

# Modeling Traffic Congestion with Spatiotemporal Big Data for An Intelligent Freeway Monitoring System

Karisma Trinanda Putra<sup>1,2\*</sup>, Jing-Doo Wang<sup>1</sup>, Eko Prasetyo<sup>1,2</sup>, Prayitno<sup>1,3</sup>

<sup>1</sup>College of Information and Electrical Engineering, Asia University, Taichung, Taiwan

<sup>2</sup>Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia

<sup>3</sup>Faculty of Electrical Engineering, Politeknik Negeri Semarang, Semarang, Indonesia

\*Corresponding Author: karisma@ft.ums.ac.id

## Article Info

Volume 83

Page Number: 6422 - 6429

Publication Issue:

March - April 2020

## Abstract

Traffic congestion is a complex, nonlinear spatiotemporal modeling problem. By collecting and analyzing a vast quantity and different categories of information, traffic flow, and road congestion can be predicted and controlled on an intelligent transportation system. This report provides an analysis of traveling time across Taiwan from North to South, vice versa. We analyze traffic in a national freeway between Tainan and Kaohsiung section, which represents the common trip of the population in Southern Taiwan. The data is recorded using the Electronic Toll Collection System (ETC) provided by Ministry of Transportation in Taiwan. We use MapReduce framework to process data into a smaller task which can be distributed on several computer clusters to speed up the process. The results show that the spatiotemporal model of traffic flow is strongly influenced by direction, working hour, and holidays with a recurring pattern for each week. The distinctive pattern inside the spatiotemporal dataset can be used on an AI-powered decision-making system for future development.

**Keywords:** Spatiotemporal, ETC, MapReduce.

## Article History

Article Received: 24 July 2019

Revised: 12 September 2019

Accepted: 15 February 2020

Publication: 04 April 2020

## I. INTRODUCTION

Transportation is one of the supporting factors for developing countries to support the growth of the national economy. An indicator of the success of a country in developing their transportation sector is how they can ensure a smooth movement of residents, goods, and services to reach all regions. In this globalization era, national roads have not been able to meet the needs of good transportation facilities due to several obstacles, some of which are a large number of obstacles and intersections within them. There are several approaches that are built to support citizen mobilization in a country. One of them is by increasing the development of infrastructure, with the use of freeway. There are many improvements regarding to freeway

technology i.e. are the implementation of a smart system that can provide accurate traffic information, lane density forecasting, and a contactless payment system without using manual payment gates. By learning from several developed countries, this system has been implemented and gained success in its implementation over the past decade, especially in Taiwan. A freeway is a road with many lanes, two directions with separators and without intersections. A freeway network requires a large investment at the outset. However, the long-term functions obtained will be proportional to the initial investment costs. An ideal freeway is a separate two-way road, without intersections, short travel times, payment systems integrated with user accounts, and has an ability to smartly guide users to shorten mileage through the fastest estimated

paths [1,2].

Traffic congestion is a complex, nonlinear spatiotemporal modeling problem that has a different character in each region, type of road and time [3]. Each region has a different population density. Population density affects the number of vehicles and the level of mobility of residents in the region [4]. Areas with high density without the support of good road infrastructure will increase the risk of traffic congestion. The width and type of road also affect the smooth running of the vehicle [5]. Another factor that affects traffic density is time. At certain times such as when leaving and returning to work will increase the frequency of residents' trips [6]. Differences in days such as workdays and holidays also have a significant impact on the pattern of population mobility. These patterns can be used as suggestions for formulating public policy so that it will certainly reduce the frequency of congestion. By collecting and analyzing spatiotemporal datasets of traffic, vehicle flow and road congestion can be predicted and avoided in an intelligent transportation system [7, 8]. The transportation system that is able to provide a rapid reporting response will produce effective travel patterns that are beneficial to the population so that the population can actively participate in suppressing the rate of congestion by utilizing public transportation or by utilizing the time outside the peak hours.

Several studies were conducted to see the advantages of using spatiotemporal data to improve the accuracy of prediction systems. Compared to the use of raw data, the use of spatiotemporal data has an advantage in terms of data capacity efficiency [9]. This type of dataset is also proven to be resistant to noise and data loss [10, 11] when it is being processed further. Spatiotemporal data are shown to have stronger correlations between variables than 2-dimensional linear datasets. The spatiotemporal dataset combines the elements of time and location with changes in the measured variable. Further use has resulted in a more advanced prediction system e.g. by using fuzzy technology [12, 13], LSTM [14, 15], even Deep Learning [16, 17, 18]. The use of spatiotemporal data is suitable for use on ETC systems that produce tens of gigabytes of data in one day so that it can improve storage efficiency and produce stronger variable correlation especially for

prediction systems that are highly dependent on changes in time and location.

The case study of this research is on a toll-based transportation system that runs along the west coast of Taiwan. A freeway connects Taipei as a capital in the North and Kaohsiung, the second largest city in Taiwan in the South. The freeway utilize E-Gate technology that is integrated with Freeway Gantries and every vehicle in Taiwan. Information such as vehicle type, travel time, gantry number, and timestamp are recorded automatically when the vehicle passes through the gantries around the freeway. As a sample, we analyze traffic in a freeway section, which is located between Tainan-Kaohsiung, represents the common trip of the population from north to south and vice versa. This section is also among the two largest international airports in Taiwan located in Taoyuan and Kaohsiung. Kaohsiung City also has an international port which certainly impacts the frequency of population travel as well as goods from this country and East Asian countries. We also provide analysis of intersections connected by highways located in the middle of Kaohsiung City.

The remainder of this paper is organized as follows. The method is described in Section 2. Experiment results presented in Section 3. Section 4 provides a discussion. Section 5 concludes the paper.

## II. MATERIALS AND METHODS

### 2.1. Tools

Tools which involving hardware and software, are described in the following.

#### 2.1.1. Software

Programming tools help processing raw data into easy-to-describe data, have characteristics that are meaningful and generate comprehensive analysis. Hadoop is a Java-based open source framework or a programming tool under the Apache license to support applications related to Big Data processing. Hadoop uses Google MapReduce technology and the Google File System (GFS) as its foundation. There are 3 main challenges that make Hadoop as a popular framework for dealing with problems related to Big Data.

- **Volume**, the need to store and manage large

amounts of data, and the data is always updated every time.

- **Velocity**, data changing is so fast and the need to be able to access large data quickly
- **Variety**, the more varied the data so that with technology such as relational databases (RDBMS) currently can no longer be handled.

Hadoop is optimally used to handle large amounts of data in both structured, semi-structured and unstructured data. Hadoop replicates data on several computers (clustering), so that if one computer dies / problems then the data can be processed from one of the other computers that are still alive. The main backbone of Hadoop is MapReduce. MapReduce is a programming model for large-scale data management so that users only need to interact with this model regardless of the problems associated with a computer cluster consisting of many machines with different load balancing, shown in Figure 1.

An additional tool, i.e. Eclipse also be used as an IDE for Hadoop. To use MapReduce, a programmer simply makes two programs, i.e. a program that contains the task planning and distribute the smaller task into the clusters and a program that combines the result from each cluster. The program is written in Java programming language. Then, we upload the program to the cluster and run it on our distributed computing machine. The result is recapped into a text file that contains information such as the number of the vehicle that passes through the gate, travel time, type, and the spatiotemporal traffic pattern. Meanwhile, to help visualize the results, we use Delphi 7 with the TeeChart plugin to graph the spatiotemporal data. Delphi uses Pascal as a programming language. The program that has been made translate results from \*.txt files into the graph as shown in Figure 3. Data is presented in a 3-dimensional curve that provides information on the number of vehicles, travel time distribution and 24-hour time index in a day.

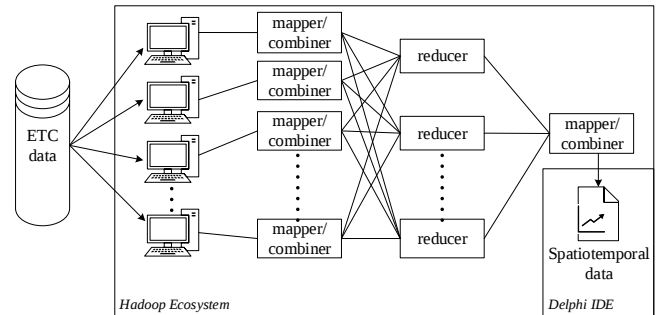


Figure 1. System architecture

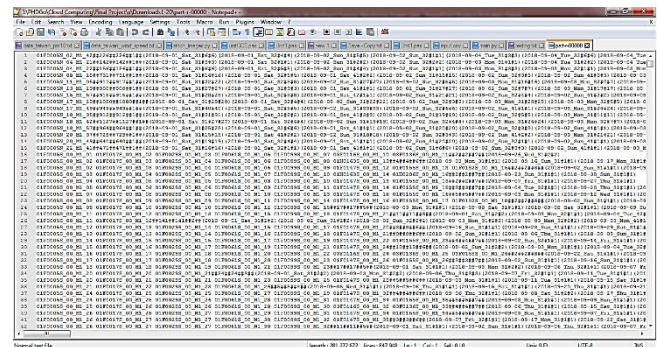


Figure 2. Raw data with \*.csv file format extracted from online database

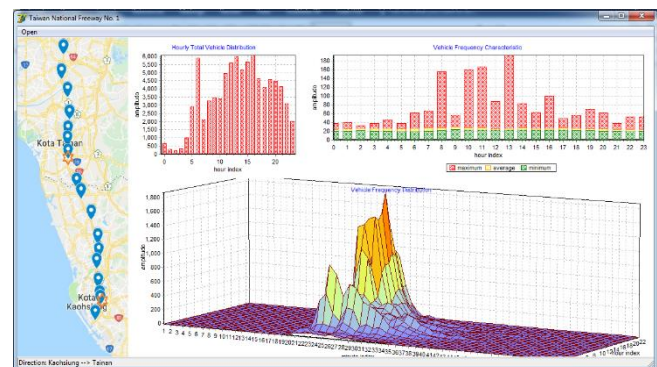


Figure 3. The used of Pascal-based software for spatiotemporal analysis

## 2.1.2. Hardware

For experiment, we use 8 computer clusters with Intel Core i7 7700 and 8GB of RAM. Meanwhile, on the software side, we use Windows 7 OS with the Hadoop programming environment. Hadoop supports MapReduce for the purpose of distributing Big Data processing load. MapReduce is a programming model that is related to processing large data sets by using distributed algorithms in parallel. Each computer is connected to each other in a network with other computers to form a cluster of computers. The cluster can share resources and data so that it is possible to carry out the process of completing tasks together at one time. The cluster

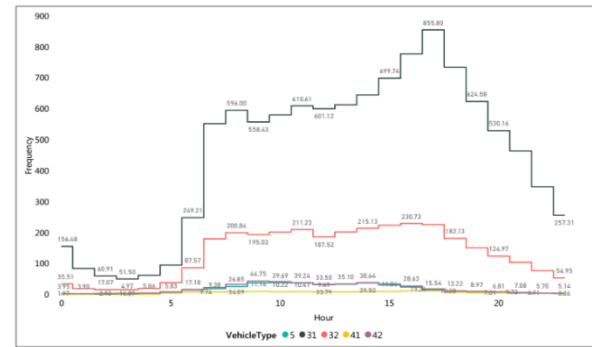
looks as the larger system that combine all resources of each node, with 64 cores processor with 64 GB of RAM in total. Then, we run the process that involves the MapReduce algorithm where large tasks will be mapped into several simpler tasks. Furthermore, these simple tasks are divided into machines available in a cluster. Each machine will work to finish the small task as fast as possible. Then, the coordinator merges every small task so that, the big, complex, and massive task can be completed efficiently.

## 2.2. Materials

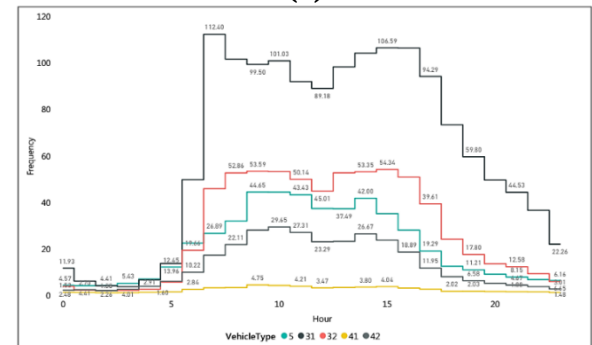
For the experiment, we use open dataset i.e. extracted from Traffic Data Collection System (TDCS) provided by Ministry of Transportation Taiwan. The data is collected for 3 months between June to August 2018. The volume of raw data is about 60 GB in \*.csv format as seen in Figure 2. The data consist of several columns i.e. describes vehicle ID, timestamp, gantry on, gantry off, and history of vehicle travel when passing through each gate. Then, the data is processed using significant time travel algorithm to parsing time sequence for each vehicle. This intersection is flanked by 4 gantries registered in the Taiwan government manifest. The Gantries include 01F3676N and 01F3696N for the Northbound direction, while 01F3676S and 01F3686S for the Southbound direction. We monitor the frequency of vehicles entering the highway or going out to the intersection as indicated by the blue pointer in Figure 3.

## III. EXPERIMENTS AND RESULTS

Several tests are conducted using Hadoop Framework. First, a segment of freeway is chosen which represent the traffic that came out and entered into the center of Kaohsiung City, considering that the city was located at the far end of the highway system in Taiwan. The pattern observed was carried out for 24 hours so that the changing of the traffic can be observed from time to time. Next, we observe the traffic in one week. Meanwhile, the latest experiments were carried out using longer sections between the cities of Tainan and Kaohsiung for both directions. The extracted data is the density of hourly vehicles and the speed of vehicles in taking this segment.

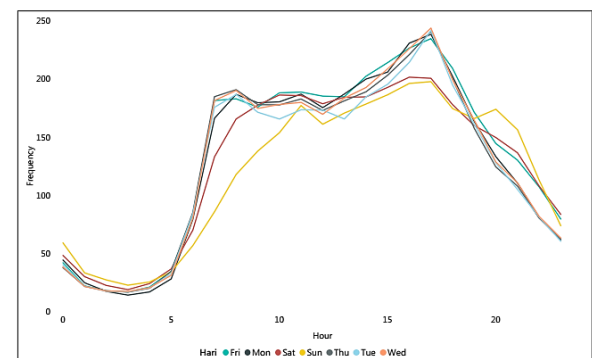


(a)

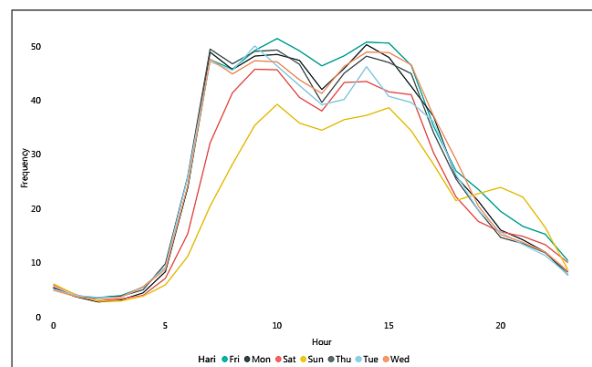


(b)

**Figure 4.** Frequency distribution of vehicle type from (a) suburb to Kaohsiung downtown (b) Kaohsiung downtown to suburb



(a)



(b)

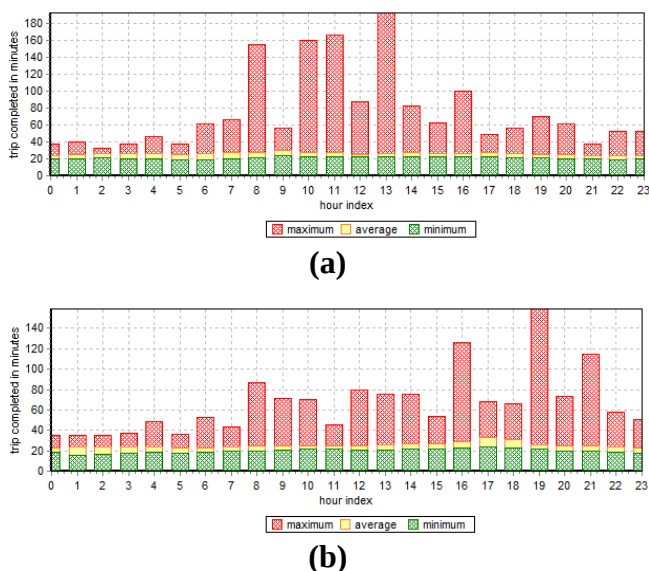
**Figure 5.** Frequency distribution of vehicle during a week from (a) suburb to Kaohsiung downtown (b) Kaohsiung downtown to suburb

In the first experiment, we observe the frequency distribution of vehicles coming out of the freeway to the intersection at the center of Kaohsiung City. Data is presented based on vehicle variations where there are 5 types of vehicles, consisting of (5) tanker, (31) small vehicle, (32) light truck, (41) bus and (42) big truck. The graph in Figure 4 (a) present average vehicle that passes every hour for 3 months during data collection. The most vehicles that use freeway are small vehicles, almost reaching 900 vehicles in peak hours. The peak hour is at 17.00. This could be due to a large number of workers returning home at these hours. The next most vehicles frequency are light trucks, in the range of one-third of the number of small vehicles. Small trucks are commonly found in the area around the airport and port for the process of shipping goods. The 3 other type of vehicles i.e. tanker, bus, and big truck do not use this intersection significantly. Vehicle-rushing hours are between 7:00 and 23:00. Outside these hours, the frequency of vehicles passing the intersection is very low. This could be due to the large number of commuters i.e. workers and students who stay out of town or suburbs of Kaohsiung. At 23:00, this intersection is still quite busy servicing vehicles, probably because this road connects the second busiest airport in Taiwan and also Kaohsiung Harbor.

obtained by summing the vehicles coming out of gantry from the South and the North direction. Similar to the previous data, the highest user is small vehicles reaching the range of 100 vehicles at rush hour between 7.00 and 18.00. This is different from the number of vehicles coming out of the freeway. In the previous graph, the maximum frequency can reach 800 vehicles, meanwhile, the incoming vehicles are only around 100 per hour. The other four types of vehicles are not too significant using the freeway. Moreover, the frequency of buses, is very rarely found through this intersection to the freeway. Vehicles that enter through the freeway are quite rare because the airport and port distances are easily accessible from the middle of the city. Meanwhile, the vehicle going out from freeway to this intersection is very large due to the flow of commuter from the airport or harbor, and also the flow of vehicles from cities in northern Kaohsiung. The peak hours of the vehicle that uses the freeway from this intersection also shift as a result, which ends at 18.00. Above this hour, the frequency of vehicles decreases below 50 vehicles per hour.

In the second experiment, we want to see the difference in freeway usage between weekends and weekdays (Figure 5). The pattern that arises is the average frequency of freeway usage at downtown Kaohsiung. Sunday has the lowest frequency of vehicles that use this road, followed by Saturday. Sunday traffic has a slow rise time during the morning and also slowly decreasing time while at night. Meanwhile, weekday is more crowded than that of in the weekend. The number of vehicles coming out of the freeway is more than vehicles entering the freeway. The peak hours for vehicles coming out of the freeway occur at 17.00 - 18.00. Meanwhile, vehicles entering the freeway have peak hours that are spread evenly between 7:00 and 18:00. The peak hour for the traffic that exit to Kaohsiung downtown is more evenly distributed than that of entering the freeway to the North. The frequency of traffic that exit the freeway is 5 times higher than that of enter the freeway.

In the last experiment, we want to see the maximum, the average, and the minimum travel time for a segment which is located between Kaohsiung and Tainan City (Figure 6). The distance of this segment is about 49.4 km. The higher the amplitude (in minutes), the lower the speed of



**Figure 6.** Average travel time for (a) Tainan (b) Kaohsiung direction

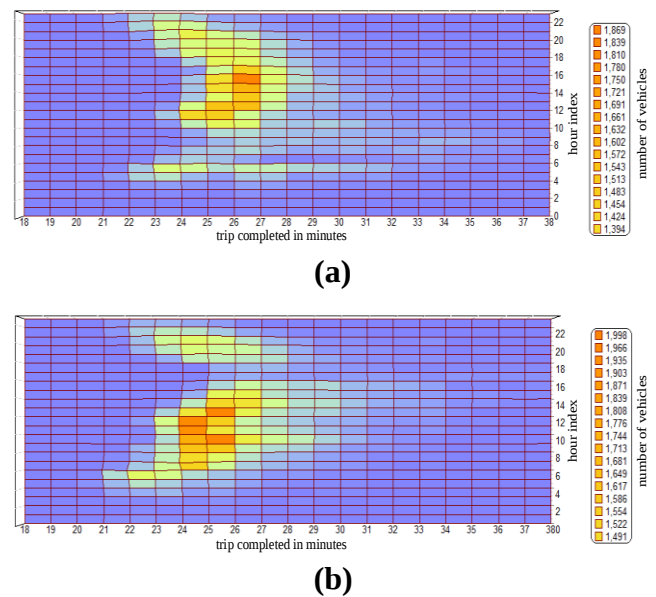
Next, we analyze the same thing with the opposite direction, from the intersection to the freeway as shown in Figure 4 (b). This data is

average traffic. In other words, the traffic is crowded and a spatiotemporal analysis needed to minimize this problem by extracting the most distinguishing pattern and provide a better solution. When we compare the total number of daily trips at one time, the pattern that appears is slightly different. This can be because this graph summarizes all trips at one period for example in three months. Temporal details will be lost because data is presented in smaller dimensions. The more we monitor and record the data, the more accurate the result provided. In figure 3, the highest travel intensity occurs at 6 a.m. and throughout the middle of the day for the direction of Tainan city and it reaches almost 1,000 vehicles. In contrast, for the direction of Kaohsiung, peak traffic occurred in the morning at 10 a.m. For the average number of travel times, we can make observations through Figure 5. It can be seen that the maximum travel time can reach around 180 minutes. This case can rarely happen, only occurs on trips carried out by heavy vehicles such as container trucks. While the average travel time is in the range of 20 - 25 minutes (yellow chart). This also means that traffic congestion occurs during midday. Whereas in the opposite direction, the density occurs at night with a maximum travel time of 150 minutes, but with similar average travel time around 20 - 25 minute.

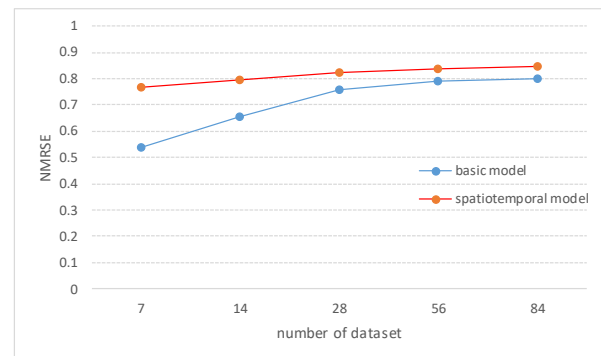
#### IV. DISCUSSION

Kaohsiung to Tainan via freeway (Figure 6 (a)). The average trip that was successfully carried out was distributed in the range of 21 to 32 minutes. This means that vehicles can pass through these two cities at that time. When compared with the distance between the two cities, you can get travel speeds in the range of 141 - 93 km/hour. This also similar to the reverse trip from Tainan to Kaohsiung (Figure 6 (b)). Interestingly, the distribution in 24 hours looks different. For trips to the south, there are the decreasing intensity of the traffic between 7 a.m. and 10 a.m. In contrast, this pattern on the North side is complement with that on the South side where the decline occurs at 4 p.m. up to 6 p.m. Meanwhile, the peak of the trip for both directions is relatively same at midday until 3 p.m. Another feature that can be extracted from the data is that the flow of vehicles on Sundays is at the lowest point compared to that of on the other days for both direction. Overall, vehicle speed to Tainan City is

faster with lower density than the reverse direction.



**Figure 7.** Frequency distribution of vehicle travel time in minutes vs the amount of traffic and daily time for (a) Tainan (b) Kaohsiung destination.



**Figure 8.** Comparison between the use of basic data and spatiotemporal data on an ANN model to forecast traffic on the next day

Spatiotemporal pattern extraction is a useful tool to identify or make a better analysis of traffic congestion, especially on an ever-changing data model. To provide comparative analysis, we test the data obtained in the basic format according to Figure 5 compared to the spatiotemporal data according to Figure 7. We use machine learning with a simple ANN method to train and evaluate the system so that it can predict patterns for one day ahead. The ANN model is built using Delphi and tested using several datasets i.e. 7, 14, 28, 56, and 84 respectively. About 80% of the dataset is used for training and the remaining 20% for testing. The results are presented as the Normalized version of Root Mean Square Error (NRMSE) which

correlated with system accuracy, shown in Figure 8. By using a simple ANN, data represented using the spatiotemporal model produces better accuracy compared to the normal model. This accuracy can be improved with the use of larger dataset with increasingly small differences for the two models. Overall, the spatiotemporal data model produces better-detailed data characteristics, the stronger correlation between variables and better accuracy when implemented on a decision-making system

## V. CONCLUSION

The vehicles flow entering Kaohsiung city from the suburb is higher than that which exits the city using the freeway. The frequency of vehicles passing the freeway is dominated by small vehicles, followed by trucks. In addition, rush hour at monitored intersection occurs from 7:00 to 20:00. This describing data can be done by parsing huge amount of Traffic Data Collection System provided by Taiwan government. We can use the MapReduce framework which is provided by Hadoop to process the data and distribute it into several nodes. By using a scheduler system such as Hadoop, the task can be finished efficiently. The data result which represented as a spatiotemporal model has benefits i.e. better-detailed data characteristics and the stronger correlation between variables. For further developments, the distinguish pattern that has been extracted can be used to predict the traffic flow by using machine learning techniques. Furthermore, these data can be used as a reference by the government for the construction of an advance freeway network by considering traffic on each intersection.

## REFERENCES

- [1] J. Wang and M. Hwang, "A novel approach to extract significant time intervals of vehicles from superhighway Gantry Timestamp sequences," 2017 International Conference on Applied System Innovation (ICASI), Sapporo, 2017, pp. 1679-1682.
- [2] F. Hagenauer, C. Sommer, R. Onishi, M. Wilhelm, F. Dressler and O. Altintas, "Interconnecting smart cities by vehicles: How feasible is it?," 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), San Francisco, CA, 2016, pp. 788-793.
- [3] Y. Chung and W. W. Recker, "Spatiotemporal Analysis of Traffic Congestion Caused by Rubbernecking at Freeway Accidents," in IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 3, pp. 1416-1422, Sept. 2013.
- [4] P. Wang et al., "Regional Detection of Traffic Congestion Using in a Large-Scale Surveillance System via Deep Residual TrafficNet," in IEEE Access, vol. 6, pp. 68910-68919, 2018.
- [5] C. Pasquale, S. Sacone, S. Siri and M. Papageorgiou, "Optimal Control for Reducing Congestion and Improving Safety in Freeway Systems," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 11, pp. 3613-3625, Nov. 2018.
- [6] P. Wang et al., "Forecasting Traffic Volume at a Designated Cross-Section Location on a Freeway From Large-Regional Toll Collection Data," in IEEE Access, vol. 7, pp. 9057-9070, 2019.
- [7] H. B. Celikoglu, "Flow-Based Freeway Travel-Time Estimation: A Comparative Evaluation Within Dynamic Path Loading," in IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 2, pp. 772-781, June 2013.
- [8] Z. Hou and X. Li, "Repeatability and Similarity of Freeway Traffic Flow and Long-Term Prediction Under Big Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 6, pp. 1786-1796, June 2016.
- [9] C. Zhang, O. Li, X. Tong, K. Ke and M. Li, "Spatiotemporal Data Gathering Based on Compressive Sensing in WSNs," in IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1252-1255, Aug. 2019.
- [10] W. Lee and S. Lee, "Improved Spatiotemporal Noise Reduction for Very Low-Light Environments," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 63, no. 9, pp. 888-892, Sept. 2016.
- [11] M. Maggioni, E. Sánchez-Monge and A. Foi, "Joint Removal of Random and Fixed-Pattern Noise Through Spatiotemporal Video Filtering," in IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 4282-4296, Oct. 2014.
- [12] S. Nurlatifani, A. Retnowardhani and D. N. Utama, "Fuzzy ecoDSS Development for Road Traffic Engineering," 2018 International Conference on Information Management and Technology (ICIMTech), Jakarta, 2018, pp. 345-349.
- [13] Y. Li, T. Guo, R. Xia and W. Xie, "Road Traffic Anomaly Detection Based on Fuzzy Theory," in IEEE Access, vol. 6, pp. 40281-40288, 2018.
- [14] L. Zhang, G. Zhu, P. Shen and J. Song, "Learning Spatiotemporal Features Using 3DCNN and Convolutional LSTM for Gesture Recognition," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, 2017, pp.

3120-3128.

- [15] Hou, Z., & Li, X. (2016). Repeatability and Similarity of Freeway Traffic Flow and Long-Term Prediction Under Big Data. *IEEE Transactions on Intelligent Transportation Systems*, 17(6), 1786–1796.
- [16] M. Das and S. K. Ghosh, "Deep-STEP: A Deep Learning Approach for Spatiotemporal Prediction of Remote Sensing Data," in *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1984-1988, Dec. 2016.
- [17] F. Tseng, J. Hsueh, C. Tseng, Y. Yang, H. Chao and L. Chou, "Congestion Prediction With Big Data for Real-Time Highway Traffic," in *IEEE Access*, vol. 6, pp. 57311-57323, 2018.
- [18] Z. Bai, B. Cai, W. ShangGuan and L. Chai, "Deep Learning Based Motion Planning For Autonomous Vehicle Using Spatiotemporal LSTM Network," 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 1610-1614.
- [19] J. Sewall, J. van den Berg, M. Lin and D. Manocha, "Virtualized Traffic: Reconstructing Traffic Flows from Discrete Spatiotemporal Data," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 1, pp. 26-37, Jan. 2011.