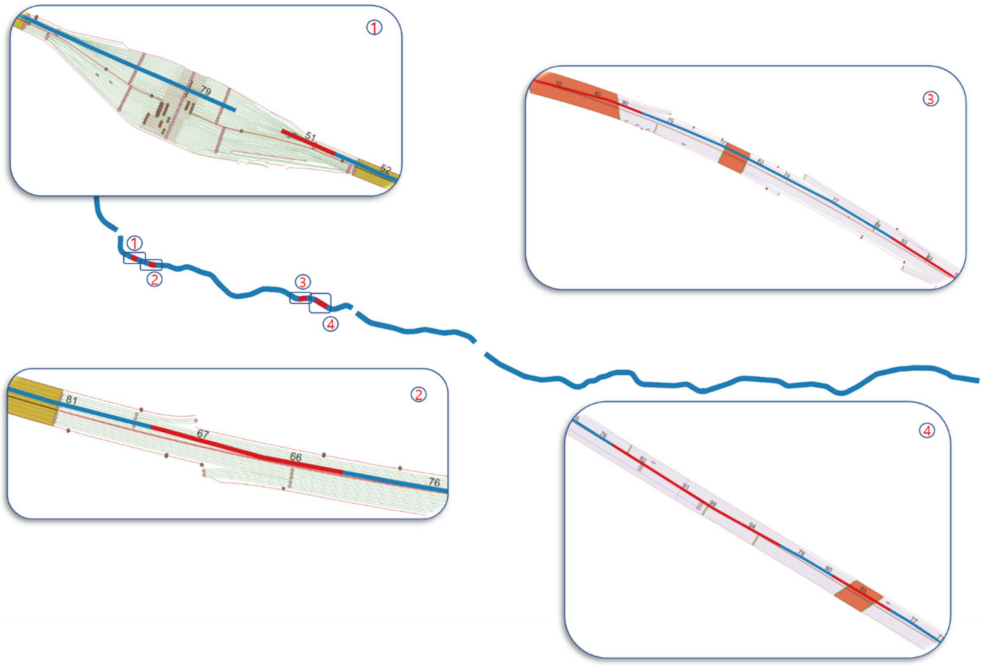


Figure 10. Analysis of the precise road network map of the UF segment.

5. Conclusions

This study introduced a methodology aimed at enhancing the precision of road information for fine-grained micro-segments through the utilization of individual vehicle big data and a criterion for categorizing UF segments characterized by speed variations across lanes. The presented approach involves establishing criteria for identifying segments with uneven speed distributions, leveraging inter-vehicle speed deviation based on the relationship between SMS and TMS, which can indicate the speed difference of individual vehicles within the same segment. Segment divisions were determined by computing standard deviation in speed discrepancies among vehicles through a geohash-coded approach. This approach, in turn, facilitated the identification and classification of UF segments exhibiting lane-specific speed differences, thereby providing detailed traffic information.

In this investigation, DTG data encompassing vehicle identifiers, coordinates, azimuth angles, and speed measurements from freight vehicles were harnessed within the Gyeonggi region of Expressway No. 50. As a result, a total of 301 segments were identified, including 178 eastbound and 123 westbound segments. UF segments corresponded to partitions falling beyond the reference standard deviation range. Compared with VDS and conzone speeds, the proposed method provided more precise and continuous speed information, surpassing those obtained from conventional link-based approaches.

A salient outcome of this study resides in its ability to pinpoint specific sections characterized by speed differentials and lane-specific velocity fluctuations. This demonstrated that UF segments can be segmented while under the effect of classification or merging, yielding insights into sudden decelerations or lane-specific slowdowns along expressways.

Such information holds a direct correlation with rear-end and side-impact incidents, positioning it as a valuable tool for furnishing drivers with proactive traffic accident mitigation guidance. Future research endeavors should contemplate the integration of additional data sources, including autonomous vehicles, as well as the exploration of lane-specific speed variation benchmarks based on precise road maps and speed deviation standards for each vehicle in SF according to specific parameters, including road type, lane configuration, vehicle category, and road alignment.

Due to its capacity to effectively deliver speed information with higher accuracy, the methodological framework introduced in this study for generating micro-segment speed information based on segmentation using individual vehicle big data can be used for real-time micro-segment speed information services. As a result, this approach holds the potential to improve the efficiency of the collection, analysis, segmentation, and dissemination of traffic information, leveraging driving data streams from autonomous, connected, and shared vehicles. Moreover, its applicability extends to supporting secure vehicle operation in dynamic driving environments encompassing both autonomous and non-autonomous vehicles.

Funding: This study was supported by a research grant from the Ministry of Land, Infrastructure, and Transport (20220293-001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Ahn, J.W.; Park, T.J.; Kweon, T.J.; Han, C.S. A path tracking control algorithm for autonomous vehicles. *J. Korean Soc. Precis. Eng.* **2000**, *17*, 121–128.
2. Maduro, C.; Batista, K.; Peixoto, P.; Batista, J. Estimating Vehicle Velocity Using Rectified Images. In Proceedings of the VISAPP 2008—International Conference on Computer Vision Theory and Applications, Madeira, Portugal, 22–25 January 2008; pp. 551–558.
3. Lee, S.K. The Development Direction of Highway C-ITS for Smart Autonomous Driving. *Transp. Technol. Policy* **2021**, *18*, 4–6.
4. MOLIT. *Korean Highway Capacity Manual*; MOLIT: Sejong City, Republic of Korea, 2013; pp. 78–86.
5. Lee, S.; Ahn, W.; Kang, H. A Study on Forecasting Traffic Congestion Using IMA (Integrated Moving Average) of Speed Sequence Array. *J. Civ. Environ. Eng. Res.* **2010**, *30*, 113–118.
6. Kim, Y.; Lee, S. Dynamic O-D Trip estimation Using Real-Time Traffic Data in congestion. *J. Korea Inst. Intell. Transp. Syst.* **2006**, *5*, 1–12.
7. Kim, S.; Ryu, J. Development of Impulse Propagation Model between Lanes through Temporal-Spatial Analysis. *J. Korean Soc. Transp.* **2011**, *29*, 123–137.
8. Do, C.W. *Transportation Engineering Principle*, 3rd ed.; Cheungmungak: Seoul, Republic of Korea, 2009; pp. 53–59.
9. Park, J.I. Road Policy Utilization Plan of Vehicle Driving Route Big Data. *Krihs Policy Brief* **2017**, *624*, 1–8.
10. Kim, D.; Jung, T.; Yi, K.S. Lane Map-based Vehicle Localization for Robust Lateral Control of an Automated Vehicle. *J. Inst. Control. Robot. Syst.* **2015**, *21*, 108–114. [CrossRef]
11. Lee, S.; Chang, H.; Kang, T. Analysis Method for Speeding Risk Exposure using Mobility Trajectory Big Data. *J. Soc. Disaster Inf.* **2021**, *17*, 655–666.
12. Lim, D.H.; Ko, E.J.; Seo, Y.H.; Kim, H.J. Spatiotemporal Traffic Density Estimation Based on Low Frequency ADAS Probe Data on Freeway. *Transp. Technol. Policy* **2020**, *19*, 208–221. [CrossRef]
13. Kwak, N.J.; Jeong, J.S. Trend of Image Recognition Technology for Autonomous Vehicles. *J. Korean Soc. Automot. Eng.* **2018**, *40*, 32–36.
14. Bernstein, D.; Kornhauser, A. *An Introduction to Map Matching for Personal Navigation Assistants*; Technical Report; TIDE Center: South River, NJ, USA, 1996.
15. White, C.E.; Bernstein, D.; Kornhauser, A.L. Some map matching algorithms for personal navigation assistants. *Transp. Res. Part C Emerg. Technol.* **2000**, *8*, 91–108. [CrossRef]
16. Han, E.; Kim, S.B.; Rho, J.H.; Yun, I.S. Comparison of the Methodologies for Calculating Expressway Space Mean Speed Using Vehicular Trajectory Information from a Radar Detector. *Korea Inst. Intell. Transp. Syst.* **2016**, *15*, 34–44. [CrossRef]
17. Ko, E.J.; Kim, S.H.; Kim, H.J. Microscopic Traffic Analysis of Freeway Based on Vehicle Trajectory Data Using Drone Images. *Transp. Technol. Policy* **2021**, *20*, 66–83. [CrossRef]

18. Liu, Y.; Yan, X.; Wang, Y.; Yang, Z.; Wu, J. Grid Mapping for Spatial Pattern Analysis of Recurrent Urban Traffic Congestion Based on Taxi GPS Sensing Data. *Sustainability* **2017**, *9*, 533. [CrossRef]
19. Wang, X.; Zhou, Q.; Quddus, M.; Fan, T. Speed, speed variation and crash relationships for urban arterials. *Accid. Anal. Prev.* **2018**, *113*, 236–243. [CrossRef] [PubMed]
20. Zhao, J.; Gao, Y.; Yang, Z.; Li, J.; Feng, Y.; Qin, Z.; Bai, Z. Truck traffic speed prediction under non-recurrent congestion: Based on optimized deep learning algorithms and GPS data. *IEEE Access* **2019**, *7*, 9116–9127. [CrossRef]
21. Ibarra-Espinosa, S.; Ynoue, R.; Giannotti, M.; Ropkins, K.; de Freitas, E.D. Generating Traffic Flow and Speed Regional Model Data Using Internet GPS Vehicle Records. *MethodsX* **2019**, *6*, 2065–2075. [CrossRef] [PubMed]
22. Drake, J.; Schofe, R.J.; May, A. A statistical analysis of speed-density hypotheses. *Highw. Res. Rec.* **1965**, *154*, 53–87.
23. Garber, N.J.; Hoel, L.A. *Traffic and Highway Engineering*, 3rd ed.; Cengage Learning: Boston, MA, USA, 2002; pp. 173–176.
24. Leong, L.V.; Azahar, A.M. Estimating Space-Mean Speed for Rural and Suburban Highways in Malaysia. In Proceedings of the 9th International Conference of Eastern Asia Society for Transportation Studies, Jeju, Republic of Korea, 20–23 June 2011; Volume 8, p. 295.
25. MOCT. National Spacial Data Infrastructure Portal. 2023. Available online: http://www.nsdi.go.kr/lxportal/?menuno=2679 (accessed on 10 July 2023).