A STUDY ON SEGMENTATION AND MAPPING METHODS ACCORDING TO THE CHARACTERISTICS OF DYNAMIC SOC INFORMATION

Kyu Soo Chong* and Junhee Youn
Research Fellow, Department of Future & Smart Construction Research, Korea Institute of Civil Engineering and Building Technology
283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do, 10223, Korea
Email: *ksc@kict.re.kr, younj@kict.re.kr

KEY WORDS: Dynamic SOC information, Geohash, DTG, Mean speed, Traffic density

ABSTRACT:

The diversification of urban mobility options, such as self-driving cars, personal mobility, and drones, and the sudden increase in logistics transportation have caused various safety hazards to burgeon in cities. Therefore, the need is also increasing for quick and accurate spatial recognition, along with information update-based dynamic object monitoring. Geospatial data, which are data from buildings, roads, and so on, do not change frequently, and the necessary time interval for updates in the case of changes is quite large. Hence, obtaining time- or date-specific information on changes in traffic, certain vehicles or pedestrians, changes in facilities, etc., is difficult. Currently, traffic is monitored through driver/pedestrian reports, CCTV, and probe vehicles, but owing to the limitations in update time interval and spatial recognition range, there are many cases in which severe accidents cannot be tended to within the golden hour. Therefore, dynamic mapping technology is being developed to consistently track and monitor dynamic SOC information in almost real time, including objects moving within cities and changing phenomena, through a fixed or dynamic platform. It is difficult to provide real-time information because the collected information must be matched with the street network. Furthermore, it is difficult to split the street network's composition of nodes and links into characteristic-specific areas based on the collected information. Therefore, in this regard, this study proposes a classification method for homogeneous speed spaces using a grid division method based on Geohash and various attributes to map the real-time big data generated using various sensors of the location, speed, and type of cars moving along urban roads. Digital tacho graphs(DTG) of trucks were used for the location-based information of cars, while the spatial range was Expressway 1, which is included in Seoul and Seongnam, South Korea. After binarizing and coding all coordinates, spatial division was conducted by setting the speed information obtained from each car's big data as a variable. To determine the level of spatial division, the difference between Time Mean Speed (TMS) and Space Mean Speed (SMS) was used, and thus, spaces with similar speed characteristics were identified. Divided spaces could be linearly mapped using high definition road map information and overlay analysis. This study is expected to enable the real-time mapping of urban data from various sensors, selection and automatic separation of similar areas based on certain characteristics of information, and provision of real-time customization services based on a urban's dynamic spatial information.

1. INTRODUCTION

Because urban mobility has become diversified (including autonomous driving cars, personal mobility, and drones), and the logistics movement has dramatically increased, increasingly more urban safety issues are arising. This requires object monitoring based on quick and precise recognition and update of dynamic spatial information. Static spatial information (such as buildings, roads, etc.) is quite unchanging, and the update cycle of change information prolongs. This makes it difficult to provide temporary information such as traffic changes, vehicle/pedestrian information, and facility changes. Currently, ground dynamic information is monitored through driver/pedestrian reports, CCTV(Closed-circuit Television), probe cars, etc. Nonetheless, owing to the limitations of the update cycle and the scope of spatial information recognition, the golden hour of an emergency is often missed. Addressing this limitation, a dynamic thematic map system is being developed to solve various social problems. This can be accomplished by detecting and tracking dynamic spatial information of live SOC(Social Overhead Capital). These include moving objects and changing phenomena in the city through the fixed or mobile platforms in quasi-real time. However, the produced information should be matched with the road network. This makes it difficult to provide the result in real time and to divide the road network of nodes and links into partitions by applying the information characteristics.

Various methods are used for map matching based on coordinates to analyze data. Considering map matching, several algorithms are used: point-to-point algorithm matching points-to-points (Bernstein, 1998), point-to-curve algorithms (Bernstein, 1998; White, 2000) connecting points to lines, and curve-to-curve algorithm (Bernstein, 1998; White, 2000) matching lines to lines. In addition, there are the Map Matching algorithm (Quddus, 2007) which is based on topological analysis and the Probabilistic algorithm (Quddus, 2007), a probabilistic map matching method. Nevertheless, these map-matching methods are inappropriate for location-based processing of large data. This is because of the error introduced by each algorithm and the need for spatial analysis between individual vehicle data and road alignments.

Spatial indexing techniques are a way to quickly index your location on a map using individual location coordinates. Traditional spatial index approaches are broadly categorized as uniform, non-disjoint, and disjoint decompositions, which correspond to grid-, R-Tree-, and QuadTree-based indexes, respectively (Singh, 2017). Moreover, Geohash, the latest method for location-based indexing, converts the coordinates of two values of global longitude and latitude into a single number. This is to speed up spatial information retrieval from spatial big data, using variants of geocoding methods—32-decimal (Niemeyer, 2000), 64-decimal, and Hilbert (Vukovic, 2016). However, these indexing approaches represent individual locations on a grid or another scale and have not been used for linear traffic analysis beyond regional traffic pattern analysis.

This study aims to present a method of grid division based on Geohash and spatial partitioning for homogeneous space. The study employs each property information to map big data, including the location, speed, and type of moving vehicles in the city extracted with various sensors in real time.

2. METHODOLOGY

2.1 Space Partitioning Methods

Although Geohash is a method of coding the entire world by dividing it into longitude/latitude, this study presents a method of spatial partitioning, coding, and mapping within the longitude/latitude range of Korea. The range of Korea's mainland is between 126° and 130° longitudes and 34° and 38° latitudes. Furthermore, spatial partitioning is performed by converting GPS coordinates of individual vehicles into transverse meractor (TM) coordinates and generating one code. Because mapping large data for a small area in a longitude-latitude range that fits Korea's territory may require excessive computation, it is necessary to apply peripheral geo-coordinates and reduce the scale of segmentation for spatial partitioning. Regarding each latitude and longitude, the corresponding spaces were listed by assigning a binary number of zero or one and changed to quaternary notation to generate one code. To minimize the difference in partitioning steps, we used the quaternary notation to apply a near-square spatial partitioning that fitted the shape of Korea and divided it into two equal spaces in the latitude-longitude direction. Figure 1 shows the process of quaternary spatial partitioning and code assigning.

(1) If a space is divided into four parts at a time, considering code 2 in <Fig. 1>, the latitude is '1' (down), the longitude is '0' (left), and the binary number '10' becomes the quaternary number '2'.

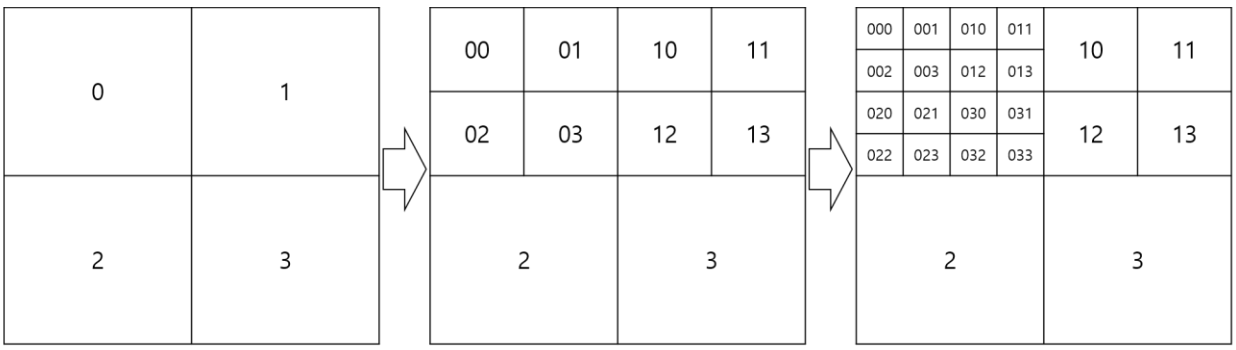


Figure 1. Method of space partition based on quaternary notation.

(2) Generate codes by splitting the given GNSS(Global Navigation Satellite System) coordinates and compare the codes per digit to use them.

(3) If the coordinates are coded 123123123123, the initial space is one. The code for the five-division space is 12312, and coordinates with the same code are counted as the same space.

Therefore, the space using the location information of individual vehicles is progressively divided from the divided space into four spaces. It can be calculated based on the coordinates of the four corners of each space.

Considering the above partitioning method, a spatial code with the same number of digits as the number of partitions is generated. When partitioned once, four spaces are generated (including CH0, CH1, CH2, and CH3) and when partitioned twice, 16 spaces are generated, including CH00, CH01, CH02, and CH03. When partitioned 12 times, 16,777,216 spaces are generated including CH000000000000, CH000000000001, CH000000000002, and CH000000000003.

2.2 Spatial Partitioning Method Based on Dynamic Information Characteristics

We observed dynamic SOC information and built spatial information to develop a dynamic thematic map system. This system can solve various social problems by detecting and tracking dynamic spatial information of life SOC. This includes moving objects and changing phenomena in the city through the fixed or mobile platforms in quasi-real time. The objects and sensors used are listed in Table 1.

Table 1 Dynamic SOC information characteristics

Category	Definition	Function
Life SOC	Living infrastructure such as transportation, administration, welfare, environment, etc.	Analysis
Dynamic Spatial Information	Changes in traffic conditions, pedestrian/vehicle movement and traffic density, changes in facilities, etc.	Analysis
Fixed Platforms	CCTV, detectors, communication APs, IoT sensors, etc.	Collection
Mobile Platforms	Unmanned Aerial Vehicles	Collection

The average speed of vehicles in a segment can be used by dividing the segment by vehicle speed. This does not reflect the difference in driving speed in the same segment owing to individual driver characteristics. Therefore, the temporal average speed and spatial average speed in the segment can be used to classify homogeneous segments and can divide the space. The condition for spatial partitioning to continue follows the Garber function: $\boldsymbol{u_s = 1.035u_t - 3.666}$:

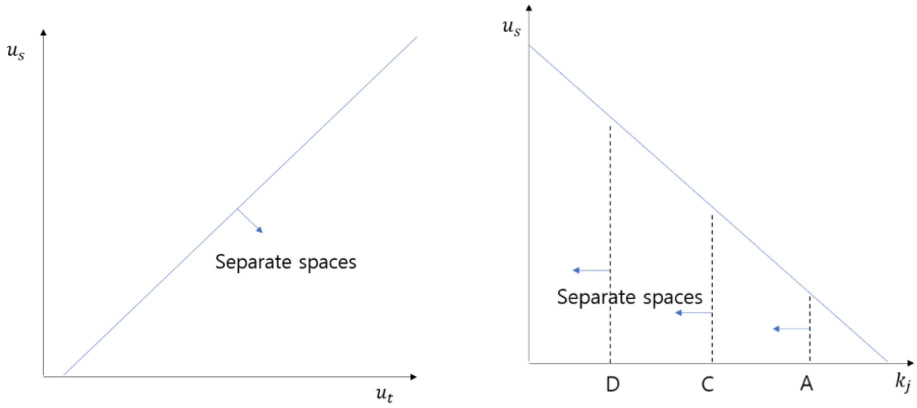
$$u_t = \frac{1}{n}\sum_{i=1}^{n}\frac{d}{t_i} \quad (1)$$

$$u_s = \frac{nd}{\sum_{i=1}^{n} t_i}. \quad (2)$$

The density of traffic flow is the number of vehicles within a unit length of a road or single lane at any given moment. Contrary to the criteria of traffic congestion or speed and traffic volume, density is difficult to measure on site and can be calculated by using aerial photographs or installing a vehicle detection system. In this study, the number and speed of vehicles in the same section using the location of individual vehicles generated by the mobile platform is employed to divide the same density section based on Greenshields (straight line model). The conditions for spatial partitioning to continue are the density values at service levels A, B, C, D, E, and F, considering the design speed of the road:

$$u_s = u_f\left(1 - \frac{k}{k_j}\right) \quad (3)$$

where u_t = time mean speed (km/h)
u_s = space mean speed (km/h)
t_i = Passage time of car i (hour)
d = segment distance (km)
u_f = free flow speed (km/h)
k_j = maximum density (v/km)



(a) (b)

Figure 2. (a) Equal-rate spatial partitioning (b) Equal density spatial partitioning.

3. RESULTS AND ANALYSIS

The locational information of vehicles was obtained from the Digital Tacho Graph of freight vehicles, and the spatial scope was Expressway 1. This is included in Seoul, Korea, and Seongnam (Gyeonggi-do). The coordinates were binarized and coded, based on which spatial partitioning was performed with speed information as a variable among the big data of each vehicle. To determine the level of spatial partitioning, we used the difference between temporal and spatial average speed and determined spaces with similar speed characteristics. The partitioned space was linearly mapped using a precise road map of the city and overlay analysis.

3.1 Data Preprocessing

The target of this study is Expressway 1, which is included in the city of Seoul. Regarding this study, data such as location and speed information of each vehicle should be collected. Moreover, data collection on the highway should be available, and the impact of driving lanes such as designated lanes should be small. We used freight vehicle Digital Tacho Graph (DGT) data because the taxi DTG data is mainly distributed in urban areas. Random stops may occur owing to the influence of passengers getting on, off, and waiting. The bus DTG is also mainly distributed in urban areas, and considering intercity buses, it may appear different from the surrounding traffic flow owing to the influence of bus-only lanes. Although freight vehicles have relatively low driving speeds and they often drive on the right side of the road, they are evenly distributed on roads across the country. This includes highways; therefore, relatively accurate data can be collected. In addition, it is obligatory for freight vehicles to install DTG and publish the information, making it possible to collect the data in real time. Nonetheless, there are limitations because it cannot represent general vehicles, considering real-time collection, driving speed, and vehicle characteristics.

Because there is no driving route information in DTG data, we had to extract data by highway route. To achieve this, a standard node link of the Intelligent Transportation System Management was used. The extracted data was buffered by 3.5 m per highway lane. We excluded cases where the speed data was 0 kph, considering stops such as sleeping shelters, rest areas, and cases where the speed limit exceeded 20%. The location information in the DTG data is GPS longitude and latitude coordinates, which require conversion to plane orthogonal coordinates. The longitude and latitude coordinate system is WGS84, which is converted to the central origin (N38, E127) and TM coordinate system. The error of the actual GPS can exceed the lane standard. This limits the data aggregation; nevertheless, the possible GPS error is not considered because we used the average value of large data. Moreover, the up and down direction of the vehicle was separated. When the azimuth was zero degrees from the true north, between 90 and 270 degrees was considered as down and the opposite was considered as up. Only the down data are applied in this study.

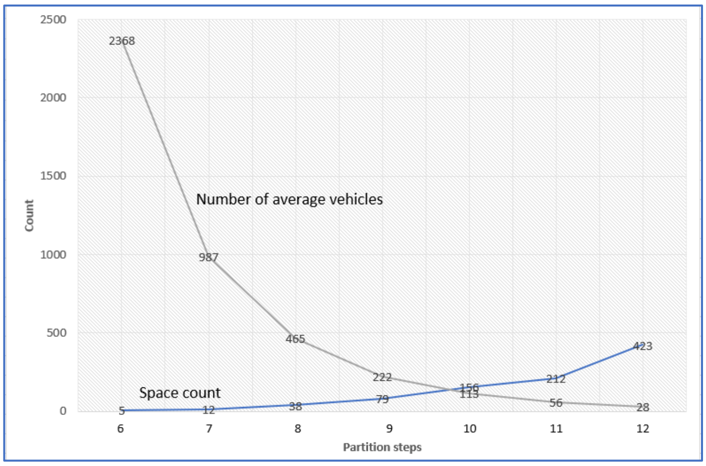
3.2 Real-time Big Data Mapping

This study aims to map large amounts of individual vehicle information and use it to analyze traffic information including speed change zones, vehicle driving patterns, and individual driving characteristics. The precision of spatial separation can be differentiated for different purposes. We applied spatial partitioning to the speed information of individual vehicles to reduce the speed deviation between vehicles and increase the efficiency of geocoding. The DTG data in the Gyeonggi-do region of Expressway 1 was geocoded using GPS coordinates. We included all the DTG data attributes of individual vehicles and applied cumulative data for 5 min from 9:30 a.m. at one-second intervals for each vehicle. Table 2 shows the number of spaces and individual vehicles geocoded to the same space. It also demonstrates the sum of the average speeds within each space divided by the total number of spaces. The sum of the standard deviations within each space divided by the total number of spaces by spatial partitioning steps has also been presented.

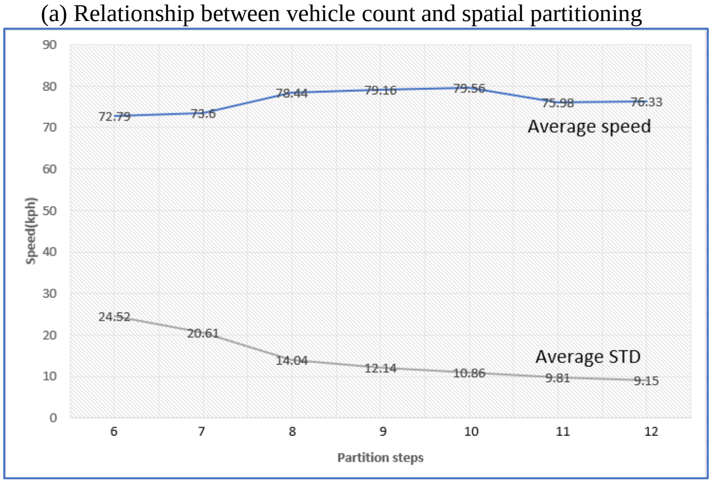
Table 2. Real-time mapping results of vehicular big data

Partition steps	Number of spaces	Number of average vehicles	Average speed (km/h)	Average standard deviation (km/h)
6	5	2368	72.79	24.52
7	12	987	73.60	20.61
8	38	465	78.44	14.04
9	79	222	79.16	12.14
10	156	113	79.56	10.86
11	212	56	75.98	9.81
12	423	28	76.33	9.15

Considering Figure 3(a), the number of spaces increases with the number of spatial partitioning, whereas the number of vehicles geocoded in the same space decreases. To achieve efficient geocoding, it is necessary to consider the appropriate number of vehicles and spaces. If there is low traffic volume (i.e., a smooth traffic flow), the number of spaces can be reduced. If there is a high traffic volume, the number of spaces can be determined based on whether there is an unstable or stable traffic flow. Figure 3(a) shows the average number of vehicles per space and the total number of spaces based on the spatial partitioning steps. Figure 3(b) shows the average values of the average speed and standard deviation per space based on the spatial partitioning steps. The average speed was constant regardless of the number of spatial partitioning. Moreover, the standard deviation of the speed in the same space declined with the number of spatial partitioning, and the difference in standard deviation was insignificant after nine partitions.



(a) Relationship between vehicle count and spatial partitioning



(b) Relationship between vehicle speed and spatial partitioning.

Figure 3. Relationship between partition steps, count and speed

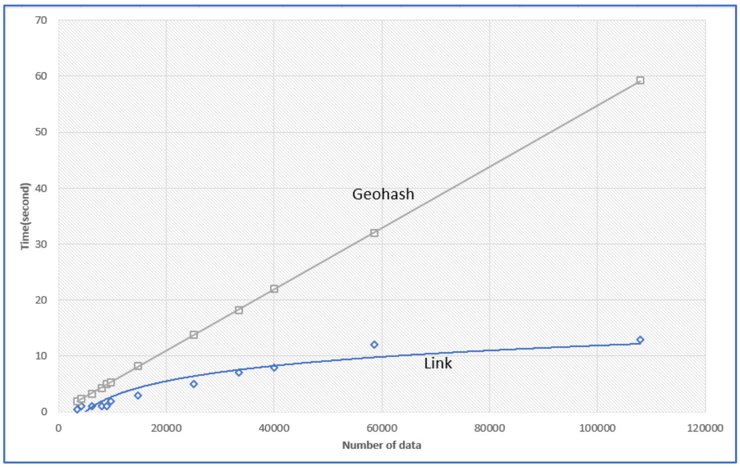


Figure 4. Comparison of the mapping speed: Geohash-based spatial partitioning vs. link matching-based mapping.

We develop a spatial-based mapping of vehicle information that is linearly based on roads, enabling rapid processing of large data. To verify the speed of data processing, we compare it to the method that uses the location data of individual vehicles to map to the nearest link, a point-to-curve method. We applied different numbers of DTG data and mapping routes for comparison. Our method can process approximately 4.7 times faster for 100,000 data. Figure 4 compares the processing time of the proposed method to the matching methods using the data processed. Although the link-based spatial processing method increases linearly, the spatial processing method increases processing speed less than the increase in data volume. This proves that it is possible to efficiently map large amounts of individual vehicular information. This is mainly because the link-based processing method requires spatial analysis of links for each individual vehicle location. However, this study's method computes the locations of individual vehicles as spatial codes and maps them without spatial analysis. This is an advantageous feature of this study, and it is an optimal spatial partitioning for the analysis of attribute criteria, such as vehicle speed. Despite these advantages, this method has some limitations. Owing to the nature of spatial separation, if it is close to a separating boundary, it may be separated into different spaces even if it has similar characteristics. If it is on a boundary line, it is difficult to map the space.

3.3 Adjusting Spatial Partitioning Steps by Data Characteristics

We conducted spatial partitioning considering the location-based speed and density data of vehicles. The DTG data of Expressway 1 sections in Seoul and Seongnam (Gyeonggi-do) were extracted based on road links. The spatial and temporal average speed of individual vehicles were extracted, and the partitioning was repeated based on Equation (1), (2), (3) and Garber function. The partitioning results in sections with the same characteristics as shown in Figure 5.

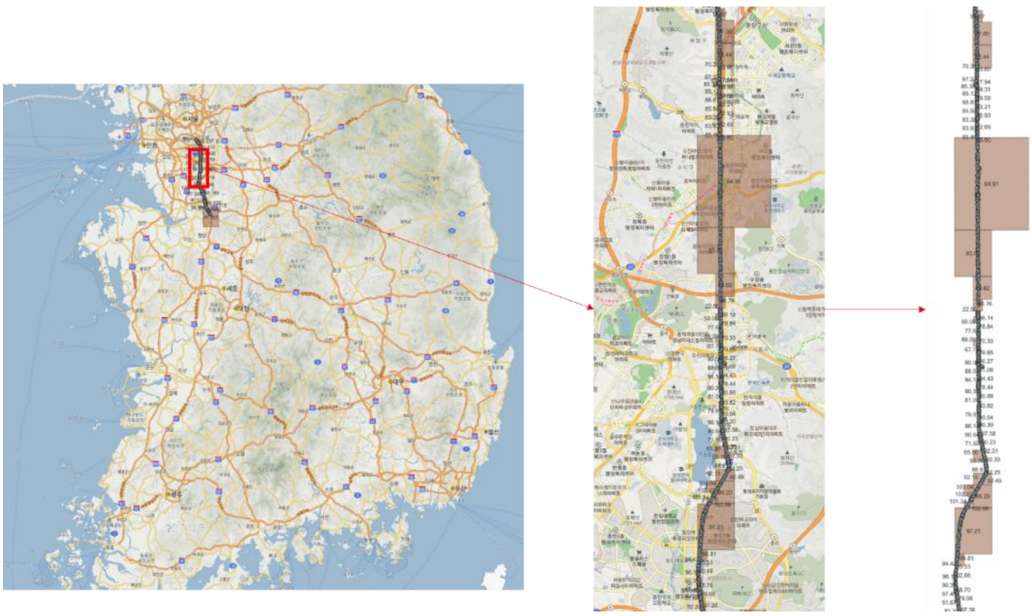


Figure 5. Spatial partitioning based on the speed data.

Considering Table 3, the total 63.46 km route is divided into 323 grids. It has a minimum and maximum average speeds of 12 kph and 107 kph per grid, respectively, with a population standard deviation of 11.52 kph.

Table 3. Spatial partitioning results by the data characteristics

Step	No. of grids	DTG count	Grid ID (exam.)
5	1	3850	23001
7	3	6273	2120030
8	2	2901	21202310
9	12	5435	212232231
10	14	3103	2120033212
11	16	988	21200330320
12	291	19733	212203210133

4. CONCLUSION

This study presents a method of grid division based on Geohash and spatial partitioning for a homogeneous space. It uses each property information to map big data, including the location, speed, and type of moving vehicles in the city extracted with various sensors in real time. The locational information of vehicles was obtained from the Digital Tacho Graph of freight vehicles. The spatial scope was Expressway 1, located in Seoul, Korea, and Seongnam (Gyeonggi-do). The coordinates were binarized and coded based on which spatial partitioning was performed using speed information as a variable among the big data of each vehicle. To determine the level of spatial partitioning, we used the difference between temporal and spatial average speed and determined spaces with similar speed characteristics. The partitioned space was linearly mapped using a precise road map of the city and overlay analysis.

We observe that our method processes 100,000 pieces of data approximately 4.7 times faster than the point-to-curve method. It uses individual vehicle location data to map to the nearest link. Furthermore, we have been able to spatially partition areas with similar driving speeds and densities using real-time location-based vehicle data. We have also mapped them on a road-by-road and lane-by-lane linear basis using a precise network map. This study lays the foundation for providing customized services of dynamic real time spatial information in cities. This can be done by enabling real-time mapping of big data, including the location, speed, and vehicle type of vehicles traveling on city roads extracted from various sensors. It can also be achieved by determining and automatically partitioning the same space based on the characteristics of the information.

ACKNOWLEDGEMENTS

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143782).

REFERENCES

Bernstein, D., and Kornhauser, A. 1998. An Introduction to Map Matching for Personal Navigation Assistants. Technical report, New Jersey TIDE Center.

White, C.E., Bernstein, D., and Kornhauser, AL. 2000. Some Map Matching Algorithms for Personal Navigation Assistants. Transportation Research Part C: 8, pp. 91-108.

Quddus, M.A., Ochieng, W.Y., and Noland, R.B. 2007. Integrity of Map-matching Algorithms. Transportation Research Part C: Emerging Technologies, Vol.14, pp. 283-302.

Singh, H., and Bawa, S. 2017. A Survey of Traditional and MapReduce Based Spatial Query Processing Approaches. ACM SIGMOD Record, Vol.46, Issue 2, pp. 18-29.

Chen, S., Ooi, B.C., Tan, K.L., and Nascimento, M. 2008. ST2B-tree: a Self-tunable Spatio-temporal b+-tree Index for Moving Objects. ACM SIGMOD international conference on Management of data, pp. 29-42.

Niemeyer, G. (2020.12.10.), Wikipedia, https://en.wikipedia.org/-wiki/Geohash

Vukovic, T. 2016. Hilbert-Geohash-Hashing Geographical Point Data Using the Hilbert Space-Filling Curve. M.S. thesis, NTNU, Trondheim, Norway.